Analysing Students' Behaviour Patterns in Online Assessment

Mario Jadrić, Maja Ćukušić, Željko Garača
Faculty of Economics
University of Split
Cvite Fiskovića 5, 21000 Split
{jadric, maja.cukusic, garaca}@efst.hr

Abstract*. The aim of the study was to discover students' behaviour patterns based on the data recorded in a learning management system and the links to their results achieved in a specific e-learning course. In order to identify behavioural patterns during formative online assessment test (multiple access allowed), the collected data are analysed both within and between two generations of students who participated in the e-courses. The students in 2016/17 accessed the e-course more often, achieved better results and accessed the content of the e-course to a greater extent. At the same time, for the same generation, fewer test attempts are noted. Based on the insight into the structure of the content that students have accessed, the best results are particularly positively linked with access to different e-course content.*

Keywords. e-course, online assessment, students' behaviour patterns, educational data mining, learning analytics.

1 Introduction

Latest reports confirm a steady increase in the demand and online course offerings in higher education. As an example, almost 5.3 million students took at least one online course in the year 2013/14 (Murphy & Stewart, 2017). Some of the reasons for this increase in online course offering are related to institutional limitations (e.g. a lack of classroom space, educational costs) but also to a steady student demand for flexible learning options and expectations.

In parallel, and in part as a result of this increase in online course offerings, there is a growing interest of researchers to automatically analyse data generated by students in an online education environment. One of the reasons for this interest is the expanding availability of data (i.e. logs of student activities), which provide ample opportunities to discover behavioural patterns, and any deviations in the student's expected behaviour. Furthermore, it is possible to build prediction models that can calculate probabilities of students' behaviour, all in order to provide timely support to students. The results are typically of more use to teachers since they can use them to tailor/adjust/optimise their teaching and learning strategies and adapt the online education environment.

Without adequate ICT support (usually in the form of data mining tools) the analysis of the generated data from a course with lots of students and countless activities can be a challenging task for the instructor (Burgos et al., 2017). In addition to a great volume of the data, to get the full picture, an additional technical requirement is to combine data from various data sources and from different users (course assessment, lecturer assessment, student assessment, etc.). For some time, new methods for exploring the unique types of data that come from educational settings and their use to better understand learners and the settings is explored as a part of Educational Data Mining (EDM) field (Chalaris et al. 2014). The advancement in terms of data mining methods and tools makes it possible to analyse increased volumes of educational data in order to improve the quality of the educational processes (Asif et al., 2017). One of the most frequent uses of EDM is for examining students' (learning) behaviour in online learning environments (Siti Khadijah & Zaidatun, 2013).

In the paper, the focus in on the analysis of data collected within an e-learning system (Moodle) before and during the online assessment. The students were given the possibility to access the test as many times they wanted. Online testing has become a common way to organize formative assessment in higher education environments. The studies show that when student participation is stimulated by scoring formative tests held in an unproctored, online environment, issues of academic dishonesty occur (Arnold, 2016). Scoring is controversial in formative testing, as chasing the score may distract from deep learning (Wolsey, 2008 in Arnold, 2016). In that line, we hope to provide additional insight into these particular issues.

2 Analyses of student activities' logs in learning management systems

Online courses are delivered through e-learning platforms that allow users to create virtual learning environments. One of the basic functions of the system is to manage e-learning courses. In addition to that, the system enables the creation and presentation of different types of learning media, recording of user data, virtual communication among participants and so on (Ertl et al., 2007). Learning Management Systems (LMSs), often referred to as virtual learning environments, developed a great deal over the last twenty years (cf. McCormack & Jones, 1997). This resulted in the growth of e-learning implementation projects at universities, schools and business organisations (Mazza & Botturi, 2009). These institutions usually rely heavily on the whole range of tools that enable centralization and automation of different aspects of learning through the following functions, among others (Morrison, 2003): user registration, user profile management, e-course catalogue management, storing and delivering e-learning courses, integrating modules and tools required for e-learning, tracking and recording the progress of users, learning assessment, tracking and storing assessment results, and generating various types of reports to manage different processes. The functions are often grouped into ones that support (i) management of learning resources, (ii) communication and collaboration between students and instructors, (iii) assessment of learning, (iv) system support, and (v) access and role management (Coffey, 2007). LMSs either support (i) traditional courses (to a lesser extent, usually for online material delivery), (ii) hybrid approach, and/or (iii) learning that is fully online. The last approach uses the greatest number of features provided by the system.

