Computers & Education 72 (2014) 100–109

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Online self-assessment and students' success in higher education institutions

Maja Ćukušić*, Željko Garača [1], Mario Jadrić [2]

University of Split, Faculty of Economics, Cvite Fiskovića 5, 21000 Split, Croatia

ARTICLE INFO

Article history:
Received 19 July 2013
Received in revised form
28 October 2013
Accepted 31 October 2013

Keywords:
Higher education
Online self-assessment
Students' success
Dropout rates
Simulation

ABSTRACT

This paper validates effects of online self-assessment tests as a formative assessment strategy in one of the first year undergraduate courses. Achieved students' results such as test scores and pass rates are compared for three different generations for the same course but also judged against the exam results of other courses taught in the same semester. The analysis points out that there is a statistically significant difference between the groups for half-semester tests and exam pass rates after online self-assessment tests were introduced. Positive effects on students' success are approximated for the overall institution using a simulation model. Results point out that a small increase in pass rates could significantly impact the overall success i.e. decrease of dropout rates.

1. Introduction

The selection of assessment techniques and appropriate assessment tools is an integral part of planning the e-learning processes. Most modern Learning Management Systems (LMS) and Virtual Learning Environments (VLE) incorporate functionalities for developing and processing online tests. It is customary to use such system functions for self-assessment purposes and other types of evaluation in both full distance learning and hybrid learning courses. The use of computer-supported online assessment has been increasing in both summative and formative assessment areas (Bull & Dalziel, 2003; Bälter, Enström, & Klingenberg, 2013).

Interacting online requires educators to rethink online pedagogy so as to support meaningful (higher-order) learning and its assessment (Gikandi, Morrow, & Davis, 2011). The development of an assessment strategy should take into account the purpose of the evaluation results plus it should always be planned simultaneously with the preparation of learning activities. During the process, assessment can be perceived as one of the enablers of innovation and change in an educational setting. Sometimes, this can be a tough challenge in the context of compulsory education where the norm is even now the traditional assessment that is to say summative and teacher centered assessment (Granić, Mifsud, & Ćukušić, 2009). Growing in popularity, formative assessment is to a greater extent promoted as "a moment of learning" – it allows beneficial comparisons between the actual and referral levels of achievement and so the results can be used to identify gaps between the compared levels (Black & Wiliam, 1998). E-learning scenarios should therefore aim to maximize the potential of an LMS to implement a series of formative assessment strategies (i.e. assessments "for" learning).

It is not essential or even possible to integrate online assessment activities into every lesson since this process requires a significant amount of time and creativity in order to create good series of questions (Mödritscher, 2006) with respect to: the lesson topic; learning outcomes; the type of questions; number of participants; cheating prevention mechanisms; the time for assessment etc. Nevertheless, there are tests that may be incorporated as a part of almost any learning scenario such as quizzes at the beginning of the lesson (in terms of pointing out the gaps in foreknowledge of a group), during the lesson (measuring progress in understanding) or the end of the lesson (to assist the recap). In that respect, an important assessment strategy is self-assessment.

* Corresponding author. Tel.: +385 21430758; fax: +385 21430701.
E-mail addresses: maja.cukusic@efst.hr (M. Ćukušić), garaca@efst.hr (Ž. Garača), jadric@efst.hr (M. Jadrić).
[1] Tel.: +385 21430601; fax: +385 21430701.
[2] Tel.: +385 21430739; fax: +385 21430701.

0360-1315/$ – see front matter
http://dx.doi.org/10.1016/j.compedu.2013.10.018

The study presented hereinafter deals with effects of online self-assessment in one university course of hybrid type and discusses the potential extrapolation of the identified result to the whole system, to be exact higher education institution (HEI) by simulation method. Simulation modeling can support managerial approach (Greasley, 2003) to change key educational processes in order to understand and measure variations in performance indicators such as student success. An important characteristic of the method is the possibility of repeating the simulation runs with changing input parameters (i.e. increased course pass rates as a result of introducing self-assessment tests) and monitoring the impact of changes in outputs (overall HEI success in terms of number of students completing the study programs).

After presenting the theoretical and empirical background of the study in the second part of the paper, results of quantitative and comparative study are presented in the third part. Fourth part illustrates and discuses the potential impact of online self-assessments to the whole institution and the overall educational process outcomes. The final, fifth part of the paper concludes the study summarizing the results.

2. Effectiveness of online (self-) assessment tools

Self-assessment can result in major benefits both for teachers and students (McConnell, 2006), specifically it is more oriented to students, reduces some of the teachers' load, provides instant feedback and helps to remove certain "barriers" between teachers and students. Furthermore, the students become less dependent on their teachers, responsible and autonomous; they take on a more proactive role and develop self-confidence, while the teachers can evaluate the effects of their teaching efforts more accurately and objectively. Responsibility can be additionally promoted by permitting students to agree the rules for self-assessment among themselves or together with the teacher (McConnell, 2000). Hence students are actively involved in decision-making process about the evaluation criteria and the evaluation process of their own and other students' works. Involvement of students in their own assessment is an important part of preparation for life and work settings.

In order to prove the effectiveness of online teaching and learning tools and their impact to students' success many studies have been conducted in recent times. Examining available research on the subject of online formative assessment in higher education settings over the years, Gikandi et al. (2011) present a thorough literature review of 18 studies. Differing the mode (online and/or blended), theories and methodologies of the studies they conclude that effective online formative assessment can foster a learner and assessment centered focus through formative feedback. Only few of the presented studies are focusing on self-assessment tools such as quizzes that while other study impact of tools such as e-portfolios and online discussions.

Subsequently, only several studies focusing on the effectiveness of online self-assessment tools are presented in short as a research setting. Experimenting in a university course González, Jover, Cobo, and Muñoz (2010) randomly assigned students ($N = 121$) to different groups and instructed them on how to use the online tools. Afterward, they inspected the difference in the exam results where the exam consisted of two blocks of questions related to "online" and "offline" part of the curriculum. Their results clearly indicate that students improved their exam results for about 5% related to questions about the material presented online. Similarly, in a hybrid learning environment students ($N = 162$) that used the VLE system to prepare for the exam or to refresh their memory to the course content, in as little as two or more hours a week, performed lot better on the exam than those who did not use the system (Stricker, Weibel, & Wissmath, 2011).

