Different patterns of students´ interaction with Mooodle and their relationship with achievement

Rebeca Cerezo
University of Oviedo
Faculty of Psychology
+34 627607021
cerezorebeca@uniovi.es

M. Sanchez-Santillan
University of Oviedo
Computer Science
+34 669 094015
melsanchezsantillan@gmail.com

J.C. Núñez
University of Oviedo
Faculty of Psychology
+34 985 103224
jcarlosn@uniovi.es

M. Puerto Paule
University of Oviedo
Computer Science
+34 689384409
paule@uniovi.es

ABSTRACT
This work tends to broaden the knowledge about the learning process in LMSs from an EDM approach. We examine students' interactions with Moodle and their relationship with achievement. We analyzed the log data gathered from a Moodle 2.0 course corresponding to the different interaction patterns of 140 undergraduate students with the LMS in an authentic learning context. We found out 4 different patterns of learning related to different academic achievement.

Keywords
Learning process, LMSs, Moodle, higher education, log analysis.

1. INTRODUCTION
In traditional learning settings, instructors can easily get an insight into the way that the students work and learn. However, in LMSs, it is more difficult for teachers to see how the students behave and learn in the system [2]. Since learner activities are crucial for effective online teaching-learning process, it is necessary to search for empirical methods to better observe patterns in the online environment. In recent years, researchers have investigated various data mining methods to help instructors to improve e-learning process and systems [1]. As shown in the review of Romero and Ventura [3], a good number of quality works have been conducted with techniques similar to the ones used at this work. Most of them were carried out in laboratory settings with concrete tasks, but just a few in real settings or during an extended period of time [2]. These work aims to go beyond laboratory contexts and researcher-controlled settings. Therefore we set two research questions: 1. Are there sense different patterns of students' interaction when they learn in an LMS in a real context? 2. Are those patterns related to students' final marks?

2. METHODOLOGY
2.1 Participants and procedure
The datasets used in this work have been gathered from a Moodle 2.0 course that enrolled 140 undergraduate university students in a psychology degree program at a state university in Northern Spain. The experience was an assignment in the curriculum of a third year mandatory subject. Students were asked to participate in an eTraining program about self-regulated learning related to the subject's topic. The program was composed of 11 different units that were delivered to the students on a weekly basis. Students get an extra point in their final subject grade if they complete the assignments. We have used 12 actions that make the most sense to represent the students' performance in the particular Moodle course described (See Table 1). The variables selected can be grouped into two different groups: Variables related to effort and time spent working (*Time task, Time Span, Relevant Actions,* and *Word Forums)* and Variables related to procrastination (*Day's task* and *Day's Hand-in*). *Final marks* were extracted from the performance in the subject that is the grade of the e-Training program and the sum of the grade in an objective final exam of the subject.

2.2 Data Analysis
First, as an exploratory approach to the optimal number of behavioral patterns or clusters in the LMS, the expectation-maximization (EM) algorithm was used. Second, we sought a similar solution to the one provided by EM for the cluster classification but through the k-means algorithm. The objective of these two first steps is to obtain a clustering solution based on coherence among EM and k-means. Through the clustering, we aim to get high similarity intra-cluster and maximize the differences between them. Finally, ANOVA analyses were run to observe if there were differences between the inter-clusters, and the predictive validity of those clusters to predict final marks.

3. RESULTS
After analyzing the data with the EM algorithm, with k-means and with the elbow method, k = 4 was found to be the optimal number of clusters for this sample. Fig. 1 graphically represents the characteristics of the four groups. The second question was to bring up the chances of those patterns being related to students' final marks. For this purpose, an ANOVA analysis was carried out. The results obtained with final marks as the dependent variable and the different clusters the independent ones where $F(3,136) = 13.31$; $p < .00$; η^2_p .227, indicates that there are statistically significant differences between the four student groups in final marks. The post hoc comparisons showed the following statistically significant differences: cluster 1 vs cluster 2 (d = 0.82, large effect), cluster 2 vs cluster 4 (d = 1.43, very large effect), and cluster 3 vs cluster 4 (d = 1.01, large effect).