Analysis of user data generated through interaction with resources in an LMS can be compared with web (data) mining, but with a special emphasis on learning and pedagogic information that is commonly not available in a standard web analytic approach or tools. Although more research has been conducted on this subject lately, there were a number challenged related to systematic approach to analysing a large number of logs generated by student activities and a common architecture (Scheuer et al., 2009). Increasing volumes of data about learning and teaching processes generated in different educational contexts (whether formal or informal, higher education or lifelong learning) led to advent of concepts such as learning analytics (LA) and EDM. Learning analytics is often defined as the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs (Siemens, 2011). The definition is adopted and promoted by Society for Learning Analytics Research (SoLAR).

LA and EDM develop rapidly due to advances in data management (Elias, 2011). In addition to modern day tools for data management, a large number of tools for (big) data analysis are available on the market.

In higher education, one of the key questions is how to increase student engagement and, in the process, achieve transformative learning outcomes. LA and EDM are crucial tools for answering such a complex question (Siemens & Baker, 2012). Approaches behind LA and EDM are very similar. The key difference is that LA relies on human interpretation of the data, data visualization and social network analysis, whereas EDM is based on automated machine learning i.e. data mining methods.

Drawing on number of success stories and research reports, we recently explored and assessed appropriateness of LA and EDM concepts and tools in one higher education institution (HEI) in Croatia and advocated an analysis that goes beyond basic reports provided within centrally managed academic information systems. As a proof of concept, a data mart that combines data from couple of data sources (Moodle LMS and custom academic information system) was presented as a basis for systemic, real-time analysis of educational data in HEIs (Maršić et al., 2016).

Here, we focus on the potential of log files that result from the automatic tracking of all interactions within a LMS. As stated earlier, this data can be used to analyse and evaluate learning activities with the purpose to improve the activities or the learning environment (Avouris et al., 2009). By using logs, it is possible to explore how the student's behaviour in the e-learning system affects their success, as studies show that more successful students spend more time and are more engaged in e-learning courses compared to less successful students (Campbell et al., 2006). Knowledge discovered from log data can be used by students, teachers, and system administrators (Romero & Ventura, 2007). For students, it is possible to recommend activities, teaching materials and assignments in a way that facilitates and improves the learning process. Recommendations can be made on the basis of the student's behaviour as recorded in the system and the behaviour of other similar students. Teachers can get a more objective feedback, evaluate the structure of teaching content/course and determine the effectiveness of the program. Furthermore, teachers are offered the ability to classify students into groups based on their needs for additional help and guidance, to explore behavioural patterns in the system, to look for the most common errors. At the same time, the administrators can monitor the parameters important to improve system performance (optimal server size, network traffic distribution, and so on).

The link between using the course material and the success achieved in the final test has been studied for a while now (e.g. Rafaeli & Ravid, 1997). The limitation of such studies is that they do not analyze the activities that students spend offline, due to the fact that a large number of online materials will be printed and used in a paper version whenever possible. It has been confirmed that time spent on assignments and frequency of participation is important for successful online learning (Morris et al., 2005); the good predictors of the final grade are the number of discussion posts posted, the number of visited pages with content, and the time spent in viewing the discussion pages. Differently, Ramos and Yudko (2008) confirm that that opening and reading of online course pages is a good predictor of success, but posting in discussions and its reading are not. To explore the case in detail, we have conducted a study presented hereinafter.

3 Research questions

In order to discover student behaviour patterns (based on the data recorded by the LMS) which could be related to their score required to pass the specific e-course, the research questions have been posed:

- Is there a correlation between accessing the content of the e-course and the results obtained in the final test of the e-course?
- Is there a correlation between the number of accesses to the final test and the results achieved in the final test of the e-course?
- Which of the two links is stronger when observing the result achieved during the first access to the test?
- Which of the previous two links is stronger when observing the best achieved result on the test?