Over a period of three academic years Llorente and Morant (2011) monitored the result and the effect of introducing interactive seminars through a VLE: starting from the preparatory period (the year without practicing this technique) to two years of conducting interactive seminars with 406 students. Introducing the online collaborative seminar resulted in an increase of the exam pass rate of 9.4%. At the same time, the results of student satisfaction increased from 6.86 (out of 10) to 8.24 to be precise over 20.12% compared to the year without the online seminars.

Due to their advantages the online multiple-choice (MC) questions are extensively used and their importance is increasing. But, with regards to the assessment phase of the e-learning process, an important question is the justification of MC online quizzes as an official, formal method of assessment especially compared to an oral exam. Ventouras, Triantis, Tsiakas, and Stergiopoulos (2011) studied the possibility of using online MC questions in high-stakes assessments such as final exams and proved that online MC quizzes as a method of learning assessment result in statistically equivalent results as oral examinations. It is important to note that authors actually demonstrate how particular types of MC assessment techniques (i.e. paired MC questions) taking place in controlled settings can be equivalent to a structured oral examination of students.

Still, online assessment in VLE is mostly used for informal purposes as part of formative assessment strategies. For that purpose, Henly (2003) found that the performance and level of use of self-assessment online questionnaires containing multiple-choice questions, short questions and matching questions decline during the semester. The author assumes that as the material becomes progressively more difficult the usage style changes as well – the undergraduate students use the self-assessment tools at first for recap and prepare before taking the self-assessment quiz. Later, students use the tests as guides to prepare for final exams. Despite the decline in the performance and degree of use, the top 10% of students used the self-assessment tools significantly more than the lowest performing 10%. Of course, it is debatable whether the use of tools has contributed to improved performance or the better performing students take on a more active approach. The very same question is investigated in a study by Buchanan (2001) who found that the level of use of online self-assessment tools is a very good predictor of exam results, even if the degree of class attendance as a reflection of a student's approach and work ethics is constant. Thus, the results indicate that online formative self-assessment has the potential to improve learning outcomes.

Correspondingly, Klecker (2007) found that her randomly grouped postgraduate students have different results. One group had to take one weekly online formative test and other one was a control group. Specifically, the students who had to take weekly self-assessment tests performed significantly better on the exam.

Wilson, Boyd, Chen, and Jamal (2011) analyze the effectiveness of computer-assisted assessment by providing informal and formal feedback to students and relate it to exam performance. Other than achieving positive student feedback (e.g. they liked gaining control over own learning, setting own pace and comfortable testing area with having enough time for reflection) the exam results showed that students who conducted self-assessment achieved over 10% better results than the others. de Marcos et al. (2010) point out the benefits in terms of user satisfaction and better results. Three different experimental student groups were involved in the research, age 14 to 21. The results give reason to believe that online self-assessment tools affect and improve student performance, particularly among the younger students.

The above-presented studies highlight the effectiveness of certain teaching and learning methods, but none of the research presented inspects the link between the use of self-assessment tests via VLEs in higher education setting over the course of several years.

The use of simulation to model and analyze variations in HEI performance indicators such as students' success is novel and is used as an addendum in order to explore the potential for institution-wide implementation of self-assessment tests.

3. Research methodology

The overall aim of the research was to design, implement and monitor the changes in the teaching process in the first-year undergraduate course Information Technology at Faculty of Economics, University of Split. Significant improvement of the exam results and test scores was expected after introduction of online self-assessment tests.

3.1. Introducing online self-assessment tests via Moodle system

During the year 2008/2009 the Faculty introduced the Moodle system as a preferred VLE. The core module "Quiz" implemented and extensively used for purposes of this study and development of self-assessment tests is one of the most complex and the most flexible parts of the Moodle system.

Even before introducing Moodle the course Information Technology was set up to facilitate continuous tracking of students' progress by using a model of accumulating points allowing students to collect points for different course activities such as lab exercises, half-semester tests and different bonuses. In 2009/2010 online self-assessment tests were introduced. A total of 8 scheduled online tests, one per topic/book chapter were set up. The tests could be accessed immediately after the face to face lectures within the timeframe of two weeks. In 2009/2010 each test consisted of 10 MC questions totaling in maximum 10 points and limited to 30 min while in 2010/2011 the tests consisted of 20 questions totaling in maximum of 20 points and limited to 40 min.

Each MC question presents 6 choices with the number of correct responses varying from 1 to 6. Every time a test is started by a student, both questions and the order of choices are generated randomly at startup from 320 questions in question bank.

Moodle automatically scores answers immediately after the test is submitted with incorrect answers bringing negative points so the balance per question cannot be less than zero. Number of test attempts within two weeks is unlimited but with a time delay of 1 h between two attempts. Students are required to achieve at least 60% in 6 out of 8 tests.

3.2. Collecting and analyzing data from databases

The dataset contains data from two different systems in order to verify the hypothesized improvement of the exam results and test scores (illustrated in Fig. 1):

- Moodle database provides data about tests and user behavior such as test results, number of attempts, number of resource views, assignment submissions etc. For the purposes of reporting on the results and user behavior two different modules are used – the Grader report and the Log.
- Information System of Faculty of Economics (ISFE) database contains the grades and exam results. Although certain reports about exam success rates and results are available within the system, the data presented in this analysis was largely obtained via custom queries to the central database.

