

## Student Activity Analytics in an e-Learning Platform: Anticipating Potential Failing Students

Bertil P. Marques <sup>1\*</sup>, Jaime E. Villate <sup>2</sup>, Carlos Vaz Carvalho <sup>1</sup>

<sup>1</sup> GILT/ISEP, Porto, PORTUGAL

<sup>2</sup> FEUP, Porto, PORTUGAL

\*Corresponding Author: [bpm@isep.ipp.pt](mailto:bpm@isep.ipp.pt)

**Citation:** Marques, B. P., Villate, J. E. and Carvalho, C. V. (2018). Student Activity Analytics in an e-Learning Platform: Anticipating Potential Failing Students. *Journal of Information Systems Engineering & Management*, 3(2), 12. <https://doi.org/10.20897/jisem.201812>

**Published:** April 06, 2018

### ABSTRACT

The evolution of learning technology and tools changed the way students access information and build their knowledge. Registering the interaction of students with these tools generates a large amount of data that, once critically analysed, can provide important clues about the students' learning progress. Nevertheless research has still to be conducted to fully understand how (and if) the students' interaction with the learning technologies relates to their learning success. In parallel, new analytical tools must be developed to allow teachers to fully exploit the information embedded in this data, in a friendly but flexible way. This article is a contribution to this effort and presents a study where the use of a learning management system (LMS) in a specific semester-long course in an Engineering school produced data that was analysed and correlated to the students' success. Results indicate that some correlation exists between the effective use of some of the tools integrated in the LMS and student success which points the way to build specific applications to provide teachers with indicators of students in danger of failing.

**Keywords:** MOODLE, academic portal, ICT, higher education

### INTRODUCTION

For some time now, a multitude of technology-based learning tools have been used to support online and offline courses. From these tools, Learning Management Systems (LMS) are probably the most used platforms, especially in Higher Education. These applications provide teachers students with an extensive range of information and communication tools, depending on the structure defined by the teacher (Mota et al., 2014). Furthermore, access to the LMS is ubiquitous, in time and location, which necessarily changes the way students approach the learning process.

These platforms record and store all the user activity, from entry to exit, like the number of accesses, duration of accesses, paths traversed in the platform, tools used, resources used or downloaded, access to files and folders, performed tasks and activities, messages and posts read and sent, quizzes attempted and answered, assignments submitted, etc (Marques et al., 2010) (Preidys and Sakalauskas, 2010) (Mostow and Beck, 2006).

Teachers have access to this data but the sheer size of the collected information, the lack of synthetic views over this data and the inability to apply adequate techniques and tools to mine this data usually drives teachers away from making an effective use of it. Furthermore, data is normally obtained from three different sources: (1) recorded text, (2) web server log files, and (3) learning software log files (Black et al., 2008) and as such it is not stored in a systematic way so its thorough analysis requires long and tedious preprocessing (Kruger et al., 2010).

Therefore, it is mostly researchers that apply concepts of Data Analysis, Big Data and Learning Analytics to exploit the data (Alves et al., 2015), (Black et al., 2008), (Garcia and Secada, 2013), (Kruger et al., 2010), (Lino et al., 2017) as teachers do not have the required knowledge to apply these techniques and they don't have friendly tools to do that analysis and provide them with the processed information. These techniques have been applied to the assessment of student's performance, to support course adaptation, to scaffold recommender systems, detect atypical student's behavior and even to detect learning styles of students (Romero et al., 2007) (Liyanae et al., 2015) (Khribi et al., 2009).

For instance, Alves studied the access to virtual learning environments (VLEs) and reported on the large quantities of data resulting from both students and teachers' activities developed in those environments (Alves et al., 2015). Black used e-learning tools to generate relevant information, for the teacher and the students, to optimize their learning process. The study combines data-processing and learning analytical to improve higher education learning processes. Authors concluded that activity logs of virtual learning environments provided real knowledge of the use of these environments, but also identified the need for new pedagogical approaches (Black et al., 2008) to exploit that data.

The AAT tool was created to "...access and analyse student behaviour data in learning systems by enabling users [learning designers and/or teachers] to extract detailed information about how students interact with and learn from online courses in a learning system" (Graf et al., 2011). The Moodle Data Mining (MDM) tool addressed the knowledge discovery process from student data registered in Moodle courses (Luna et al., 2017). CourseVis is a visualization tool that shows a graphical representations of the students access data (Mazza and Dimitrova, 2007). GISMO complements the previous tool by giving information on the actual use of contents and activities (Mazza and Milani, 2004). Mostow et al. (2005) created a tool to represent the teacher-student interaction based on their communication logs. MATEP gets data from the LMS log files but also from the academic portal to generate dynamic reports (Zorrilla and Álvarez, 2008).

Some authors have used content analysis methods to study the interaction in discussion forums (Lin et al., 2009), (Dringus and Ellis, 2005). Using text analysis methods (a process that uses algorithms capable of analyzing collections of text documents in order to extract knowledge), authors found online discussion types. The results of this experiment helped teachers monitor online activities that took place in the discussion forums. Other authors have created software agents based on mathematical methods and statistical analysis to perform that data analysis (Castro et al., 2007), (Mamčenko and Šileikienė, 2006) and (Preidys and Sakalauskas, 2010).

In this paper, we present the process of data analysis of the behavior of students in a LMS to measure and contextualize their access (where, how many times, and in what form), their digital paths in the platform (what tools are used, actions or queries, use of resources, forum participation, etc.) and the correlation with student learning success. The intention was to determine significant correlations that leads to the creation of a tool that allows teachers to make an early identification of students with problems. This work was developed in the context of a PhD project whose objective was to study, discuss, propose and validate a support model for the adoption of Information and Communication Technologies (ICT) for pedagogical purposes in the HE scenario, and to propose a coherent and consistent model of institutional and pedagogical activity, centered on teachers (Marques, 2015). This article is an improved, revised and extended version of the publication by Marques (2017).