These questions are set in the context of the online test where the possibility of multiple access is enabled. The answers to these questions will be based on the data analysis within and between the two generations of students who have accessed the e-course.

4 Methodology

4.1 Research setting

The LMS Moodle that has been used at the Faculty of Economics, University in Split since 2008 was the platform that delivered the course. Students enrolled in the first-year course “Information Technology” were able to access the e-course “Information Security” for 4 weeks. The objective of this e-course is to educate students about the concepts of information security and the measures of protection of information resources. The students were able to access the resources and the activities (reading text, watching video material, complete the surveys, etc.) in the sequence and dynamics that suited them. To successfully complete the e-course, the students were required to achieve 70% score on the final test but there was no limit on the maximum number of accesses to the test or the time between taking the test.

Figure 1. Test, video and infographic within e-course

The Information technology course is structured around various activities where continuous monitoring of student progress is employed through a model of accumulation of points. Students who successfully passed the e-course were awarded bonus points. The Quiz module which was used to develop the test is one of the most complex and most flexible parts of the Moodle system. For the test, we opted for randomly generated questions from the databank. The feedback is immediate. For the questions with more than one correct answer, the correct answers were scored as positive, and the incorrect as negative, so the sum of positive and negative points gives the final result on that question. The result could not be less than zero.

4.2 Participants

The participants of the study were the first-year students of the Faculty of Economics. In the academic year 2015/16, 271 students participated in the hybrid course Information technology. The age of the students was between 18 and 22 years, 72% were female and 29% male. In the academic year 2016/17, 269 students participated in the same hybrid course. The age of these students was between 18 and 22, 69% were female and 32% male. The respondents come from a relatively homogeneous group (first-year students) and share a similar background in terms of education, economic situation etc. The age and gender structure between the two generations of students is very similar.

4.3 Collecting and analysing data

The data about student behaviour in the e-learning course was collected from the Moodle system where detailed logs on students' activities are tracked. Data was collected from two groups of students who had access to the e-course “Information Security”. For the purposes of reporting on the results and student behaviour, there are two different modules in Moodle that allow easy downloading of textual or tabular files – the Grader report for the results of activities that are scored, and the Log that captures activity data for available resources within the system.

Selected data from the system were exported to the .xls file whereby pivoting the data, it was reduced to a format suitable for further analysis. From the first generation (15/16) over 168,000 records were collected. In 16/17, a new version of the Moodle was used so the number of collected records was over 299,000. The data collected from the Grader and the Log was merged into a flat file with the ID of the student and all the activities the student had done in the course. Using the IBM SPSS tool, descriptive and correlative analysis was performed for each generation and the differences between the generations were tested using t-test.

5 Results and discussion

The results are first presented for the 15/16 generation, and then for 16/17, followed by the comparison of the results for the two groups. Table 1 presents the descriptive statistics for the first result/attempt that students achieved when they submitted the online test (in 15/16). The average value is 74.94, above the 70-point threshold.

Table 1. Descriptive statistics for the first result/attempt on the test (generation 2015/16)

	N	Min.	Max.	Mean	St. Dev.
First result	263	6.67	100.00	74.94	15.10

Figures 2 and 3 shows data sorted by the number of points achieved on the test in their first attempt. Figure 2 shows a significant increase in the number of accesses to the test (log test) after the number of points falls below 70 (which is the threshold to pass the e-course), while Figure 3 does not indicate a significant increase in the activity of accessing the content of the e-course when the result falls below 70. This suggests that in the conditions of multiple-access to online tests, the students are more likely to opt for trial & error system than for reading the content.