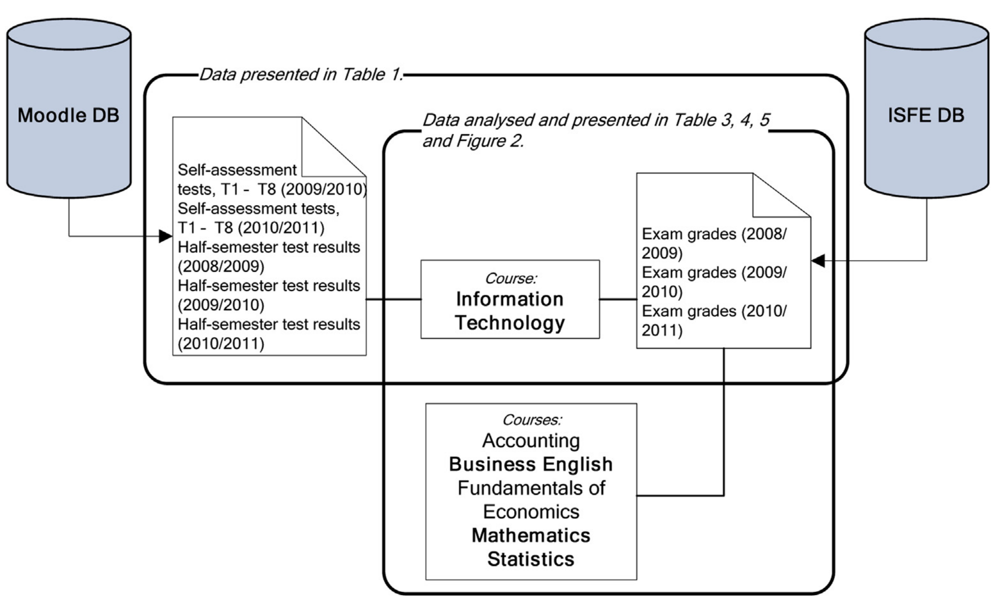

Fig. 1. Dataset used in the study mapped to data sources.

The final dataset for one academic year contains information about the number of attempts and the results of each self-assessment test (in total 8), then the number of points on the first and second half-semester test, the results achieved in practical tests, exam grades from the first semester for 6 courses including Information Technology. With the aggregate indicators such as total number of activities within the Moodle system the dataset for one year exceeds 20.000 entries. Results of three groups of students enlisted in the first semester of undergraduate studies in three academic years were tracked. During the three years the course was organized in the following way:

- two half-semester tests and an exam (2008/2009),
- two half-semester tests, 8 self-assessment tests with 10 questions and an exam (2009/2010),
- two half-semester tests, 8 self-assessment tests with 20 questions and an exam (2010/2011).

Controlling the conditions of a study as required by the experimental method is rarely possible outside the laboratory. For practical reasons, it was not possible to randomly determine the control and experimental group. Testing the effectiveness of educational interventions like this one for an entire generation of students (and not only for a part of it) is justified not only because of practical but for ethical reasons as well. In quasi-experiments like this one the shortcomings of the design should be well specified, but the lack of true control should be mitigated using statistical control (usually by analysis of covariance) so as to assess and eliminate the impact of otherwise uncontrollable external factors on the outcome of the study. As there is no control group the results are weighted against the academic year 2008/2009 where there was no self-assessment but also judged against the results of the other courses taught the same academic year that have not changed their course syllabi, teachers or assessment strategy. With that regard control is in that sense both horizontal (in relation to exam results of other courses in the same generation of students) and vertical (with respect to the results among different generations of students for the same course and other courses as well). To be more precise, horizontally within the same academic year the structure of subjects is the same (the same students taking different courses); the courses are equivalent in terms of numbers of students, group sizes, ECTS points, teaching hours and all are part of the same study program curricula. Obviously, teachers are not the same for different courses. Vertically, within several academic years only four courses were the same in terms of teachers and teaching material while two courses were eliminated due to some changes. Students are different but with very similar background in terms of gender and age structure. Exam results and pass rates are considered only for students who were enrolled in the course for the first time; the results for the re-enrolled students regardless of their success in course activities are excluded.

In addition to calculating the correlation coefficients r and ρ, differences between groups are tested with regards to half-semester tests and exams for all three generations of students (total of 1.379 students). For that purpose Kruskal–Wallis's H test and one-way ANOVA F-test were used along with post-hoc test analyses.

3.3. *Extrapolating results using simulation experiments*

The second part of the study draws on the results of the first one by creating a simulation model that allows the optimization of key performance indicators of educational processes in higher education, taking into account positive effects of formative self-assessment and the potential for institution-wide implementation. Thus, simulation modeling is used as a cost-effective method of analyzing “what-if” scenarios such as what would happen if similar effects were achieved in other courses as well. In case similar effects for other courses are assumed an increase in pass rates and consequently lower drop our rates are to be expected. The simulation model enables us the verification of this thesis. In the simulation experiment we assume that it is possible to implement self-assessment tests in circa half of HEI courses to test the impact on the entire system (overall HEI success in terms of number of students completing the study programs) by a relatively small increase in the course pass rates.

Considering the specific types of simulation modeling, the capabilities of modern simulation tools and several rules and recommendations regarding the simulation steps, the last part of the study progressed through the following stages: planning the simulation process, data collection and analysis, the construction of a simulation model, model verification and validity assessment, experimentation, and results' analysis. These stages are shortly presented here and in the chapter 4.

The planning process started with setting the objective of the model in terms of the analysis of potential opportunities specifically the examination of the impact of empirically validated increased pass rates in particular courses, all formulated through key indicators such as dropout rates and time to complete the studies. To develop and test the model the scope, the level of detail, the type of experimentation and the format of the expected results were defined. The model was conceptualized based on the Faculty's regulations and policies thus mirroring the logic of enrolling into courses (with associated ECTS), attending the courses (lectures and exercises) and taking exams. Enrolling in the HEI is out of the scope. One generation of undergraduate students is observed through 5 years. During that period the students are required to enroll and get/pass 180 ECTS overall, enrolling maximum 60 ECTS every year. The students are allowed to take the exam from the same course up to four times and in case they do not succeed they re-enroll in the course next year. They are allowed to re-enroll just once. In case they do not pass the exam in additional four attempts they drop out. Leaving the system can be twofold – finishing studies or dropping out. Likewise, students who re-enroll courses should complete the whole program within five years, otherwise their status expires. Accordingly, only the basic components of the system contributing to the objective of simulated process are modeled thus limiting the scope and the model complexity. Certain activities are abstracted so the model does not replicate course requirements such as for example attendance or course prerequisites in terms of coupled courses, etc. Likewise, the model does not cater for different grades; it only simulates pass/fail exam success based on available data. Because of simplicity and universality the courses are given generic names (Subject 01 to Subject 30) and all hold 6 ECTS.

Data for simulation and validation of the model is based on transactional data on students and exam results, study completion and dropout rates, all stored in ISFE. The input variables such as average pass rates per courses and years of study shape the model while the output variables describe how the system responds to a specific set of input variables and represent the result of simulation comparable to data from the real system.

