

INTED **2020**

14th International
Technology, Education and
Development Conference

2-4 March, 2020 - Valencia (Spain)

CONFERENCE PROCEEDINGS



Sharing the Passion for Learning

INTED **2020**

**14th International
Technology, Education and
Development Conference**

2-4 March, 2020 - Valencia (Spain)

**CONFERENCE
PROCEEDINGS**

Published by
IATED Academy
iated.org

INTED2020 Proceedings
14th International Technology, Education and Development Conference
March 2nd-4th, 2020 — Valencia, Spain

Edited by
L. Gómez Chova, A. López Martínez, I. Candel Torres
IATED Academy

ISBN: 978-84-09-17939-8
ISSN: 2340-1079
V-153-2020

Book cover designed by
J.L. Bernat

All rights reserved. Copyright © 2020, IATED

The papers published in these proceedings reflect the views only of the authors. The publisher cannot be held responsible for the validity or use of the information therein contained.

APPLICATION OF DATA MINING IN MOODLE PLATFORM FOR THE ANALYSIS OF THE ACADEMIC PERFORMANCE OF A COMPULSORY SUBJECT IN UNIVERSITY STUDENTS

Ana Álvarez-Méndez¹, Teresa Angulo Carrere¹, Jesús Cristobal Barrios²,
Carmen Bravo-Llatas³, M^a Pilar Álvarez Vázquez⁴

¹*Nursing Department, School of Nursing, Physiotherapy and Podiatry, Universidad Complutense de Madrid (SPAIN)*

²*Corporate Software and Academic Management Service, Universidad Complutense de Madrid (SPAIN)*

³*Teaching and Research Support Service, Universidad Complutense de Madrid (SPAIN)*

⁴*Cell Biology Department, School of Medicine, Universidad Complutense de Madrid (SPAIN)*

Abstract

E-learning platforms used in Higher Education Institutions store valuable information that can be analysed. Data mining is a multidisciplinary technique that integrates computer science, education, and statistics, and that could serve to interpret results and predict academic performance through the virtualized subjects.

Objective: To identify the use of the interactive platform based on Moodle LMS (Learning Management System) in the Occupational Health compulsory subject in the curriculum of the podiatry degree and its relationship with academic success from 2017 to 2019 courses.

Materials and methods: Logs (files of the interactions between a system and the students in the virtual system) for analyzing a variety of information from different activities carried out through the virtual campus on a Moodle platform were extracted, depurated and prepared for analysis. Finally, 33,776 (13,818 in 2017 and 19,958 in 2018) logs were used to perform the statistical analysis using the RStudio program and the SPSS v.22 software. A descriptive analysis, Pearson correlations and continuous variable decision trees diagrams were performed to determine the use of the Moodle classroom activities and resources and their relationship with academic results obtained in this subject.

Results: 62 students enrolled in the academic course 2017-18 and 59 in the 2018-19 were studied. In the academic course of 2017-18, 62.9% were women, and mean and SD of academic results was 7.5 ± 1.09 . In 2018-19 academic course, 76.3% were women, and the mean results was 7.2 ± 1.06 . The highest peak of activity registered on the virtual subject was 200 visits in each academic course, with differences by months in relation to the distribution of tasks. The highest activity recorded was on Tuesdays and Sundays, in both years, but with more activity in the 2018-19 academic year. Lessons were the most used tools in both courses (40.3% in 2017 vs 46.2% in 2018) followed by the participation in forums (32.7% in 2017 vs 12.8% in 2018). Participation in the forums was 100% vs 93.1% and URLs entries 71.43% vs 87.93%; comparing both academic courses. In 2018, 3 new tools were introduced with high participation: Self-assessment Test 96.55%, glossary 87.93% and a wiki activity with a 48.28% of participation. Tools that significantly correlated with better test scores in year 2017 was the participation in the forums ($p=0.016$), while in 2018 test scores were significant correlated with the participation in the tasks ($p=0.041$) and self-assessment tests ($p=0.044$) carried out on the virtual classroom. Tree-like graphs identified two clusters of students related with forums and URLs entries, in the virtual classroom.

