Development of a methodological framework for strategic decision-making in higher education – a case of open and distance learning implementation

IP-2014-09-7854

Analysing students' behaviour patterns in online assessment

Mario Jadrić, Maja Ćukušić, Željko Garača

Our aim was to discover:

1. students' behavioural patterns during a formative online assessment test, and
2. links to their results achieved in a specific e-learning course.

The focus of the paper is on the analysis of data collected within an e-learning system (Moodle) before and during the online assessment.

Motivation

- Growing interest of researchers to automatically analyse data generated by students in an online education environment due to:

  - 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,

  - potential to build prediction models that can calculate probabilities of students' behaviour, all in order to provide timely support to students.

LA & EDM

- 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 Educational Data Mining (EDM).

- In HE 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).

LA vs EDM

- 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.

- One of the most frequent uses of EDM is for examining students' (learning) behaviour in online learning environments. We focus on the potential of log files that result from the automatic tracking of all interactions within a LMS.

The potential uses of log data

- Knowledge discovered from log data can be used by students, teachers, and sys administrators (Romero & Ventura, 2007).

  – 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.

  – Administrators: can monitor the parameters important to improve system performance (optimal server size, network traffic distribution, and so on).

higher DECISION

Issues in formative online assessment

- Scoring is controversial in formative testing, as chasing the score may distract from deep learning (Wolsey, 2008 in Arnold, 2016).

- 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).

Research questions

1. Is there a correlation between accessing the content of the e-course and the results obtained in the final test of the e-course?

2. 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?

3. Which of the two links is stronger when observing the result achieved during the first access to the test?

4. Which of the previous two links is stronger when observing the best achieved result on the test?

Research setting

- Faculty of Economics, University in Split.

- Moodle used since 2008.

- First-year course "Information Technology"

- Within the course, an e-course "Information Security" is available for 4 weeks.

E-course Information security

- The objective of this e-course is to educate students about the concepts of information security and the measures of protection of information resources.

- Accessing the resources and the activities (reading text, watching video material, complete the surveys, etc.) in the sequence and dynamics that suits students.

- 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.

- Students who successfully passed the e-course were awarded bonus points.

Participants & Data

- First-year students of the Faculty of Economics in Split.
- 2015/16 - 271 students
  - between 18 and 22 years, 72% were female and 29% male.
- 2016/17 - 269 students
  - between 18 and 22, 69% were female and 32% male.
- Data from Grader and Log modules of the Moodle system.
- More than 450.000 data records.

higher**DECISION**

Disclaimer

- To inspect the link between the variables, correlation coefficients were calculated. Since the sample was large (>200), Pearsons's correlation coefficient was calculated and presented in the paper.

- However, as data is not normally distributed, it would be more appropriate to use Spearman's rank correlation coefficient as a nonparametric (distribution-free) rank statistic.

- The correlations are still significant, but are of different intensity.

- Therefore, the tables presented hereinafter differ slightly from the ones presented in the paper available from conference website http://www.ceciis.foi.hr/.