4. Results

4.1. Examining student- and course-level indicators

Table 1 shows the mean and the deviation of students' results for self-assessment test, half-semester tests and the exams in academic year 2009/2010 and 2010/2011. The maximum number of points per self-assessment test was 10 (2009/2010) and 20 (2010/2011) with students achieving on average 25.46 (63.65%) and 49.02 (61.27%) respectively in the first four self-assessment tests, and much less 21.01 (52.53%) and 37.56 (46.96%) respectively in other four test. Overall, the students' average scores were 46.37 (57.96%) and 86.58 (54.12%) out of maximum number of points in self-assessment test.

The correlations between students' results in formative (online self-assessment tests) and summative assessments (half-semester tests and exams) are presented in Table 2. Several statistically significant correlations were found. Note that all correlations are positive, indicating that higher scores on the formative assessments are associated with higher scores in summative assessments. Highest correlation is between exam scores and total number of points in self-assessment tests $\rho(183) = .397$, $p < .001$. Lower scores in other four self-assessment tests (T5-T8) could be accounted for a slight correlation $\rho(237) = .154$, $p < .005$ between exam scores and number of points in other four self-assessment tests (T5–T8). This can be the result of decreasing use of self-assessment tests. Faced with similar results Henly (2003) assumes that as the material progressively becomes more difficult the purpose and the usage changes. Students initially use the test for repeating what they have learnt and prepare themselves for the self-assessment test, while later they use it mostly as guidance for examination or final exams. Another reason for lower scores may be the fact that large number of students solved first six tests with the score above 60% thus completing the course requirement so some of them have not even started the last two tests.

In order to inspect the effect of self-assessment tests and exam results in relation to results of other courses ANOVA was used. Other than the course Information Technology courses Mathematics, Fundamentals of Economics, Business English, Accounting and Statistics are taught within the same semester (first semester of the first year of undergraduate studies). To compare the three academic years it was very important that other courses have not changed anything with regards to course planning or delivery such as their course syllabi, teachers in charge or assessment strategy. After careful inspection two courses were excluded from the analysis due to changes. The course Fundamentals of Economics introduced a different assessment method while in Accounting one teaching unit was completely dropped from the course plan in 2009/2010. Moreover, not all students have to enroll into Accounting in the first semester. In Mathematics the title of the assessment changed, but in reality the assessment was organized in the same manner and for the same learning material although the colloquiums were now labeled as half-semester tests organized in the same timeframe.

The pass rates were collected for the first exam period (February 2009, 2010 and 2011) since it is to expect the effects of online self-assessment tests would have a short-term impact on the exam results. As for the pass rates for the four courses the highest mean pass rate is in Business English (45%) and lowest in Mathematics (14%). Information Technology and Statistics have similar pass rates, 17% and 20% respectively.

Kruskal–Wallis non-parametric ANOVA was used to determine if the pass rates for the three academic years are statistically different. The results of a Kruskal–Wallis test (Table 3) were significant - the mean pass rates are significantly different for the three courses: Information Technology ($H = 12.3$, 2 d.f., $p = .002$), Mathematics ($H = 30.285$, 2 d.f., $p < .001$), Statistics ($H = 48.699$, 2 d.f., $p < .001$). The test was not significant for Business English course ($H = 3.627$, 2 d.f., $p = .163$). Descriptives are presented in Table 4 and illustrated in Fig. 2.

Since the test is statistically significant, the least significant difference (LSD) as a post-hoc test was used (Table 5). Post-hoc test indicates that the pass rates for Information Technology course within the first year of introducing the self-assessment tests (2009) were significantly better (+8.7%, $p < .001$), the same as in the second year (2010) (+5%, $p = .049$) compared to a year without the self-assessment tests (2008) (13%). There are no significant differences between 2009 and 2010 when the number of questions increased from 10 to 20 ($p = .131$). At the same time the pass rates for the course Information Technology increased, there is no significant difference for Mathematics in 2009, but in 2010 there is a significant difference, a decrease of 11% ($p < .001$). Significant difference for the course Statistics is observed in 2009, a drop of 15% pass rate ($p < .001$).

Other than analyzing mean pass rates, numbers of points on two half-semester tests were analyzed using one-way ANOVA as well. There is a statistically significant difference between the groups according to ANOVA test results for the first half-semester test ($F(2, 1323) = 7.937$, $p < .001$) and the second half-semester test ($F(2, 1082) = 17.805$, $p < .001$). Post-hoc tests (Tukey HSD) indicate that the results of the half-semester tests in 2009 and 2010 after the self-assessment tests were introduced were significantly higher, to be exact +.71 points ($p = .001$) in the first and +1.64 points ($p < .001$) in the second test compared to a year without the self-assessment tests (average 7.62 and 8.47 points). Again, there is no significant difference between the number of points 2009 and 2010 when number of questions increased from 10 to 20.

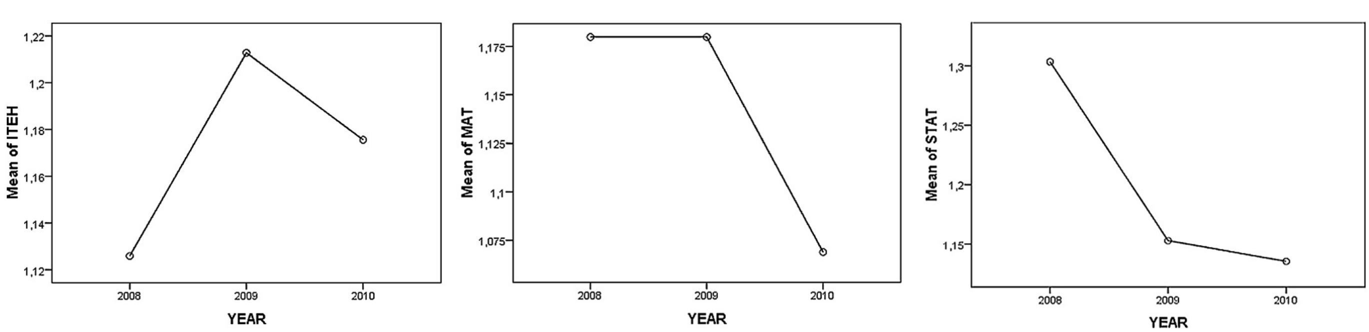

Fig. 2. Means plots of the pass rates for the three academic years.