## STUDENT INTERACTION WITH LEARNING TOOLS VS. STUDENT SUCCESS

The LMS MOODLE collects a set of data that, when analyzed, can give teachers indicators to follow the student's behavior in order to identify critical situations and preventing dropouts. For this study, data was collected during a period of 6 months (from September 1, 2014 to February 29, 2015) from the course on Computer Principles (PRCMP) of the BSc in Computer Engineering (LEI) of the Department of Informatics Engineering (DEI). The population consisted of 364 individuals.

This study followed the proposal of Gaudioso and Talavera (2006) in the sense that the work started from a question derived from intuition based on the own experience and data was collected to confirm it. In this case, the hypothesis was that there was a correlation between the level of involvement of the students with the learning tools and their final learning success.

**Figure 1** shows the histogram of the final grades obtained by the students. Approximately 3% of the students failed the course because they did not obtain the minimum grade required. Nevertheless it is possible to see that the majority of students (85%) achieved a grade of 10 or higher which means they were successful. Analysing the access data is important to prevent the 15% of failing students, particularly the 3% of students that dropped out.

For the collection of data related to the use of the MOODLE LMS and its tools it was necessary to combine the internal functionality of the platform that allows to generate some usage reports with the creation of a set of applications that extracted directly the information from the MOODLE databases. This was a time-consuming

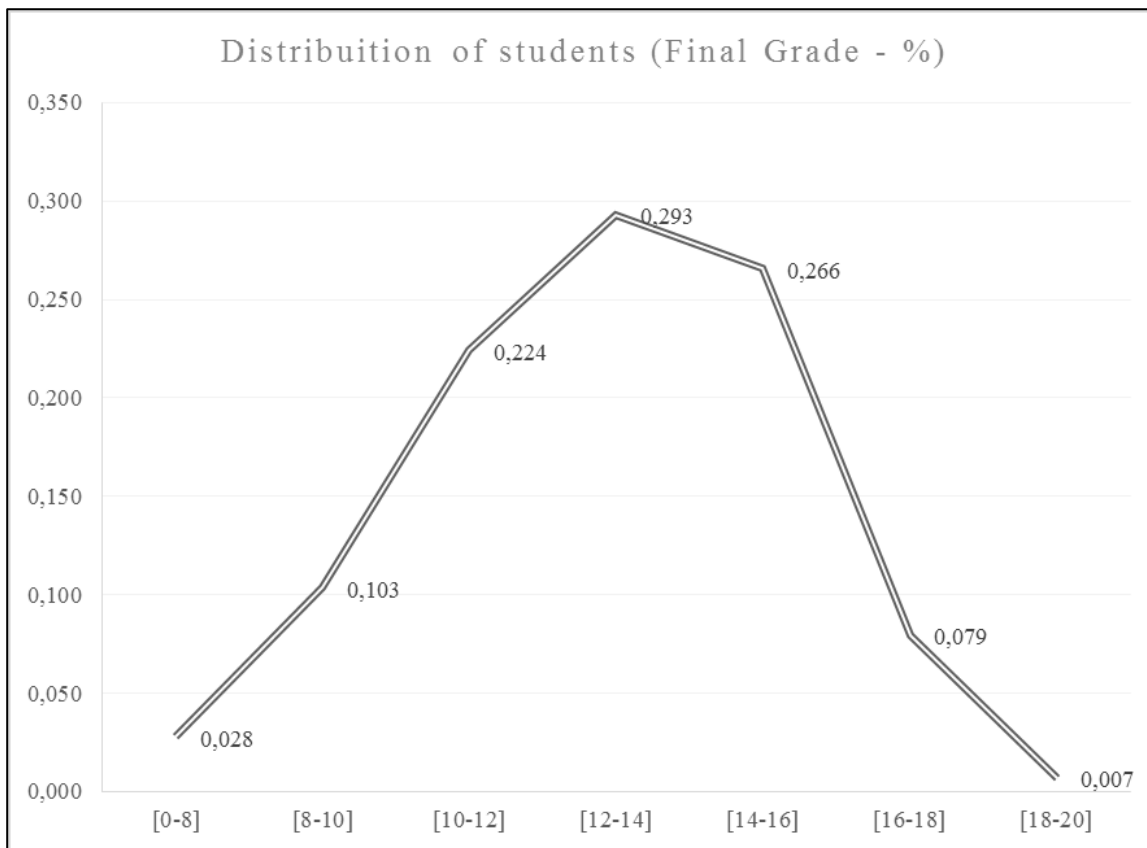


Figure 1. Distribution of students according to the final grade of the course

Table 1. Accesses by Students (Working days and Weekends)

Access per hour	Total students per hour (Workdays)	Total students per hour (Weekend)	Average accesses per student (Workdays)	Average accesses per student (Weekend)
0-1	208	80	2,95	1,63
1-2	129	56	2,20	1,71
2-3	69	32	1,77	1,41
3-4	43	22	1,72	1,41
4-5	21	5	1,62	1,20
5-6	35	6	1,86	1,33
6-7	86	2	1,88	1,00
7-8	299	17	5,02	1,35
8-9	325	54	10,56	1,67
9-10	329	119	9,07	2,27
10-11	338	149	11,17	2,62
11-12	333	181	9,55	2,55
12-13	325	199	6,63	2,34
13-14	324	218	8,08	2,38
14-15	337	251	9,19	2,83
15-16	340	272	9,97	3,54
16-17	335	273	9,21	3,07
17-18	338	266	8,12	3,05
18-19	341	249	7,54	2,74
19-20	328	219	6,47	2,36
20-21	331	206	6,29	2,42
21-22	329	218	6,42	2,25
22-23	317	206	5,83	2,12
23-24	281	170	4,37	2,16

and specialized programming work that is clearly out of reach for most of the teachers, even at the Higher Education level.

Data related to the distribution of accesses per hour and day was collected to characterize the students' use of the platform (Table 1).