Conclusion: This study reveals the importance of identifying and selecting tools with the capacity to improve and stimulate active and significant learning.

Keywords: Higher education, Moodle platform, e-learning, student performances, Occupational health, learning analytics.

1 INTRODUCTION

The incorporation of virtual educational models allows adapting the teaching environment for improving the student's learning and therefore the academic performance [1]. Some researches on

learning process have shown that different students prefer and learn according to different types of resources, which should be integrated in the respectively subjects, for that it is important to build models for adapting the learning space to different student's needs [2]. Online learning platforms are a technology framework used to create online courses that usually allow to register, monitor and evaluate activities and manage contents, as well as to exchange information among dispersed students through synchronous and asynchronous communication, and are currently very used in Higher Education Institutions [3]. Moodle (Dynamic learning environment oriented to modular objects) is the Open Source Learning Management System (LMS) that has been selected by the Complutense University of Madrid (Spain) for supporting the teaching and learning process as it favors contact between teachers and students and helps to create interactive learning environments that generate knowledge effectively [1,4]. These e-learning platforms store huge amounts of data that can be analyzed for getting value information [5]. Data mining is a multidisciplinary technique that integrates computer science, education and statistics, and that could serve to interpret results and to predict academic performance through the virtualized subjects [6]. Participant's tracking data in the virtual space (VS) can be extracted applying this technology. Moodle has very wide options with different access possibilities, modular structure, and advanced backup tools [5]. Once extracted and properly debugged and processed, data can be analyzed in order to be converted into the useful information that allows recognizing patterns in students' performance [7,8].

In this article, we present a study was conducted in two academic courses, 2017-18 and 2018-19 in the school of Podiatry at the Complutense University of Madrid. We describe the students' interactions with the available resources and different tasks in the VS of Occupational Health. Also, we analyze the relationships between the actions and activities carried out by the students within this online learning process and their academic performance.

2 METHODOLOGY

Occupational health is a compulsory subject in the curriculum of the podiatry degree in the Complutense University of Madrid. It is taught in the first quarter of the fourth academic year and its content focuses on the prevention of occupational hazards. It presents a workload of six European Credit Transfer System (ECTS) distributed in attending classrooms, taught 3 days a week, and some activities organized in four seminars throughout the semester. Classes are scheduled in the morning. Classroom activities are complemented with online activities through the Moodle platform. Moodle platform provides a complementary set of tools to support teaching and to improve learning process through resources, modules and activities to facilitate interaction among students and between students and teachers.

The subject's VS contained different resources and tasks created and uploaded by the teachers: word and PDF files, PowerPoint files summarizing the lessons and a wiki, a glossary of terms and a quiz for self-evaluation. Also, some Moodle modules to provide interactions such as chats and forums were available. Finally, the teachers included several complementary educational materials, such as image, videos and audio files, links to blogs and educational pages, as well as articles and supporting books. All these resources were available throughout the semester.

In the 2017-18 course, the scheduled activities in the VS were the presentation of a group work and the participation in two forums, while in the 2018-19 course, a glossary, a wiki and a quiz for self-evaluation were added. The deadlines were 1 month for each task.

The final mark was obtained as the addition of the continuous virtual activities such as teamwork tasks (10%) and participation in forums (10%), and the classroom final exam grade (80%).

Data stored in Moodle platform were extracted as logs or files of the participants' interactions carried out in the VS. They were debugged, anonymized and prepared for analysis with the R program, a free software environment for statistical computing, graphics and RStudio, an integrated development environment for R.

The variables chosen to be analyzed were: the total number of logins in the virtualized course, the average flows of log-in frequency by the hour of the day, according to the days of the week and for the months of the semester, as well as the average of visits to the different resources and tasks. These data were compared with the data of the global activity of the Moodle platform in the community of the Complutense University of Madrid, obtained from Google Analytics (unpublished data).