Table 1
Descriptive statistics for the observed variables in 2009/2010 and 2010/2011.

Year	Variable	Mean	Standard deviation	%	*N*
2009/2010	Number of points in first four self-assessment tests (T1–T4)	25.4590	4.99711	63.65	484
	Number of points in other four self-assessment tests (T5–T8)	21.0141	6.54966	52.54	484
	Total number of points in self-assessment tests	46.37	9.434	57.96	484
	Number of points in first half-semester test	8.25	2.839	41.25	471
	Number of points in second half-semester test	9.25	3.840	46.25	374
	Exam score	1.73	1.144		237
2010/2011	Number of points in first four self-assessment tests (T1–T4)	49.0198	12.57597	61.27	450
	Number of points in other four self-assessment tests (T5–T8)	37.5649	16.92456	46.96	450
	Total number of points in self-assessment tests	86.5848	25.61631	54.12	450
	Number of points in first half-semester test	8.33	2.910	41.65	421
	Number of points in second half-semester test	10.11	3.307	50.55	328
	Exam score	1.69	1.029		183

Table 2
Correlation coefficients of observed variables in 2009/2010 and 2010/2011.

Year	Variables	*r*	*ρ*	Sig.	*N*
2009/2010	Number of points in first four self-assessment tests (T1–T4) ↔ Number of points in first half-semester test	**0.295**[b]		.000	471
	Number of points in other four self-assessment tests (T5–T8) ↔ Number of points in second half-semester test	0.176[b]		.001	374
	Exam score ↔ Number of points in first four self-assessment tests (T1–T4)		**0.311**[b]	.000	237
	Exam score ↔ Number of points in other four self-assessment tests (T5–T8)		0.154[a]	.018	237
	Exam score ↔ Total number of points in self-assessment tests		**0.271**[b]	.000	237
2010/2011	Number of points in first four self-assessment tests (T1–T4) ↔ Number of points in first half-semester test	**0.354**[b]		.000	421
	Number of points in other four self-assessment tests (T5–T8) ↔ Number of points in second half-semester test	**0.290**[b]		.000	328
	Exam score ↔ Number of points in first four self-assessment tests (T1–T4)		**0.277**[b]	.000	183
	Exam score ↔ Number of points in other four self-assessment tests (T5–T8)		**0.369**[b]	.000	183
	Exam score ↔ Total number of points in self-assessment tests		**0.397**[b]	.000	183

Bold text is significant at the .01 level.
[b] Correlation is significant at the .01 level (2-tailed).
[a] Correlation is significant at the .05 level (2-tailed).

Table 3
Kruskal–Wallis test comparing pass rates for the three academic years.

	Information Technology	Mathematics	Business English	Statistics
Chi-square	12.311	30.285	3.627	48.699
df	2	2	2	2
Asymp. Sig.	.002	.000	.163	.000

4.2. *Examining institution-level indicators*

The simulation model allowing extrapolation of validated positive effects of formative self-assessment to the whole institution was developed as a second part of the study. The model creatively explores the potential for institution-wide implementation of such course modifications.

The simulation model illustrated in Fig. 3 was developed in ExtendSim8 from Imagine That Inc. It consists of 15 different ExtendSim8 types of blocks. In addition to *create* and *exit* blocks there are those that set attribute values, control statuses or queues and provide numerical or graphical outputs used for model verification and present the results of simulations. The final model is made from more than

Table 4
Distribution of the pass rates for the three academic years.

						95% confidence interval for mean	
		N	Mean	Std. deviation	Std. error	Lower bound	Upper bound
ITEH	2008	445	1.13	.332	.016	1.09	1.16
	2009	484	1.21	.410	.019	1.18	1.25
	2010	450	1.18	.381	.018	1.14	1.21
	Total	1379	1.17	.378	.010	1.15	1.19
MAT	2008	445	1.18	.384	.018	1.14	1.22
	2009	484	1.18	.384	.017	1.15	1.21
	2010	450	1.07	.254	.012	1.05	1.09
	Total	1379	1.14	.351	.009	1.13	1.16
STAT	2008	445	1.30	.460	.022	1.26	1.35
	2009	484	1.15	.360	.016	1.12	1.19
	2010	450	1.14	.343	.016	1.10	1.17
	Total	1379	1.20	.397	.011	1.17	1.22

Table 5
Multiple comparisons of the pass rates for the three academic years.

			Mean difference (I-J)	Std. error	Sig.	95% Confidence interval	
						Lower bound	Upper bound
ITEH	2008	2009	−.087[a]	.025	.000	−.14	−.04
		2010	−.050[a]	.025	.049	−.10	.00
	2009	2008	.087[a]	.025	.000	.04	.14
		2010	.037	.025	.131	−.01	.09
	2010	2008	.050[a]	.025	.049	.00	.10
		2009	−.037	.025	.131	−.09	.01
MAT	2008	2009	.000	.023	.999	−.04	.04
		2010	.111[a]	.023	.000	.07	.16
	2009	2008	.000	.023	.999	−.04	.04
		2010	.111[a]	.023	.000	.07	.16
	2010	2008	−.111[a]	.023	.000	−.16	−.07
		2009	−.111[a]	.023	.000	−.16	−.07
STAT	2008	2009	.150[a]	.026	.000	.10	.20
		2010	.168[a]	.026	.000	.12	.22
	2009	2008	−.150[a]	.026	.000	−.20	−.10
		2010	.017	.026	.498	−.03	.07
	2010	2008	−.168[a]	.026	.000	−.22	−.12
		2009	−.017	.026	.498	−.07	.03

[a] The mean difference is significant at the .05 level.

1.200 blocks. The model was made in three hierarchical levels. Only the highest-level, the basic model is displayed. The second level consists of hierarchical blocks Enrolling into courses and Taking exams, each with 30 courses. Course-level is the third-level of the model.

During the assessment phase one ascertains that the simulation model performs as expected. Verification of the model was done iteratively, checking the result/output of simulation at every simulation run. The result is presented through Plotter block. Fig. 4 illustrates results of a simulation experiment for 100 students. The left side contains data on the completion of study programmes and the right dropouts at various time points (per year) cumulatively. ExtendSim8 functionalities that are particularly useful for model verification are live animation and History block so the correctness of simulation flow can be ascertained and all values (years of study, accumulated ECTS, etc.) could be tracked for every student. The model has been modeled and verified according to rules and regulations of undergraduate studies at Faculty of Economics in Split.