This table is better viewed in the two following figures. In Figure 2, the average number of accesses per student is shown in hourly intervals, at workdays and at weekends. There are more accesses on working days than on the

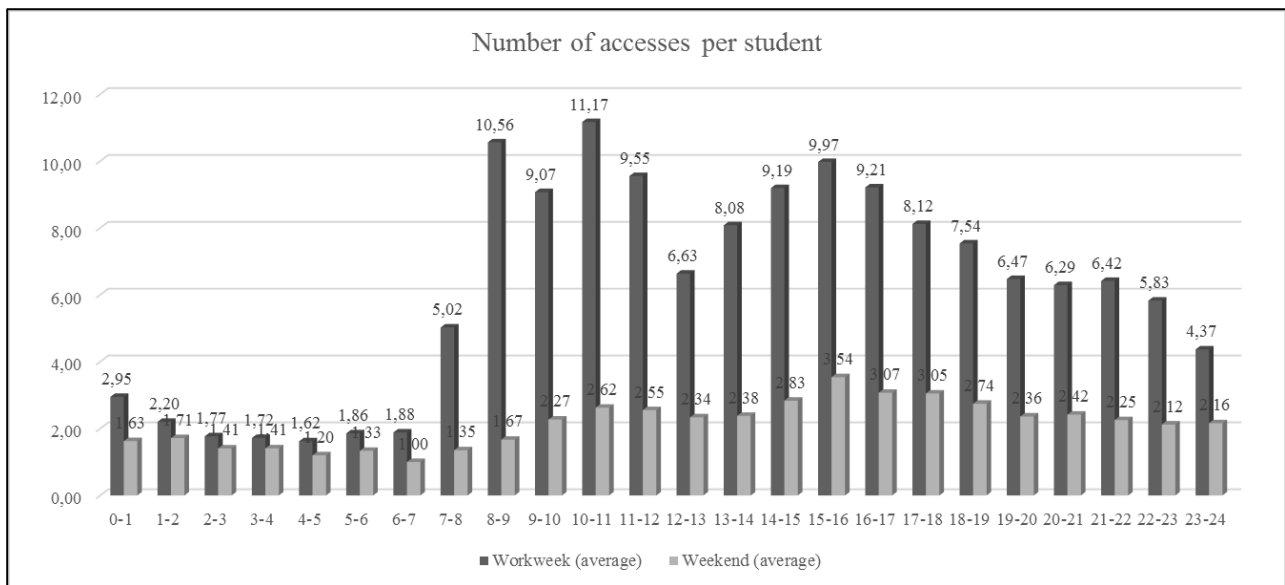


Figure 2. Average number of accesses per student

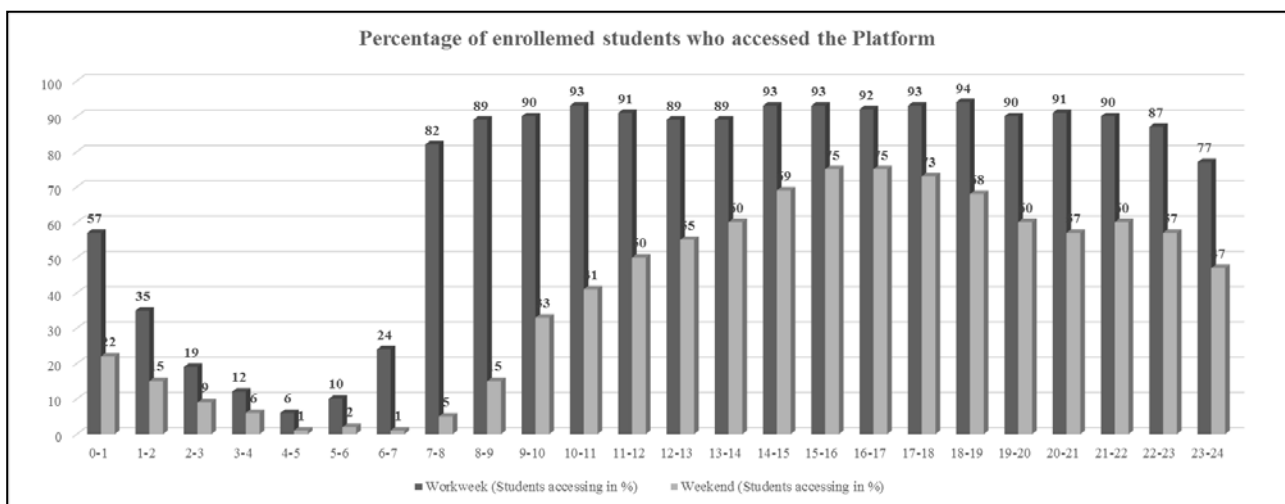


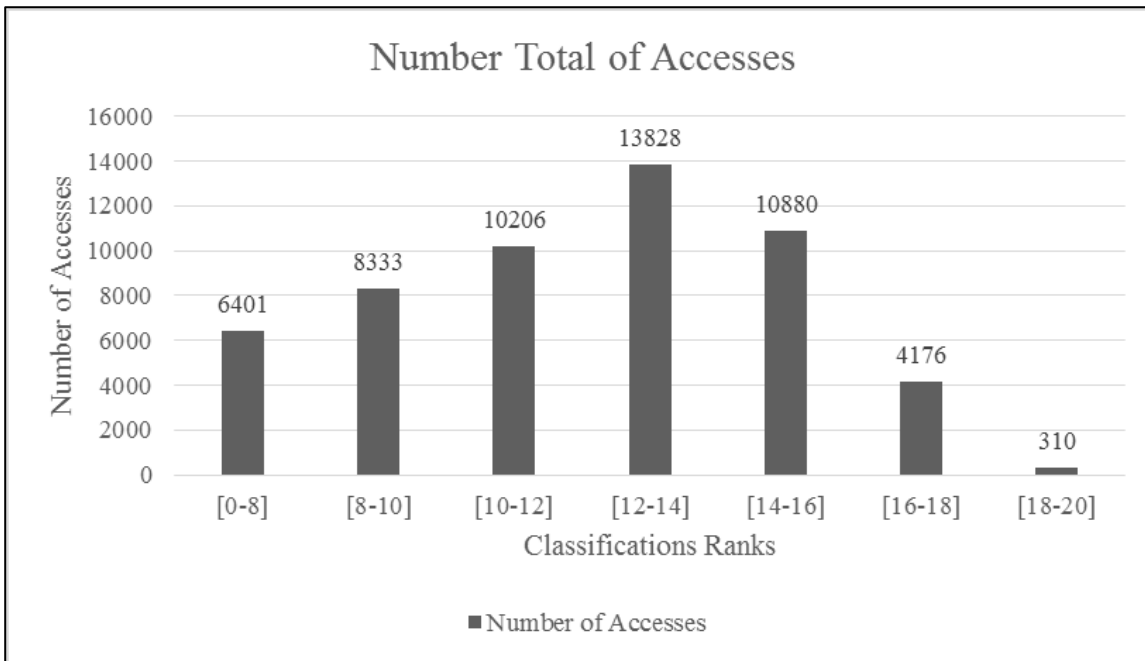
Figure 3. Percentage of enrolled students accessing the LMS in a certain hour