Results 2015/16

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

	N	Min.	Max.	Mean	St. Dev.
Best result	264	28.89	100.00	**81.95**	9.85
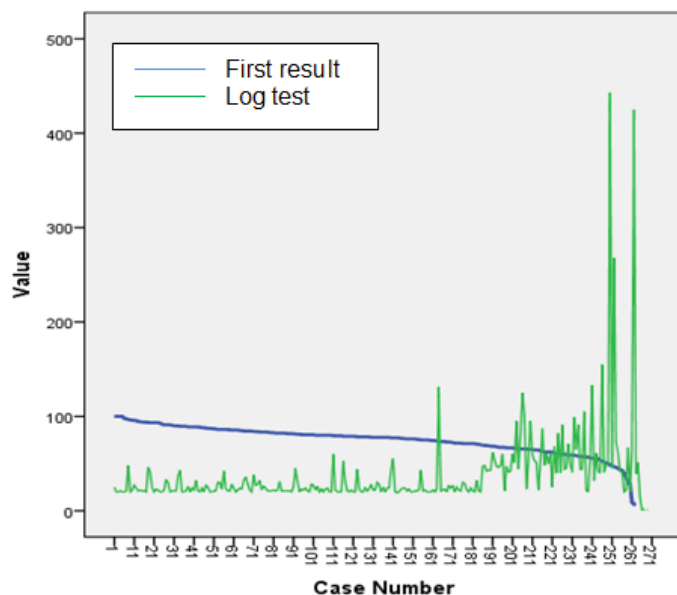
Points achieved on the first attempt and number of times accessing the test (gen. 2015/16)
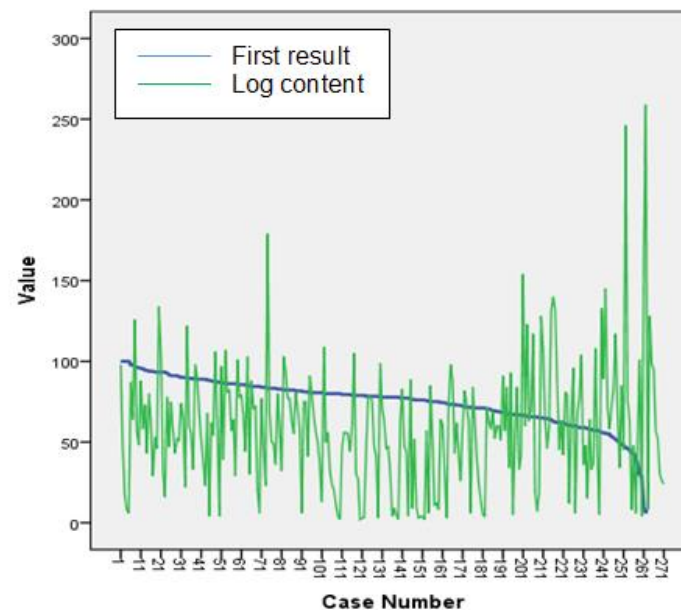
Points achieved on the first attempt and number of times accessing the content (gen. 2015/16)

higher**DECISION**

Results 2015/16

- Negative and statistically significant correlation between the activity of accessing the test and the results achieved on the first attempt.

- Repeated access to the content is positively related to the best score that a student has achieved.

- Students who achieve a higher result in the first test attempt, access the contents of the e-course more frequently.

		First result	Log test	Log content
First result	Spearman's rho Corr. Coef.	1.000	-0.509**	-0.035
	Sig. (2-tailed)	.	0.000	0.568
	N	263	263	263
Log test	Spearman's rho Corr. Coef.	-0.509**	1.000	0.339**
	Sig. (2-tailed)	0.000	.	0.000
	N	263	268	268
Log content	Spearman's rho Corr. Coef.	-0.035	0.339**	1.000
	Sig. (2-tailed)	0.568	0.000	.
	N	263	268	271

		Best result	Log test	Log content
Best result	Spearman's rho Corr. Coef.	1.000	0.118	0.275**
	Sig. (2-tailed)	.	0.055	0.000
	N	264	264	264
Log test	Spearman's rho Corr. Coef.	0.118	1.000	0.339**
	Sig. (2-tailed)	0.055	.	0.000
	N	264	268	268
Log content	Spearman's rho Corr. Coef.	0.275**	0.339**	1.000
	Sig. (2-tailed)	0.000	0.000	.
	N	264	268	271