In addition to verification the reliability with which the model reflects the simulated system was validated. Real data from the ISFE system for 2006/2007 generation was acquired (Table 6) since in 2011 all students from that generation had to either complete the study or drop out, all consistent with the Bologna regime of studying.

Since the simulation model contains a number of average parameters some deviations in the simulation results are possible. Probabilities of passing/failing the exams for a given year of the study programme are the same for all courses of the year, but for the three different levels of students. The categorization of students and probabilities are consistent with Faculty's data. Three simulations were run (Table 7) each simulating a period of six years for 100 students with the same probabilities of passing the course exams in all four exam periods for:

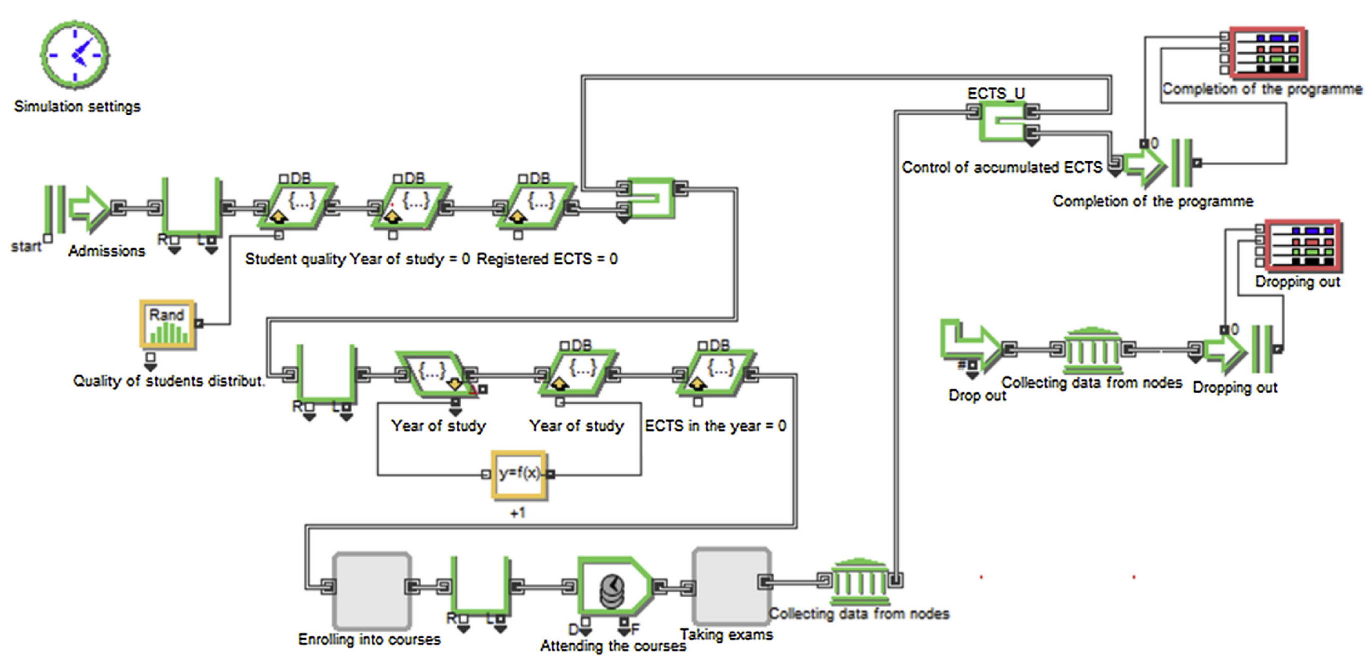

Fig. 3. First level of HEI simulation model.

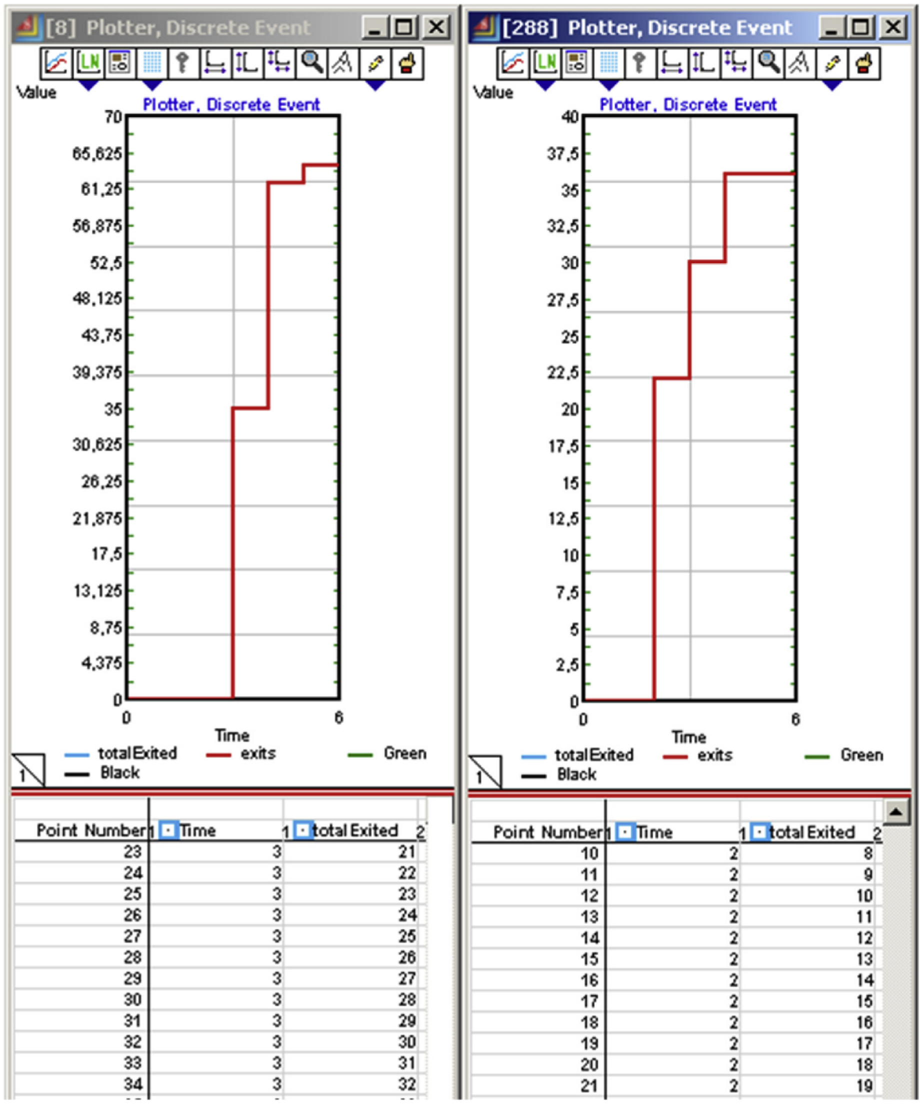

Fig. 4. Result of a simulation experiment.