Table 2. Access data and number of students per grade range

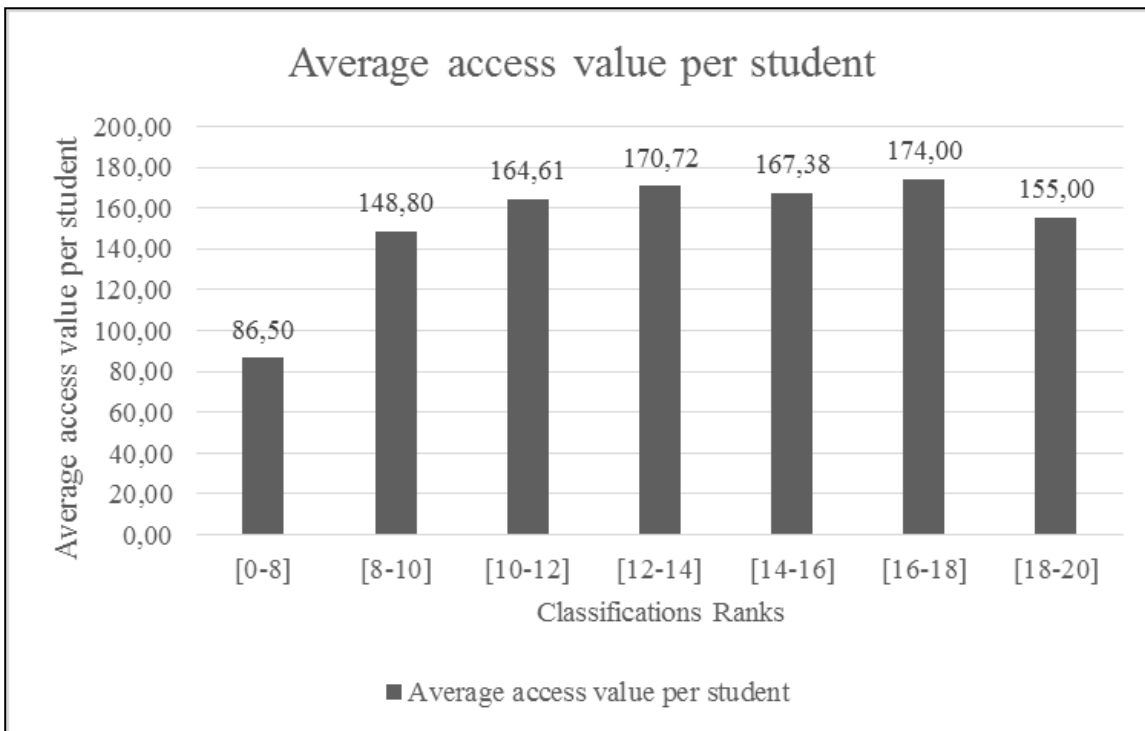
Range of Grades	Students	Cumulative number of students	Total Accesses	Cumulative number of accesses	Average number of accesses
[0-8]	74	74	6401	6401	86,5
[8-10]	56	130	8333	14734	148,80
[10-12]	62	192	10206	24940	164,61
[12-14]	81	273	13828	38768	170,72
[14-16]	65	338	10880	49648	167,38
[16-18]	24	362	4176	53824	174,00
[18-20]	2	364	310	54134	155,00

weekend, and mainly between 8:00 and 24:00 (with a drop in the lunch period and during the dinner period). This hourly trend is also visible at the weekend, but less significant. Figure 3 shows the number of enrolled students that accessed the platform in each hour.

This distribution can be explained by the fact that, on working days, students access the LMS to accompany the classes (night classes go until 23:30). In the weekend, however, students enter the LMS to work autonomously. This explains why the total number of accesses is higher in working days than in the weekend. This ratio between working days and weekend is also visible in Figure 3, which shows the percentage of students who accessed the platform throughout the whole semester. Naturally, the use of the tool on weekends is more significant as it reflects the autonomous and self-motivated use of the platform contrary to most working days' use. Therefore it is quite relevant that about 75% of all the students accessed the platform on weekends for at least one time as it shows a very high commitment level.