A descriptive analysis of continuous variables was presented as means and SD or median. Mann-Whitney U test was used as a non-parametric test to compare the variables and to evaluate differences between groups. For qualitative parameters, absolute and relative frequencies were presented. Spearman's rho was used to evaluate the correlation between non-parametric variables. Decision tree recursive with the minimization of variance (CRT) cut/cutting criterion was performed to predict the students' academic performance using the Moodle classroom activities and resources. Decision trees are hierarchical and graphical representations of interactions between variables. A decision tree was constructed by recursive partitioning —starting from the root node (known as the first parent); each node can be split into left and right child nodes. These nodes can then be further split and they themselves become parent nodes of their resulting children nodes. This splitting process/method starts by considering each feature in data set [7]. The CRT growing method attempts to maximize within-node homogeneity. CRT splits the data into segments that are as homogeneous as possible with respect to the dependent variable. A terminal node in which all cases have the same value for the dependent variable is a homogenous node that requires no further splitting because it is "pure" [9]. R Analytics (or R programming language), and SPSS Statistics V22 software (SPSS Inc., Chicago, IL, USA) were used for data analysis. P-value <0.05 was considered statistically significant.

3 RESULTS

During the 2017-18 academic course, 62 students were enrolled in the subject, 62.99% of which were women. In 2018-19, 59 students were enrolled, being women 76.3% of them.

For the analysis, the number of the logs extracted from Moodle amounted to 33,776 (13,818 in 2017-18 and 19,958 in 2018-19). All the students participated in the VS of the subject and used different resources such as lessons, forums and other activities. Regarding to the distribution of total entries in Moodle tools, the most entries were for accessing the PowerPoint slides of the lessons, in both courses (40.3% of the students in 2017-18 vs 46.2% in 2018-19). Comparing participation in 2017-18 vs 2018-19, we found that participation in forums was 32.7% vs 12.8%, and URLs entries were 71.43% vs 87.93%. In 2018-19, 3 new tasks were added, a self-assessment test, a glossary and a wiki, each of them getting high participations: 96.55%, 87.93% and 48.28%, respectively.

In both courses, the maximum number of visits reached in the VS was 200, which is a value much higher than the UCM average, as is shown in Fig 1. This peak mainly coincided with the date of the exam. Differences in access according to the months were probably due to the distribution and dates of delivery of the tasks, given that access peaks observed during the months of October, November and December in 2018-19 course may be related to the due dates of the new tasks, such as the glossary, the quiz and the wiki (Fig. 1).

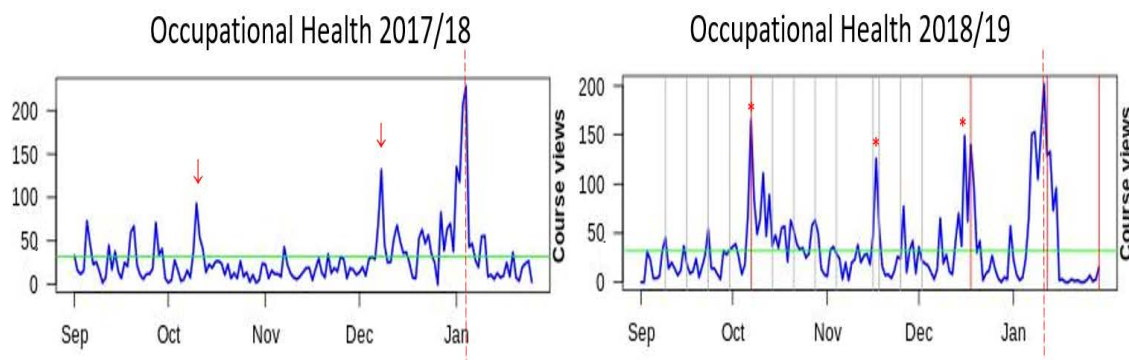


Figure 1. Number of entries to the VS per month. Green lines represent UCM average. Maximum peak (red dashed lines) was observed in January, before the exam' date, in both academic courses.

Analyzing the number of online accesses by the days of the week in both academic courses, we observed that, in general, there was a higher activity in the 2018-19 academic course. In this course, the profile of the highest activity was registered on Tuesdays and Sundays, while in 2017-18 it was on Tuesdays and Thursdays (Fig. 2A). Activity on the VS also varied along the hours of the day, as is clearly shown in Fig. 2B and 2C. These variations, according to the days of the week and the hours of the day, seemed to match with the schedule of classroom lessons and of other potential school activities.