- A level students (30% of the population) – 1st year 60%, 2nd year 70% and 3rd year 80%;
- B level students (60% of the population) – 1st year 29%, 2nd year 31% and 3rd year 38%;
- C level students (10% of the population) – 1st year 5%, 2nd year 5% and 3rd year 5%.

A simulation with 300 students was run to simulate the actual numbers. Results are presented in Table 8. The percentage of students who complete their studies (54.33%) is almost identical to the real data (55.49%) as is the number of dropouts (45.67% compared to 44.51%). The model accurately reflects the simulated system and therefore can be used for simulation experiments.

Conducting experiments is the real purpose of simulation modeling by observing the most relevant and dynamic characteristics of the system in order to find the impact of input values or structure changes on the behavior of the system. The scenario presented hereinafter only affects the values, not the structure. The previous section postulated and validated the effects of modifications of teaching processes in terms of introducing formative self-assessment in one course. The simulation model enables measuring the potential impact if similar

Table 6
Data used for validation of the simulation model.

Generation 2006/2007		Completed the programme		Dropped out	
Academic year	Students enrolled	*N*	%	*N*	%
2006/2007 (admission year)	510				
2007/2008	468			160	31.37%
2008/2009	331	171	33.53%	39	7.65%
2009/2010	144	100	19.61%	28	5.49%
2010/2011	12	12	2.35%		
In total		283	55.49%	227	44.51%

Table 7
Results of three simulations used for comparison with the real data.

	Completing the programme (out of 100 students)			Dropouts (out of 100 students)		
	After 3rd year	After 4th year	After 5th year	After 2nd year	After 3rd year	After 4th year
1st simulation	34	15	5	29	11	6
2nd simulation	32	17	7	36	3	5
3rd simulation	31	20	2	29	10	8

Table 8
Simulated data for 300 students for validation of the model.

Academic year	Students enrolled	Completing the programme		Dropouts	
		N	%	*N*	%
X (1st admission)	300				
X +1				94	31.33%
X +2		97	32.33%	24	8.00%
X +3		52	17.33%	19	6.33%
X +4		14	4.67%		
In total		163	54.33%	137	45.67%

Table 9
Results of three simulations for 100 students.

	Completing the programme (out of 100 students)			Dropouts (out of 100 students)		
	After 3rd year	After 4th year	After 5th year	After 2nd year	After 3rd year	After 4th year
1st simulation	38	23	4	24	5	6
2nd simulation	39	19	6	21	7	8
3rd simulation	37	24	2	28	5	4

Table 10
Simulated data for 300 students.

Academic year	Students enrolled	Completing the programme		Dropouts	
		N	%	*N*	%
X (1st admission)	300				
X +1				73	24.33%
X +2		114	38.00%	17	5.67%
X +3		66	22.00%	18	6.00%
X +4		12	4.00%		
In total		192	64.00%	108	36.00%

effects were achieved in other courses as well. In short, an increase in pass rates and lower drop our rates are to be expected. But, it is not reasonable to expect implementation of self-assessment tests in all HEI courses so for the purposes of the simulation experiment we assume that it is possible for half of the courses. Likewise, it is reasonable to expect that the passing rates of A-level students (30% of the population) and the C-level (10% of the population) will not significantly change. So it makes sense to test the impact on the entire system by a relatively small increase in the pass rates (conservatively 3% as opposed to 9% supported by the statistical results), for half of the courses and for the largest part of the population that would be encouraged to use formative assessments such as self-assessment tests.

With respect to new input values three simulations were run for 100 students:

- A level students (30% of the population) – 1st year 60%, 2nd year 70% and 3rd year 80%;
- B level students (60% of the population, 15 courses) – 1st year 29%, 2nd year 31% and 3rd year 38%;
- B level students (60% of the population, 15 courses) – 1st year 32%, 2nd year 34% and 3rd year 41%;
- C level students (10% of the population) – 1st year 5%, 2nd year 5% and 3rd year 5%.

Simulation results involving random variables represent only some of the possible values. Therefore, it is important to perform more simulation experiments and to statistically process the results to provide a valid basis for conclusions. Results are presented in Table 9 and Table 10. The percentage of students who complete their studies (64.00%) is significantly higher than current indicators (55.49% real life, 54.33% simulated) and correspondingly the number of dropouts is lower (36.00% compared to 44.51% in real life and 45.67% simulated).

5. Conclusion

Achieved students' results from a first year undergraduate course were compared for three different generations. After the implementation of online self-assessment tests in 2009/2010, the correlation between the exam results and the number of points in self-assessment tests was detected. By the end of the second year the self-assessment tests were used, the real and significant correlation between the points achieved in self-assessment tests and the number of points achieved in both half-semester examinations and exam results was noted.

Several similar and recent studies (Gikandi et al., 2011; de Marcos et al. (2010); Wilson et al., 2011) support the hypothesis but none correlates the effects of self-assessment tests on the population of students over several years to control variables such as other courses in the same year. Comparing the students' results with other courses statistically significant difference was detected in course pass rates for the same timeframe. In the same period when the pass rates for the course Information Technology increased, results in Mathematics and Statistics decreased significantly. Already rather low rates and overall exam results (including marks) dropped even more from one generation to another due to lower admission criteria. Therefore the results of the change in the course and introduction of formative assessment tools can be interpreted as follows: in case the criteria for the admission that is to say the "level of inputs" remained the same one could expect the same pass rates in the second year. However, because of the evident decline in the student "quality", the decline of pass rates in Information Technology is much slighter because the students put additional effort and spend more time studying by using online self-assessment tests.