**Figure 4.** Total number of platform accesses in total mode by grade



**Figure 5.** Average number of accesses per student by grade

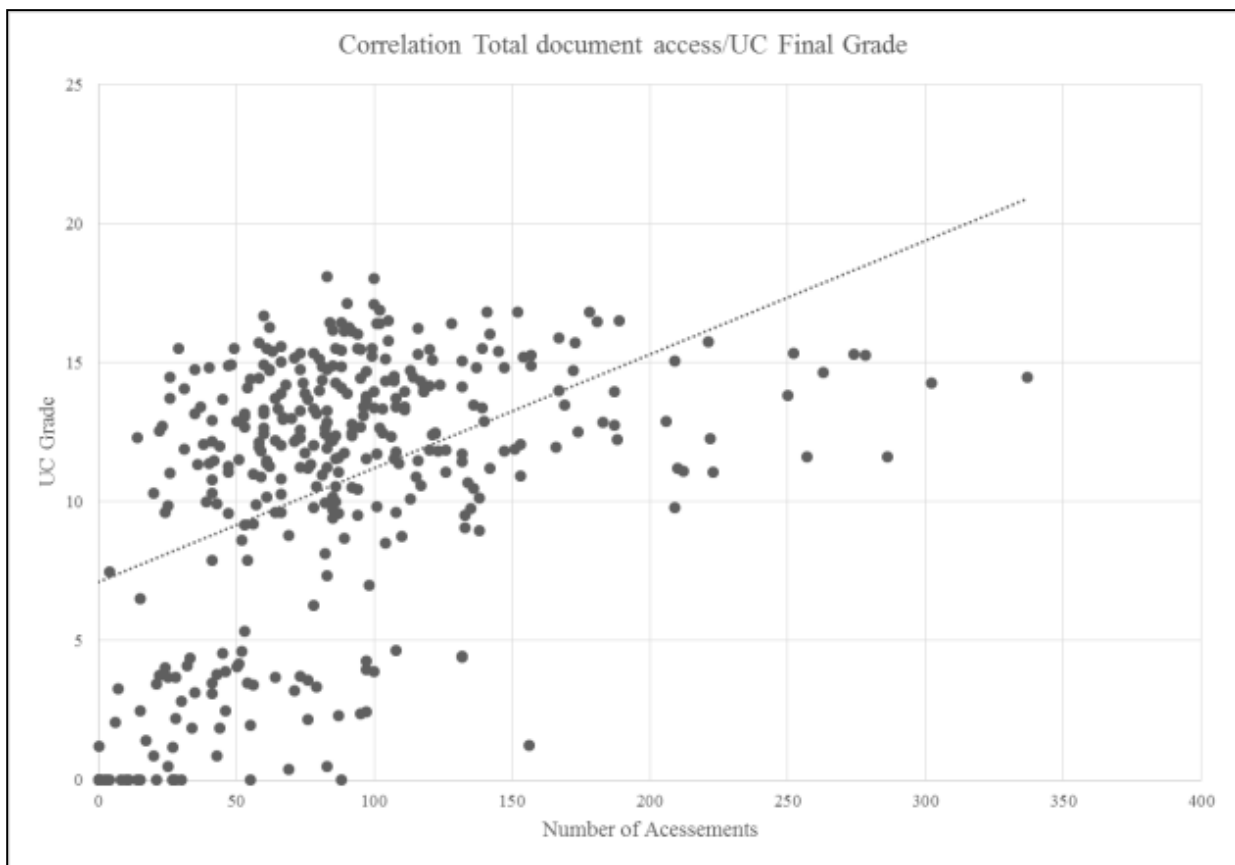
The summary of total accesses made by students is presented in [Table 2](#), ranked by their grades (the grading system uses a scale from 0 to 20 points). The table also presents the normalization (total number of accesses/number of students per range) of those ratings.

From the collected data, also represented on [Figure 4](#), we can see that the number of most relevant accesses corresponds to students who scored between 12 and 16, representing more than 40% of the accesses.

So, although it is not mandatory, a frequent use of the platform seems to lead to good results. [Figure 5](#) shows the average number of accesses per student at each classification level. It is clear that students that had success (grade of 10 or higher) have more accesses on average. Curiously the higher ranked students are not the ones with the higher access average which is a phenomenon that deserves further study. Our observation leads to the preliminary remark that this students do not feel the need to go as often as the others to the platform because they are quite confident on their knowledge of the course contents.

**Table 3.** Summary of correlations related to final grades

	Correlation
Access to documents / Final Grade	0,459403
Quizzes / Final Grade	0,275544
Submission assignments / Final Grade	0,14841
Participation Forum / Final Grade	0,2243



**Figure 6.** Distribution of accesses to course documents versus students’ score

Looking at the failing students, it is obvious that the dropouts have the lower average number of accesses which is natural as they stopped accessing the platform during the semester. Nevertheless, it is clear that the remaining failing students (those that proceeded in the course until the end but failed) had the lowest average number of accesses. Although it was not possible to get a full correlation between the average number of accesses grade and the final grade, it was possible to conclude that there is a correlation between the least number of accesses and the failing students.

**Table 3** shows several correlation scores calculated using access and participation data from LMS activities and the grades obtained by the students. The idea was to refine the previous analysis and try to identify which online learning activities would better correlate with the students’ results.