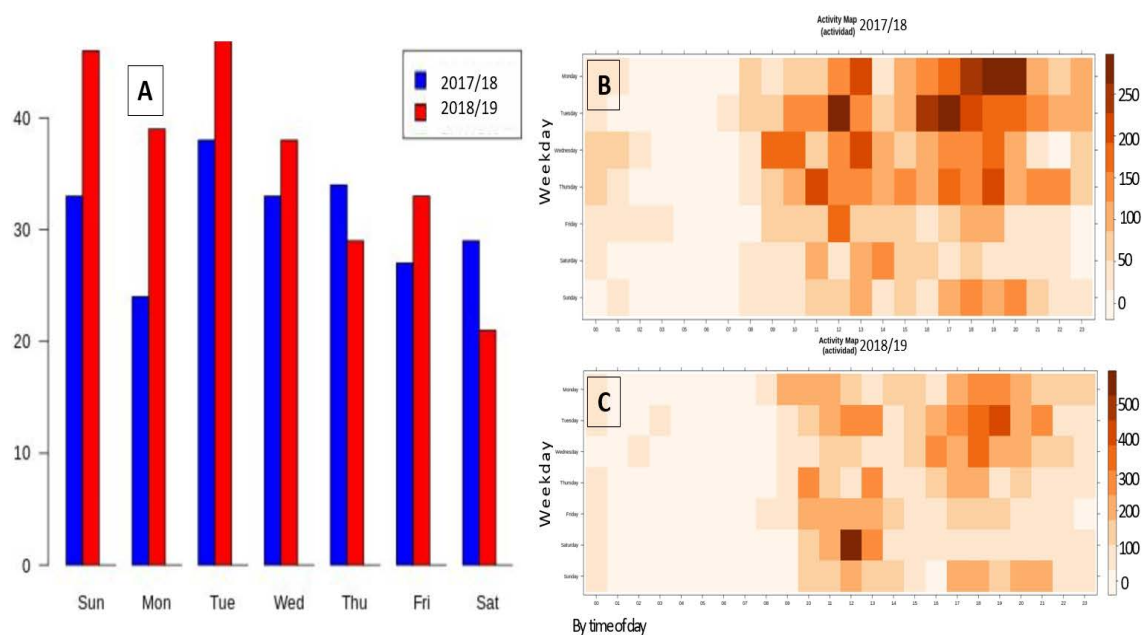


Figure 2. A: Frequency of online access by the days of the week in both courses. B and C: Maps of online activity by the time of the day and weekday in 2017-18 and 2018/19 courses, respectively.

In the subject of Occupational Health, there was one ordinary examination call and one retake exam per academic course. The multiple-choice questions exam scores represented 80% of the final grade, and the remaining 20% was obtained through the continuous work done in the VS. In the 2017-18 academic course, almost 60% passed the ordinary exam. The final average grade was of 7.5 (SD = 1.09; min 0, max 10), while in the 2018-19 course, final average grade was 7.27 (SD = 1.06; min 0, max 10) and 55% of the students passed the ordinary exam. Fig. 3 shows the outcomes in the multiple-choice test in the two academic courses as percentages of passed, failed and absent students.

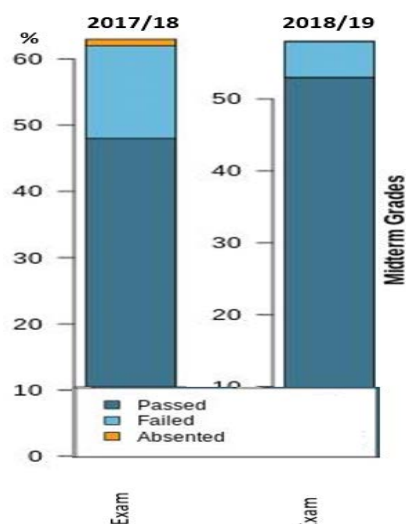


Figure 3. Scores of the ordinary exam call in 2017-18 and 2018-19 courses (48% vs 53% passed).