So, given the link between launching online self-assessment with the exam results and pass rates it can be concluded that such a teaching and ICT improvement of an educational process positively affects students' success and the overall success indicators of HEIs.

The simulation experiment investigates the effect of increasing pass rates based on presumptions that formative assessment influences students' success. The effect is inspected for only half of the study courses and the largest, average group of students. A relatively conservative 3% increase in course pass rates (as opposed to the observed 9%) has a powerful effect - according to simulated data the percentage of students who complete their studies is almost 10% higher.

It is hoped that the empirical evidence, demonstrating both the existence of a link between the online self-assessment and the student success, as well as the potential impact of such shift to formative assessment methods, provide higher educational institutions and other actors in the e-learning arena with the incentive to address more thoroughly the methodological and organizational issues of their e-learning initiatives. The message is clear – whatever one tries to teach, the success of students seems to be reinforced by the shift to formative assessment.

One of the strengths of this study lies mainly in the large number of participants (data collected for over 1.300 students) enrolled in several university level courses. Another strength is a novel combination of the methods – first, statistical method was used to inspect the effect of the quasi-experiment and then the simulation method was used to generalize and measure the implications for the whole institution. More specifically, we postulated and proved that the link between ICT support through online self-assessment tests and the observed exam results is positive. Moreover we built a simulation model that enables measuring the effect of better exam results (as a result of formative assessment strategies) on key performance indicators of a HEI (such as dropout rates and time to complete the studies). The model can be used to analyze the potential of different interventions in similar HEIs since it accurately simulates the whole end-to-end study process.

Limitations of the research are mostly related to internal and external validity of the study. We discussed this in detail in Section 3.2. Internally, the validity of the study could be disrupted by the longer timeframe from the beginning to the end of the study where different events that could affect the outcome were possible. It was impossible to control the environment and monitor all the events, especially in educational settings. External validity is reflected in lack of possibility to generalize beyond the studied group because of social, geographic and demographic differences, environment characteristics and temporal distance. In addition to two-level control, this weakness of the study could have been mitigated by incorporating some qualitative data collection such as through structured interviews which could have helped to gain deeper understanding of data obtained from the study.

Nevertheless, authors hope that the study results would contribute to the ongoing debate about the effectiveness of formative assessment strategies in higher education settings and that it will prove useful to researchers and all those who are direct stakeholders in higher education.

References

Bälter, O., Enström, E., & Klingenberg, B. (2013). The effect of short formative diagnostic web quizzes with minimal feedback. *Computers & Education, 60*(1), 234–242. http://dx.doi.org/10.1016/j.compedu.2012.08.014.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in education: principles. *Policy & Practice, 5*(1), 7–74. http://dx.doi.org/10.1080/0969595980050102.

Buchanan, T. (2001). The efficacy of a world-wide web mediated formative assessment. *Journal of Computer Assisted Learning, 16*(3), 193–200. http://dx.doi.org/10.1046/j.1365-2729.2000.00132.x.

Bull, J., & Dalziel, J. (2003). Assessing question banks. In A. Littlejohn (Ed.), *Reusing online resources: A sustainable approach to e-learning*. London: Kogan Page Limited.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: a review of the literature. *Computers & Education, 57*(4), 2333–2351. http://dx.doi.org/10.1016/j.compedu.2011.06.004.

González, J. A., Jover, L., Cobo, E., & Muñoz, P. (2010). A web-based learning tool improves student performance in statistics: a randomized masked trial. *Computers & Education, 55*(2), 704–713. http://dx.doi.org/10.1016/j.compedu.2010.03.003.

Granić, A., Mifsud, C., & Ćukušić, M. (2009). Design, implementation and validation of a Europe-wide pedagogical framework for e-Learning. *Computers & Education, 53*(4), 1052–1081. http://dx.doi.org/10.1016/j.compedu.2009.05.018.

Greasley, A. (2003). Using business-process simulation within a business-process reengineering approach. *Business Process Management Journal, 9*(4), 408–420. http://dx.doi.org/10.1108/14637150310484481.

Henly, D. C. (2003). Use of web-based formative assessment to support student learning in a metabolism/nutrition unit. *European Journal of Dental Education, 7*(3), 116–122. http://dx.doi.org/10.1034/j.1600-0579.2003.00310.x.

Klecker, B. M. (2007). The impact of formative feedback on student learning in an online classroom. *Journal of Instructional Psychology, 34*, 161–165.

Llorente, R., & Morant, M. (2011). Data mining in higher education. In K. Funatsu (Ed.), *New fundamental technologies in data mining*. Rijeka: InTech.

de-Marcos, L., Hilera, J. R., Barchino, R., Jiménez, L., Martínez, J. J., Gutiérrez, J. A., & Otón, S. (2010). An experiment for improving students performance in secondary and tertiary education by means of m-learning auto-assessment. *Computers & Education, 55*(3), 1069–1079. http://dx.doi.org/10.1016/j.compedu.2010.05.003.

McConnell, D. (2000). *Implementing computer supported cooperative learning*. London: Kogan Page.

McConnell, D. (2006). *e-learning groups and communities: Imagining learning in the age of the internet*. OU Press.

Mödritscher, F. (2006). e-Learning theories in practice: a comparison of three methods. *Journal of Universal Science and Technology of Learning*3–18. Appeared: 28/5/06.

Stricker, D., Weibel, D., & Wissmath, B. (2011). Efficient learning using a virtual learning environment in a university class. *Computers & Education, 56*(2), 495–504. http://dx.doi.org/10.1016/j.compedu.2010.09.012.

Ventouras, E., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2011). Comparison of oral examination and electronic examination using paired multiple-choice questions. *Computers & Education, 56*(3), 616–624. http://dx.doi.org/10.1016/j.compedu.2010.10.003.

Wilson, K., Boyd, C., Chen, L., & Jamal, S. (2011). Improving student performance in a first-year geography course: examining the importance of computer-assisted formative assessment. *Computers & Education, 57*(2), 1493–1500. http://dx.doi.org/10.1016/j.compedu.2011.02.011.