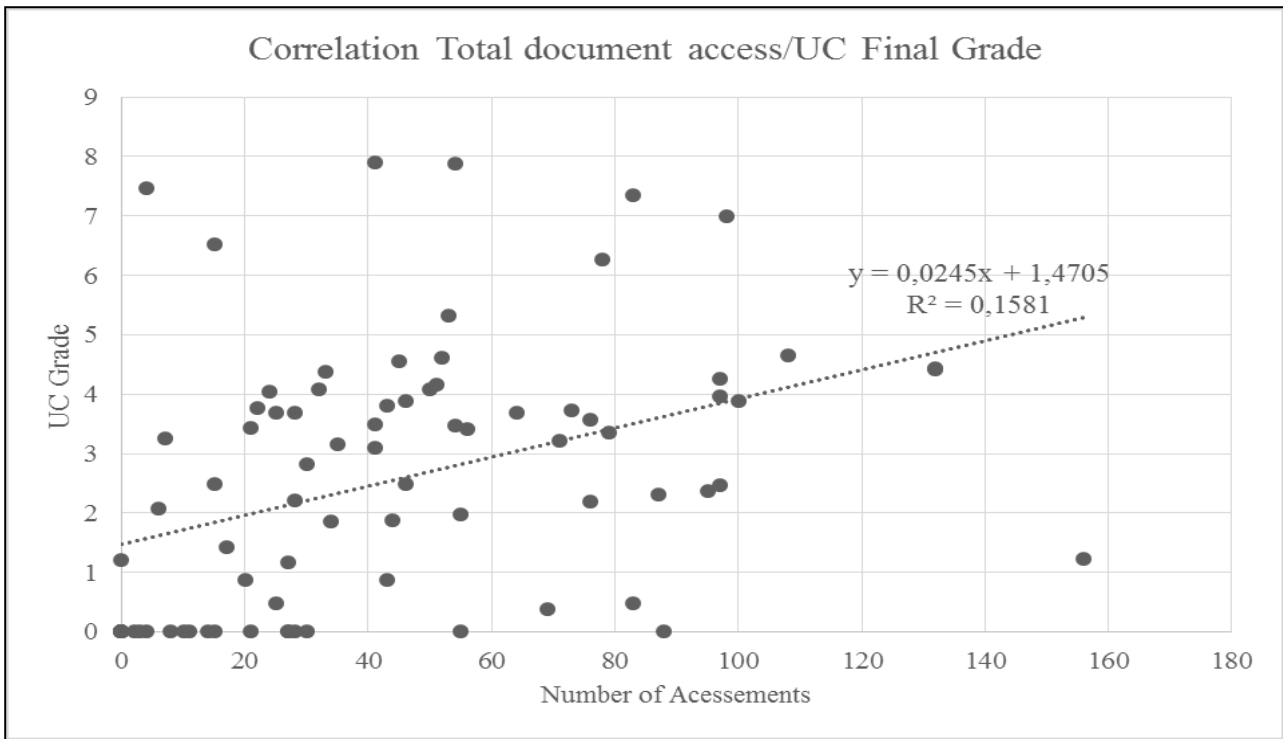
It can be seen that the access to documents (information) has the highest correlation with the learning success. In fact it is the only number with significance as the other values do not reflect any significant correlation. Therefore a more detailed analysis was conducted to evaluate that aspect.

**Figure 6** shows the students’ relation between the access to course documentation and the final grade obtained.

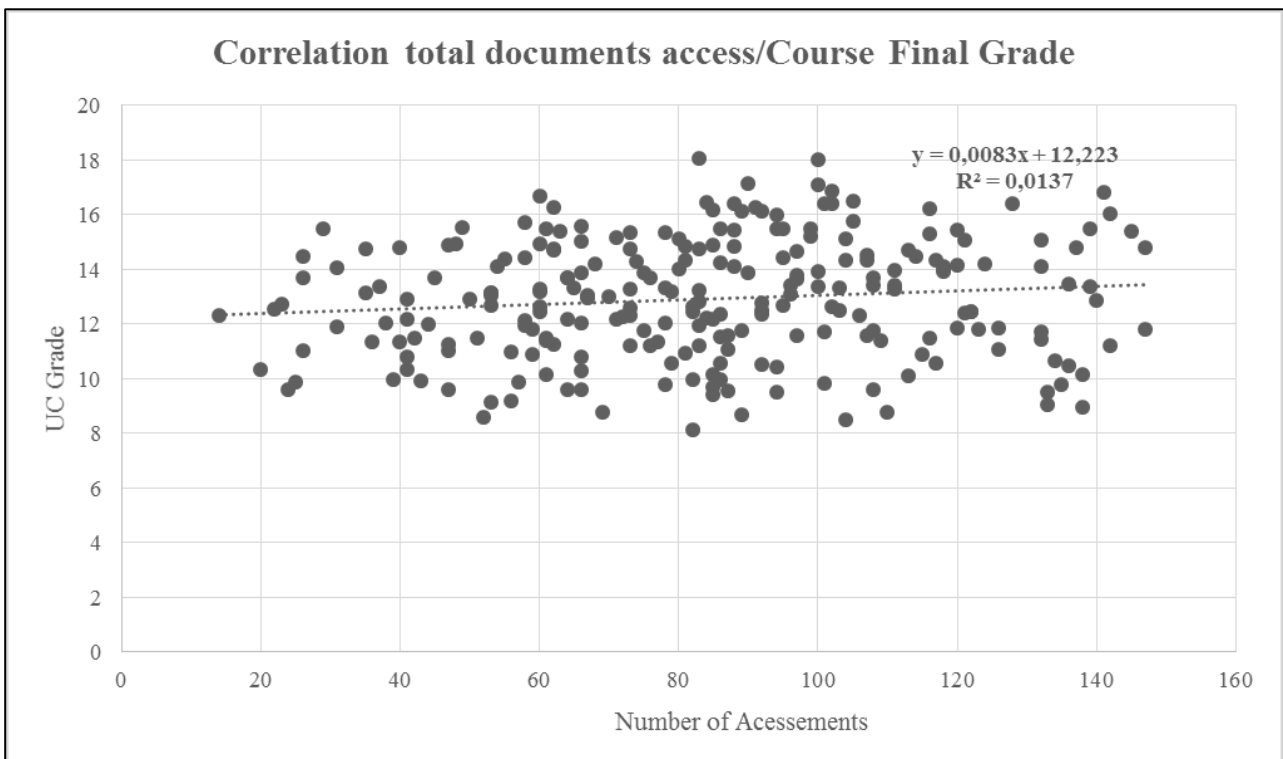
A higher granularity analysis was conducted to identify stronger correlations. Thus, we analyzed the number of accesses to the course documents considering 3 subgroups:

- A. Students who did not obtain the minimum grade of the course (8 values) and did any number of accesses;
- B. Students who obtained the minimum grade of the course (8 values) and made up to 150 accesses;
- C. Students who obtained the minimum grade of the course (8 values) and did more than 150 accesses.

The results are shown in the following figures (**Figure 7, 8** and **9**). **Figure 7** represents the students that did not obtain the minimum grade of the course (8.0 values) and did any number of accesses to the platform.