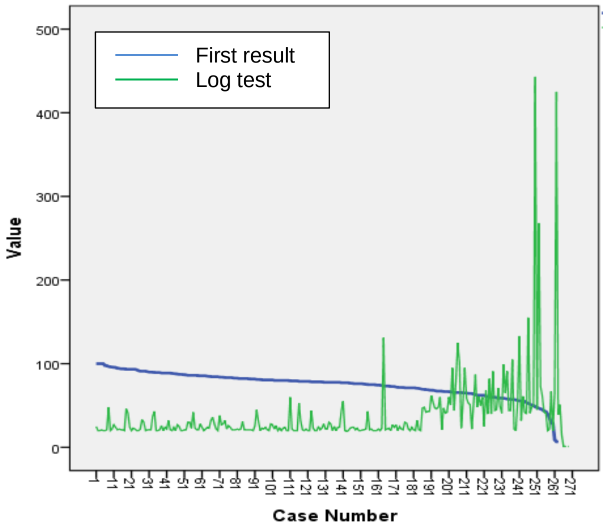

Figure 2. Points achieved on the first attempt and number of times accessing the test (gen. 2015/16)

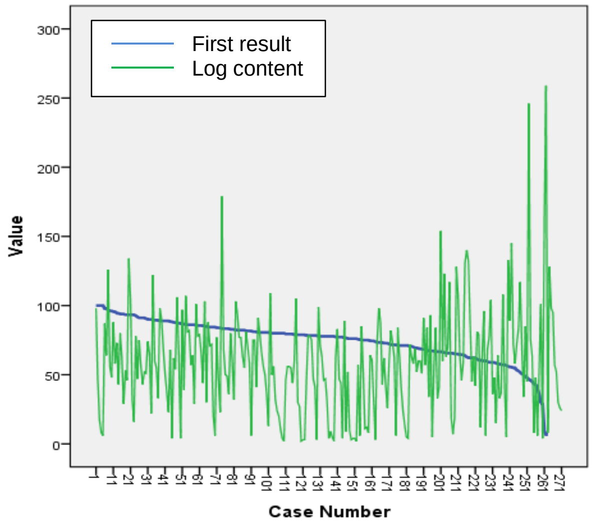

Figure 3. Points achieved on the first attempt and number of times accessing the content (gen. 2015/16)

Table 2 presents the correlation of the points achieved on the test (the first attempt) and the numbers of times accessing the test and the content. There is a negative and statistically significant correlation between the activity of accessing the test and the content and the results achieved on the first attempt.

Table 2. Correlation of the first test result and accessing the test and the content (gen. 2015/16)

		First result	Log test	Log content
First result	Pearson Corr.	1	-.485**	-.178**
	Sig. (2-tailed)		.000	.004
	N	263	263	263
Log test	Pearson Corr.	-.485**	1	.440**
	Sig. (2-tailed)	.000		.000
	N	263	268	268
Log content	Pearson Corr.	-.178**	.440**	1
	Sig. (2-tailed)	.004	.000	
	N	263	268	271

As can be seen in Fig 1 and 2, the observed negative link is stronger between multiple access to tests (log test) and the first result (-0.485**) than to accessing the content (log content) (-0.178**), meaning that students with a lower number of points (especially points below the 70-point threshold) after taking the test for the first time, re-take the test and access the content of the e-course more than students with better scores.

Table 3 presents the descriptive statistics for the best result/attempt that students achieved when they submitted the online test (in 15/16). The average value is 81.94, with standard deviation of 9.85.

Table 3. Descriptive statistics for the best result on the test (generation 2015/16)

	N	Min.	Max.	Mean	St. Dev.
Best result	264	28.89	100.00	81.95	9.85

Table 4 presents the correlation between the best test result and the total number of accesses to the test and content. The above-mentioned repeated access to the test and content is positively related to the best score that a student has achieved. A somewhat stronger link is between the best result and accessing the content (0.262**) compared to accessing the test (0.198**), both statistically significant.

Table 4. Correlation of the best test result and accessing the test and the content (gen. 2015/16)

		Best result	Log test	Log content
Best result	Pearson Corr.	1	.198**	.262**
	Sig. (2-tailed)		.001	.000
	N	264	264	264
Log test	Pearson Corr.	.198**	1	.440**
	Sig. (2-tailed)	.001		.000
	N	264	268	268
Log content	Pearson Corr.	.262**	.440**	1
	Sig. (2-tailed)	.000	.000	
	N	264	268	271

If the best results from the first attempt (presented in Table 5) are taken into account, the correlation with accessing the content is positive (0.253**), and similar to the correlation coefficient for multiple access (0.262**). A conclusion can be drawn - students who achieve a higher result in the first test attempt, access the contents of the e-course more frequently. Those students who achieve lower result than the threshold are more focused on subsequent test attempts than accessing content.