higher**DECISION**

Results 2016/17

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

	N	Min.	Max.	Mean	St. Dev.
Best result	269	56.11	97.78	**83.08**	7.37
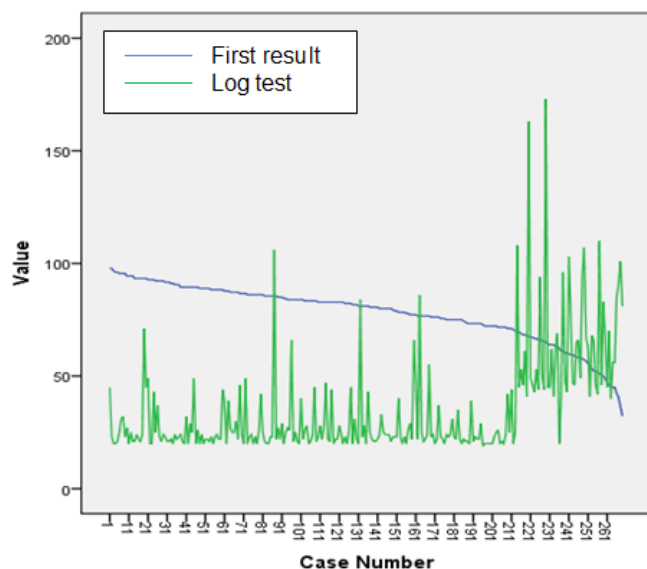
Points achieved on the first attempt and number of times accessing the test (gen. 2016/17)
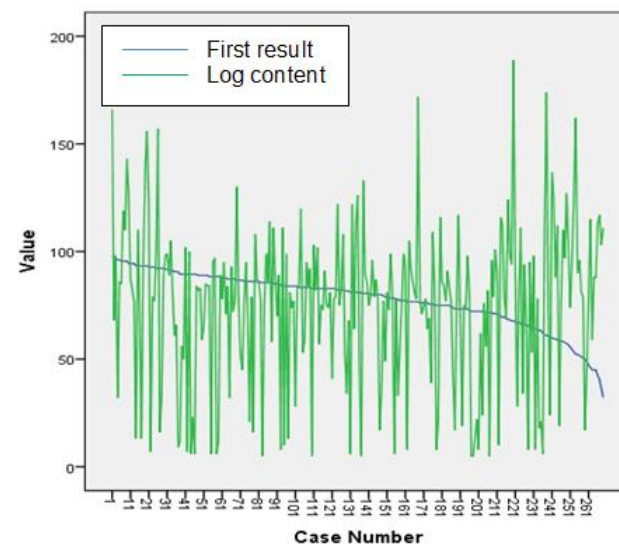
Points achieved on the first attempt and number of times accessing the content (gen. 2016/17)

HRZZ
Hrvatska zaklada za znanost

foi

higher DECISION

Results 2016/17

- Statistically significant negative correlation between numbers of times accessing the test and the first result.

- Accessing the content is positively correlated to the best score that a student has achieved. The link between accessing the test and the best test result is not statistically significant.

- If the best results from the first attempt are taken into account, the correlation with accessing the content is positive, and similar to the correlation coefficient for multiple access.

		First result	Log test	Log content
First result	Spearman's rho Corr. Coef.	1.000	-0.416**	-0.040
	Sig. (2-tailed)	.	0.000	0.517
	N	269	269	269
Log test	Spearman's rho Corr. Coef.	-0.416**	1.000	0.411**
	Sig. (2-tailed)	0.000	.	0.000
	N	269	269	269
Log content	Spearman's rho Corr. Coef.	-0.040	0.411**	1.000
	Sig. (2-tailed)	0.517	0.000	.
	N	269	269	269

		Best result	Log test	Log content
Best result	Spearman's rho Corr. Coef.	1.000	0.018	0.202**
	Sig. (2-tailed)	.	0.764	0.001
	N	269	269	269
Log test	Spearman's rho Corr. Coef.	0.018	1.000	0.411**
	Sig. (2-tailed)	.764	.	.000
	N	269	269	269
Log content	Spearman's rho Corr. Coef.	0.202**	0.411**	1.000
	Sig. (2-tailed)	0.001	0.000	.
	N	269	269	269

		Best result	Log content
Best result	Spearman's rho Corr. Coef.	1.000	0.199**
	Sig. (2-tailed)	.	0.005
	N	201	201

HRZZ Hrvatska zaklada za znanost
foi

Results 2016/17

- 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.

		Best result	First result
File	Spearman's rho Corr. Coef.	0.134^{*}	-0.099
	Sig. (2-tailed)	0.050	0.145
	N	216	216
Choice	Spearman's rho Corr. Coef.	0.253^{**}	0.139^{*}
	Sig. (2-tailed)	0.000	0.041
	N	219	219
Glossary	Spearman's rho Corr. Coef.	0.084	0.079
	Sig. (2-tailed)	0.266	0.299
	N	176	176
Page	Spearman's rho Corr. Coef.	0.250^{**}	-0.007
	Sig. (2-tailed)	0.000	0.914
	N	241	241
System	Spearman's rho Corr. Coef.	0.159^{**}	-0.027
	Sig. (2-tailed)	0.009	0.659
	N	269	269

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

HRZZ
Hrvatska zaklada za znanost

foi

higher**DECISION**

Comparison of the results for students: 2015/16 and 2016/17 generation

- 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.

- 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.

	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

		t	df	Sig. (2-tailed)	Mean Diff.
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

HRZZ Hrvatska zaklada za znanost   foi

higher DECISION

Conclusions

- The results of both studies indicate that students who visited content pages more frequently achieved better results on the test.

- The students who achieve scores lower than the threshold, 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.

Conclusions

- 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.

Conclusions

- 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).