Table 1. Name of variables considered in the study with their description and extraction method

Name	Description	Extraction Method under *Moodle nomenclature*	Additional information
Variables related to effort and time spent working			
Time Tasks	Total time spent	Sum of the periods between *quiz view/quiz attempt/quiz continue attempt/quiz close attempt* and the next different action	Students have a period of 15 days to complete the tasks.
Time Span	Total time spent working in every unit	Sum of the variables related to the time spent in the three different type of contents: *Time tasks*, *Time Theory* and *Time Forum*	Students have a period of 15 days to work in a declarative knowledge level (*Theoretical contents*), procedural knowledge level (*Practical tasks*), and conditional knowledge level (*Discussion forums*).
Words Forums	Number of words in forum posts	Extracting the number of *forum add discussion* OR *forum add reply* words	Students do not have a minimum/maximum number of words.
Relevant Actions	Number of relevant actions in the LE	Total of relevant actions considered	Actions such as log in, log out, profile updating, check calendar, refresh content, etc. are dismissed.
Variables related to procrastination			
Day's Tasks	How long students wait to check the task since it was made available in the LE (in days)	Date of task *view* since the task was made available	Students have a period of 15 days to complete the tasks.
Day's "hand-in"	The time taken to hand in the task since the task was made available at in LE (in days)	Date of *quiz close attempt* since the task was made available	Students have a period of 15 days to hand in the tasks.

Regarding the comparisons between cluster 1 vs cluster 4 and cluster 2 vs cluster 3, the inter-cluster differences' effect size was medium.

4. DISCUSSION

Four different patterns of learning with different final marks were found in this course; it is interesting how students with very different patterns in the LMS end with a very similar achievement. Cluster 1 is characterized by a small amount of time allocated to work in general but particularly in the practical task. The variables regarding procrastination and the participation in the forums are low, nevertheless, the overall number of significant actions in the LMS is high. Considering that their achievement is medium-low these results may indicate that students in this cluster work quickly but not efficiently. The students in the Cluster 2 could be described as strategic due to the small amount of time and low number of actions in the LMS that led them to very good results. The pattern for working variables is very suitable, too, with a high quantity of time invested in the tasks and they do not procrastinate. Cluster 3 is similar to the previous one in terms of achievement but not in the remaining variables. This group's achievement is a bit lower than Cluster 2's, it could be labeled as medium-high. There is nothing remarkable about procrastination variables, in contrast, the participation in the forums is really low. The number of relevant actions is also the lowest for this cluster; however, the time that they spent in the LMS was the highest. These results may indicate that they are not strategically efficient and do not make the most of the time spent, but they are still ultimately profitable in terms of achievement. Finally, Cluster 4 is characterized by the lowest marks. The most defining characteristic is that they are extreme procrastinators with really low levels in the variables related to the time spent working. Moreover, they make a significant number of relevant actions but do not benefit from them at all, which denotes a maladaptive approach to learning.

On one hand, these results may help an instructor better understand students' learning process, identify at-risk students (e.g., Cluster 1 and 4) and intervene. On the other hand, the information provided by Clusters 2 and 3 could guide the future development of recommendation systems; having a similar performance in terms of achievement the underlying interaction with the LMS denote different patterns that could be modeled by a recommendation systems in very different terms.

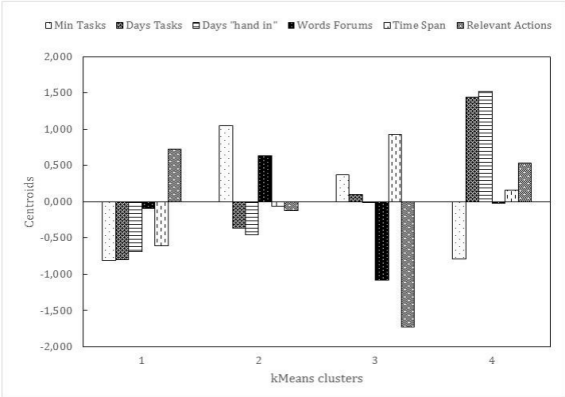

Figure 1. Graphic representation of clustering

5. ACKNOWLEDGMENTS

Our thanks to the Projects TIN2011-25978, EDU2010-16231, GRUPIN14-053 and GRUPIN14-100.

6. REFERENCES

[1] García, E., Romero, C., Ventura, S., & de Castro, C. 2006. Using rules discovery for the continuous improvement of e-learning courses. In *Intelligent Data Engineering and Automated Learning* (Burgos, Spain, September 20 - 23). IDEAL 2006. Springer, Berlin - Heidelberg, 887-895.

[2] Graf, S., & Liu, T. C. 2009. Supporting teachers in identifying students' learning styles in learning management systems: an automatic student modelling approach. *Educational Technology & Society*, 12, 4, 3.

[3] Romero, C. & Ventura, S. 2010. Educational Data Mining: A review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics*, 40, 6, 601-618.