**Figure 7.** Students who did not obtain the minimum grade of the course (8.0 values) and did any number of accesses



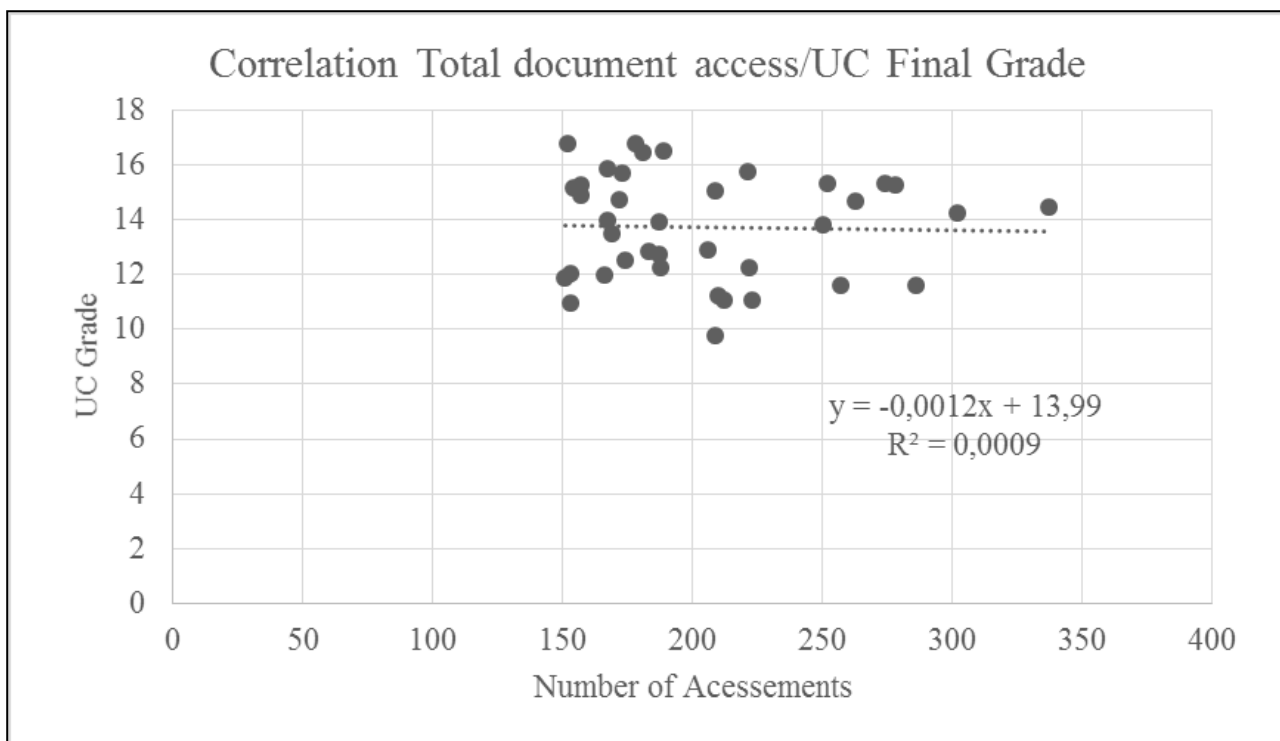
**Figure 8.** Students who obtained the minimum score of the course (8.0 values) and made up to 150 accessions

**Figure 8** represents the students who obtained at least the minimum grade of the course (8.0 values) and made up to 150 accesses to the platform during the course of the course.

**Figure 9** shows the correlation between the students who obtained the minimum score of the course (8.0 values) and did more than 150 accesses to the course platform.

The correlations obtained with this more granular subdivision were much lower than the correlation already obtained for the complete group, so it did not allow drawing additional conclusions. Nevertheless it was considered that by identifying students with a lower number of accesses particularly when these accesses did not conduct to reading or downloading the available documentation, would be possible to signal potential failing students.





**Figure 9.** Students who obtained at least the minimum score of the course (8.0 values) and did more than 150 accessions

## CONCLUSIONS

This study was part of a more general approach to identify key factors of the process of acceptance and adoption of learning technologies by the teachers in order to foster the use of technological-based pedagogical tools. In fact, although teachers sometimes do not demonstrate a strong initial motivation, it was shown that it is possible to foster that adoption. However, the need for better tools was highlighted. The use of tools to analyse the data resulting from the student's interaction with learning tools is one example. This data, when critically analysed, can provide important clues about the students' learning progress. Nevertheless new analytical tools must be developed to allow teachers to fully exploit the information embedded in this data, in a friendly but flexible way.

In particular this article focused on the need to alert teachers to potential problems of students (dropping out or failing). This alert can come from analyzing indicators resulting from the quantitative usage data of institutional learning management system. An exhaustive data collection process was organized about every aspect of access and use of the platform tools. This data was then correlated with the students' success, namely through their final grade in the course. An initial difficulty was related to the actual collection of data and the treatment process. Nevertheless, the collected results showed two potential indicators for these students: 1) lower number of accesses to the platform and 2) lower access to the provided documentation available on the platform.

Even if the extracted data allowed obtaining relevant course information, it still did not allow to design an automatic tool for the treatment of the data collected from the LMS logs. In fact, other factors not included in this study might increase the complexity of the analysis, for example by considering the presence effect in classes. Another relevant problem is that the data used (logs) is very dependent on the structural organization of the course in the learning support platform. To obtain significant results implies a compulsory use of the platform and restructuring the courses according to a given model. To sum up, although there were some positive results observed, more data must be collected to achieve more significant conclusions.

## ACKNOWLEDGEMENT

We would like to acknowledge the support of ISEP's management to this work.