Table 5. Correlation of the best test result and accessing content (generation 2015/16) [If access to test =1]

		Best result	Log content
Best result	Pearson Corr.	1	.253**
	Sig. (2-tailed)		.001
	N	181	181

With regards to generation 2016/17, Table 6 presents the descriptive statistics for the first result/attempt that students achieved when they submitted the online test. The average result is 77.98, which is above the 70-point threshold. Compared to generation 2015/16, the average result is 3.04 higher. Smaller range of points is noted (min 32.22 - max 97.78) as well as standard deviation (12.68). The statistical significance of the differences in the points achieved as well as the potential causes will be analysed further in the paper when the results of the two generations are compared.

Table 6. Descriptive statistics for the first result/attempt on the test (generation 2016/17)

	N	Min.	Max.	Mean	Std. Dev.
First result	269	32.22	97.78	77.98	12.67

Figures 4 and 5 show data sorted by the number of points achieved on the test in their first attempt. Even more significant increase in the number of test attempts (log test) after the number of points falls below 70 (the threshold) is observed in Figure 4. Similar to a year before, there is no significant increase in the activity of accessing the content of the e-course when the first result falls below 70.

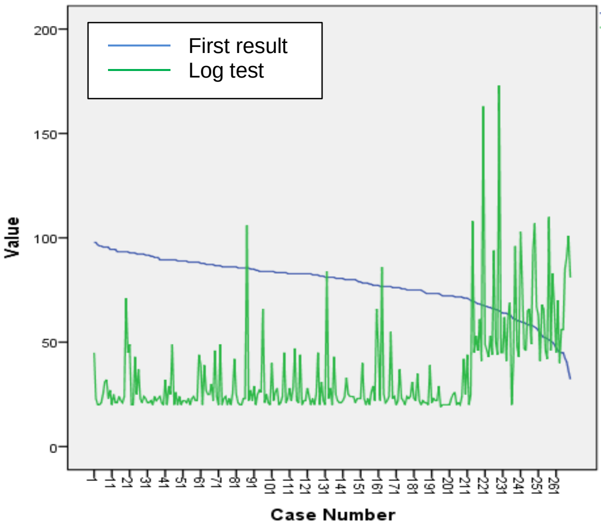

Figure 4. Points achieved on the first attempt and number of times accessing the test (gen. 2016/17)

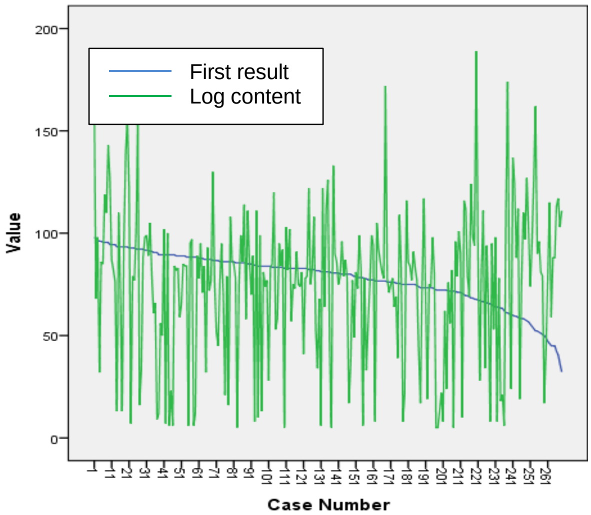

Figure 5. Points achieved on the first attempt and number of times accessing the content (gen. 2016/17)

Table 7 presents the correlations between the points achieved on the test (the first attempt) and the numbers of times accessing the test and the content (in 2016/17). As illustrated, there is a stronger statistically significant negative correlation between numbers of times accessing the test and the first result (-0.560**). Students with a lower number of points (especially points below the 70-point threshold) after the first attempt, re-take the test rather than access the content. This pattern of behaviour is even more obvious than in the previous generation.