## REFERENCES

- Alves, P., Miranda, L. and Morais, C. (2015). Record of undergraduates' activities in virtual learning environments. *ECEL2015 - 14th European Conference on e-Learning*. Hatfield. pp. 25-33. ISBN 978-1-910810-71-2.
- Black, E. W., Dawson, K., and Priem, J. (2008). Data for Free: Using LMS activity logs to measure community in online courses. *The Internet and Higher Education*, 11, 65-70. <https://doi.org/10.1016/j.iheduc.2008.03.002>
- Castro, F., Vellido, A., Nebot, A. and Mugica, F. (2007). Applying data mining techniques to e-learning problems. *Studies in Computational Intelligence (SCI)*, 62, 183–221. [https://doi.org/10.1007/978-3-540-71974-8\\_8](https://doi.org/10.1007/978-3-540-71974-8_8)
- Dringus, L. P. and Ellis. (2005) Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education*, 45, 141–160. <https://doi.org/10.1016/j.compedu.2004.05.003>
- García, O. A. and Secades, V. A. (2013). Big Data & Learning Analytics: A Potential Way to Optimize ELearning Technological Tools. *Proceedings of the LADIS International Conference e-Learning 2013* ISBN (Book): 978-972-8939-88-5, Prague, Czech Republic 22-26 July.
- Gaudio, E. and Talavera, L. (2006). Data mining to support tutoring in virtual learning communities: Experiences and challenges. In C. Romero & S. Ventura (eds.) *Data mining in e-learning*, 207–226. Southampton, UK: Wit Press. <https://doi.org/10.2495/1-84564-152-3/12>
- Graf, S., Ives, C., Rahman, N. and Ferri, A. (2011). AAT: A Tool for Accessing and Analysing Students' Behaviour Data in Learning Systems. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, pp. 174-179, 978-1-4503-0944-8. <http://doi.acm.org/10.1145/2090116.2090145>
- Khribi, M. K., Jemni, M. and Nasraoui, O. (2009). Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. *Educational Technology & Society*, 12(4), 30–42.
- Krüger, A., Merceron, A. and Wolf, B. (2010). A Data Model to Ease Analysis and Mining of Educational Data. *Proceedings of the 3rd International Conference on Educational Data Mining*, Pittsburgh, PA, July 11-13.
- Lin, F.-R., Hsieh, L.-S. and Chuang, F.-T. (2009). Discovering genres of online discussion threads via text mining. *Computers & Education*, 52, 481–495. <https://doi.org/10.1016/j.compedu.2008.10.005>
- Lino, A., Rocha, Á. and Sizo, A. (2017). Virtual teaching and learning environments: automatic evaluation with artificial neural networks. *Cluster Computing*, 1-11. (in Press) <https://doi.org/10.1007/s10586-017-1122-y>
- Luna, J. M., Castro, C. and Romero, C. (2017). MDM tool: A data mining framework integrated into Moodle. *Comp. Applic. in Engineering Education*, 25, 90-102. <https://doi.org/10.1002/cae.21782>
- Liyanage, M. P., Gunawardena, K. S. and Hirakawa, M. (2015) Detecting Learning Styles in Learning Management Systems Using Data Mining. *Journal of Information Processing*, 24(4), 740-749. <https://doi.org/10.2197/ipsjip.24.740>
- Mamčenko, J. and Šileikienė, I. (2006). Intelligent data analysis of e-learning system based on data warehouse, olap and data mining technologies. In *the Proceedings of the 5th WSEAS International Conference on Education and Educational Technology (EDU '06)*, Tenerife, Canary Islands, Spain, December 16–18, 2006 [CD]. Tenerife: WSEAS, 171–175.
- Marques, B. P. (2015). Parameters of Adoption of E-Learning Technologies in Higher Education: A Case Study. *Pb.D. final thesis. Faculty of Engineering of Porto - University of Porto*.
- Marques, B. P., Villate, J. and Vaz de Carvalho, C. (2010). Technology acceptance on higher education: The case of an engineer's school. *Proceedings of ICERI2010 - International Conference of Education Research and Innovation*, Madrid, Spain
- Marques, B. P., Villate, J. and Vaz de Carvalho, C. (2017). Analytics of student behaviour in a learning management system as a predictor of learning success. *Proceedings of CISTI 2017 - 2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, ISBN: 978-1-5090-5047-5, pp. 1121-1126, Lisbon, Portugal. <https://doi.org/10.23919/CISTI.2017.7975863>
- Mazza, R. and Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *I. Journal of Human-Computer Studies*, 65(2), 125–139. <https://doi.org/10.1016/j.jhcs.2006.08.008>
- Mazza, R. and Milani, C. (2004). GISMO: A graphical interactive student monitoring tool for course management systems. *Proceedings of International Conference on Technology Enhanced Learning '04 (T.E.L.'04)* (pp. 18-19). New York, NY: Springer.
- Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E. and Heiner, C. (2005). An educational data mining tool to browse tutorstudent interactions: Time will tell. *Proceedings of workshop on educational data mining*, pp. 15–22. Available at: <http://www.aai.org/Papers/Workshops/2005/WS-05-02/WS05-02-003.pdf>
- Mostow, J. and Beck, J. (2006). Some useful tactics to modify, map and mine data from intelligent tutors. *Natural Language Engineering*, 12(2), 195–208. <https://doi.org/10.1017/S1351324906004153>
- Mota, D., Reis, L. P. and Vaz de Carvalho, C. (2014). Design of Learning Activities - Pedagogy, Technology and Delivery Trends. *ICST Trans. e-Education e-Learning*, 4, e5. <https://doi.org/10.4108/el.1.4.e5>

- Preidys, S. and Sakalauskas, L. (2010). Analysis of students' study activities in virtual learning environments using data mining methods. *Ukio Technologinis ir Ekonominis Vystymas*, 16(1), 94-108. <https://doi.org/10.3846/tede.2010.06>
- Romero, C., Ventura, S. and García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384. <https://doi.org/10.1016/j.compedu.2007.05.016>
- Santos, L., Escudeiro, P. and Vaz de Carvalho, C. (2014). ICTWays Network: ICT in Science Classrooms. *Proceedings of CISTI 2014 - 9ª Conferencia Ibérica de Sistemas y Tecnologías de Información*, Barcelona, España, 2014, ISBN 978-989-98434-3-1.