Table 7. Correlation of the first test result and accessing the test and the content (gen. 2016/17)

		First result	Log test	Log content
First result	Pearson Corr.	1	-.560**	-.079
	Sig. (2-tailed)		.000	.197
	N	269	269	269
Log test	Pearson Corr.	-.560**	1	.360**
	Sig. (2-tailed)	.000		.000
	N	269	269	269
Log content	Pearson Corr.	-.079	.360**	1
	Sig. (2-tailed)	.197	.000	
	N	269	269	269

Table 8 presents the descriptive statistics for the best result/attempt that students achieved when they submitted the online test (in 16/17). The average value is 83.08, with standard deviation of 7.37. The average is slightly higher (1.13 percentage points) with lower standard deviation compared to year before.

Table 8. Descriptive statistics for the best result on the test (generation 2016/17)

	N	Min.	Max.	Mean	St. Dev.
Best result	269	56.11	97.78	83.08	7.37

Table 9 presents the correlation between the best test result and the total number of accessing the test and the content. Accessing the content is positively correlated to the best score that a student has achieved (0.241**). The link between accessing the test and the best test result is not statistically significant.

Table 9. Correlation of the best test result and accessing the test and the content (gen. 2016/17)

		Best result	Log test	Log content
Best result	Pearson Corr.	1	.010	.241**
	Sig. (2-tailed)		.866	.000
	N	269	269	269
Log test	Pearson Corr.	.010	1	.360**
	Sig. (2-tailed)	.866		.000
	N	269	269	269
Log content	Pearson Corr.	.241**	.360**	1
	Sig. (2-tailed)	.000	.000	
	N	269	269	269

If the best results from the first attempt (presented in Table 10) are taken into account, the correlation with accessing the content is positive (0.223**), and similar to the correlation coefficient for multiple access (0.241**).

Table 10. Correlation of the best test result and accessing content (generation 2016/17) [If access to test =1]

		Best result	Log content
Best result	Pearson Corr.	1	.223**
	Sig. (2-tailed)		.001
	N	201	201

Table 11 presents the correlations between the test results (the first and the best) and the access to different types of content. Based on this insight into the structure of the content that students have accessed, it can be stated that the best results are particularly positively linked with access to different e-course content. In contrast, the link with the first test result is not confirmed. What is more, the students with a lower number of points focus their activities on multiple test attempts.

Table 11. Correlation of the first and the best test result and accessing different types of content (generation 2016/17)

		Best result	First result
File	Pearson Corr.	.188**	-.144*
	Sig. (2-tailed)	.006	.034
	N	216	216
Choice	Pearson Corr.	.252**	.117
	Sig. (2-tailed)	.000	.085
	N	219	219
Glossary	Pearson Corr.	.009	.042
	Sig. (2-tailed)	.906	.576
	N	176	176
Page	Pearson Corr.	.281**	-.100
	Sig. (2-tailed)	.000	.120
	N	241	241
System	Pearson Corr.	.207**	-.098
	Sig. (2-tailed)	.001	.107
	N	269	269

Table 12 presents the differences between the presented results for the two generations. Overall, the students in 2016/17 accessed the e-course more often, achieved better results (both for the best and the first attempt) and accessed the content of the e-course to a greater extent. At the same time, for the same generation, fewer test attempts are noted.

Table 12. Comparison of the results for students from 2015/16 and 2016/17 generation

	Gen.	N	Mean	Std. Dev.
Total e-course access	2015/16	271	95.29	69.44
	2016/17	269	109.54	50.34
Best result	2015/16	264	81.95	9.85
	2016/17	269	83.08	7.37
First result	2015/16	263	74.95	15.10
	2016/17	269	77.98	12.68
Log test	2015/16	268	37.53	43.2
	2016/17	269	35.04	23.01
Log content	2015/16	271	58.18	38.58
	2016/17	269	74.51	37.26
Number of test attempts	2015/16	263	1.62	1.96
	2016/17	269	1.44	0.95

The differences in total e-course access, the first and the best test result and access to e-course content between the two generations are statistically significant, as presented in Table 13.

Table 13. T-test for equality of means

		t	df	Sig. (2-tailed)	Mean Difference
Total e-course access	Equal variances assumed	-2.73	538	0.007	-14.26
	Equal variances not assumed	-2.73	492.48	0.007	-14.26
Best result	Equal variances assumed	-1.51	531	0.132	-1.13
	Equal variances not assumed	-1.50	487.12	0.133	-1.13
First result	Equal variances assumed	-2.51	530	0.012	-3.03
	Equal variances not assumed	-2.50	510.47	0.013	-3.03
Log test	Equal variances assumed	.835	535	0.404	2.49
	Equal variances not assumed	.834	406.92	0.405	2.49
Log content	Equal variances assumed	-5.00	538	0.000	-16.33
	Equal variances not assumed	-5.00	537.59	0.000	-16.33
Number of test attempts	Equal variances assumed	1.39	530	0.165	0.18
	Equal variances not assumed	1.38	376.33	0.168	.1849

6 Conclusion

The changes in higher education are characterized by the increased expectations of the various implementations of ICT in educational processes. LMSs facilitate the development and management of e-learning courses as well as monitoring of student behaviour. The importance and the implications of studies focusing on student behaviour in VLE are presented in the first part of the paper (section 2). This study focused on the analysis of student activities' logs generated within one specific (integral) part of the hybrid course Information technology delivered fully online.

By using log data from the institutional LMS for two generations of students who have accessed the e-course Information security, answers to the research questions (listed in section 3) have been presented (in sections 4 and 5). The results point to correlation between certain student activities in the e-course and their test results. The results of both studies indicate that students who visited content pages more frequently achieved better results on the test. Also, it turned out that those students who achieve scores lower than the threshold, largely direct their activity to re-attempting the test instead of reading i.e. learning the content of the e-course. These results correspond with the research of Morris et al. (2005) who found that more successful students associate their online activities to what they believe is essential to achieve the passing grade. Though, in their study, the activities of successful students were, in addition to frequent visits to content pages, the participation in and following of the online discussions. Likewise, the research conducted by Macfadyen and Dawson (2010) implies that student's online activities such as using forums, sending emails, and online (self-)assessment are significant predictors of the final grade in the e-learning course.

A positive link between accessing the content and the achieved results is confirmed by testing the significance of the differences between the two generations - the students from the 2016/17 generation who accessed the e-course and the e-learning content more frequently achieved a better result while simultaneously taking the test fewer times.

When we analysed the student behaviour in the LMS focusing on the online assessment, it became apparent that students in the conditions of multiple-attempts allowed, decide to access the test more times based on the trial & error system instead of learning the content itself. Other authors have raised caution about the problem of cheating while testing students in the online environment since the focus apparently shifts from deep learning to passing the score threshold (Arnold, 2016; Wolsey, 2008). This research, looking at the results within and between generations, shows that the best results achieved in the test are still linked to learning the content rather than guessing the answers. However, students who do not achieve the threshold in the first attempt resort to guessing the questions in subsequent attempts. This issue can be resolved in the e-learning system by introducing a time delay between the two tests (this is planned in 2017/18). Notable positive outcomes of self-assessment tests with one-hour time delay within the same hybrid course (Information technology) within and between generations are presented in our earlier paper (Ćukušić et al., 2014).

Acknowledgments

This work has been partly supported by Croatian Science Foundation under the project HigherDecision IP-2014-09-7854.

References

Arnold, I.J.M. (2016): Cheating at online formative tests: Does it pay off? *Internet and Higher Education*, 29, 98–106.

Asif, R., Merceron, A., Ali, S.A., Haider, N.G. (2017): Analyzing undergraduate students' performance using educational data mining, *Computers & Education*, In press.

Avouris, N., Fiotakis, G., Kahrimanis, G., Margaritis, M., Komis, V. (2009): Beyond Logging of Fingertip Actions: Analysis of Collaborative Learning Using Multiple Sources of Data In Choquet, C., Luengo, V., Yacef, K. (Eds.): *Usage Analysis in Learning Systems*, Association for the Advancement of Computing in Education.

Burgos, C. et al. (2017), Data mining for modeling students' performance: A tutoring action plan to pre-vent academic dropout, *Computers and Electrical Engineering*, In press.

Campbell, J. P., Finnegan, C., Collins, B. (2006). Academic analytics: *Using the CMS as an early warning system*. Paper presented at the WebCT impact conference.

Chalaris, M., Gritzalis, S., Maragoudakis, M., Sgouropoulou, C., Tsolakidis, A. (2014): *Procedia - Social and Behavioral Sciences,*147, 390 – 397.

Coffey, J.W. (2007): Integrating Visual Representations of Knowledge with Learning Management Systems: Design Principles for Advanced Computer-Based Learning Support. In: Neto, F.M.M., Brasileiro, F.V. (Eds): *Advances in computer-supported learning*, Information Science Publishing (an imprint of Idea Group Inc.).

Ćukušić, M, Garača, Ž., Jadrić, M. (2014): Online self-assessment and students' success in higher education institutions, *Computers & Education*, 72, pp. 100-109, DOI:10.1016/j.compedu.2013.10.018.

Elias, T., (2011): Learning Analytics: Definitions, Processes and Potential. Retrieved from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

Ertl, B., Winkler, K., Mandl, H. (2007): E-Learning: Trends and Future Development. In: Neto, F.M.M., Brasileiro, F.V. (Eds): *Advances in computer-supported learning*, Information Science Publishing (an imprint of Idea Group Inc.).

Macfadyen, L.P., Dawson, S. (2010): Mining LMS data to develop an ''early warning system" for educators: A proof of concept, *Computers & Education*, 54, 588–599.

Maršić, A., Ćukušić, M., Jadrić, M. (2016): *Evaluating the potential of Learning Analytics and Educational Data Mining concepts and tools in higher education in Croatia*, Presentation at 16th International Conference on Operational Research KOI 2016, September 27-29, 2016.

Mazza, R., Botturi, L. (2009): Monitoring an Online Course with the GISMO Tool: A Case Study. In Choquet, C., Luengo, V., Yacef, K. (Eds): *Usage Analysis in Learning Systems*, Association for the Advancement of Computing in Education.

McCormack, C., Jones, D. (1997): Building a web-based education system. In Mazza, R., Botturi, L.: Monitoring an Online Course with the GISMO Tool: A Case Study in Choquet, C., Luengo, V., Yacef, K. (Eds.): *Usage Analysis in Learning Systems*, Association for the Advancement of Computing in Education.

Mimirinis, M., Bhattacharya, M. (2009): Design of Virtual Learning Environments for Deep Learning In Bhattacharya, M., Kommers, P. (Eds): *The Connected Learning Space*, AACE.

Morris, L.V., Finnegan, C., Wu S.-S. (2005): Tracking student behavior, persistence, and achievement in online courses, *Internet and Higher Education*, 8, pp. 221–231.

Morrison, D. (2003): *E-learning Strategies: How to Get Implementation and Delivery Right First Time*, John Wiley & Sons.

Murphy, C.A., Stewart, J.C. (2017): On-campus students taking online courses: Factors associated with unsuccessful course completion, *Internet and Higher Education*, 34, pp. 1–9.

Rafaeli, S., Ravid, G. (1997): Online, web-based learning environment for an information systems course: Access logs, linearity and performance, Retrieved from: http://www.ravid.org/gilad/isecon1997.pdf

Ramos, C., Yudko, E (2008): "Hits" (not "Discussion Posts") predict student success in online courses: A double cross-validation study, *Computers & Education*, 50, pp. 1174–1182.

Romero, C., Ventura S., Garcia, E. (2008): Data mining in course management systems: Moodle case study and tutorial, *Computers & Education*, 51, pp. 368-384.

Scheuer, O., Mühlenbrock, M., Melis, E. (2009): Results from Action Analysis in an Interactive Learning Environment, In Choquet, C., Luengo, V., Yacef, K. (Eds): Usage Analysis in Learning Systems, Association for the Advancement of Computing in Education.

Siemens, G, Baker R (2012): Learning analytics and educational data mining: towards communication and collaboration. In: *Proceedings of the 2nd international conference on learning analytics and knowledge*, pp. 252–254.

Siemens, G. (2011) Learning and Academic Analytics. Blog post. Retrieved from: http://www.learninganalytics.net/?p=131

Siti, Khadijah, M., Zaidatun, T. (2013): Educational data mining: A review, *Procedia - Social and Behavioral Sciences*, 97, pp. 320 – 324.