Comparing the use of the different tools available in the VS in both academic courses, only statistically significant differences were observed in the participation in the forums (Fig. 4A), which was significantly higher in the 2017-18 course ($p < 0.001$, MWW), and in the use of URLs (Fig. 4B), that was significantly higher in the 2018-19 course ($p = 0.03$, MWW).

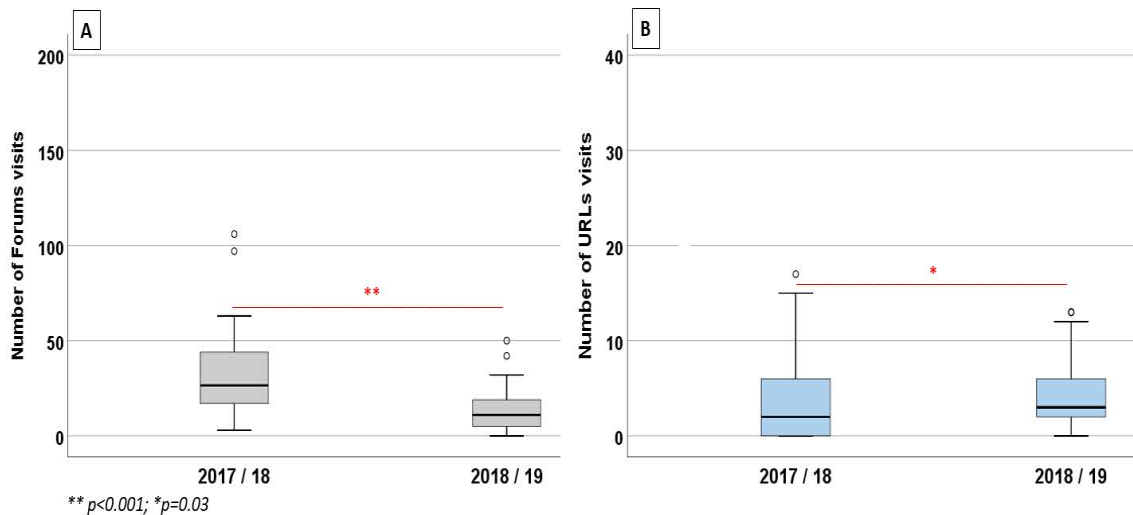


Figure 4. Statistically significant differences in the participation in forums (A) and in the number of URLs visits (B) in both academic years [Non-parametric Mann–Whitney U test (MWW)].

When performing a Spearman's correlation analysis between the final grades and the recorded activity in the subject's VS, there were no significant correlations with the use of resources/URLs nor with the participations in quiz, wiki, glossary ($p > 0.05$). In the 2017-18 course, only the use of the forums correlated significantly with the multiple-choice test score (Spearman Rho (ρ), $\rho = 0.304$; p -value = 0.016), while in the 2018-19 course two statistically significant correlations were found: between the multiple-choice test score and the number of visits to the VS ($\rho = 0.275$; p -value = 0.04) and the participation in tasks ($\rho = 0.310$; p -value = 0.02).

Decision trees that best classify students according to the multiple-choice exam score (*dependent variable*) and the use of the VS were shown in Fig. 5. In the 2017-18 course, the independent variables introduced in the decision tree cut/cutting method were three, namely, visits, resources, and forums, showing a quality (weighted terminal node variance) of 0.672 (SD = 0.105; Method of Growth: CRT). The decision tree obtained showed three groups: those with a low participation in forums (<14) had a mean score of 4.86 (SD = 1.11), those with a medium participation (<41) a mean of 5.56 (SD = 0.84) and those with a high participation (>40) a mean of 6.088 (SD = 0.63). In the 2018-2019 course, the independent variables introduced in the model were visits, resources, forums, quiz, glossary and wiki, showing a quality of 0.575 (SD = 0.106; Method of Growth: CRT). This time, only a participation in forums greater than 11 entries, increased scores up to 5.76 (SD = 0.8). In the group of students with a low participation in forums (<12) the number of visits was related to scores: students with lower number of visits (<35) had a mean score of 4.72 (SD = 0.55), and those with higher number of visits (>36) had a mean score of 5.32 (SD = 0.71).

In the Fig. 5, the blue/red text represents a condition/internal node, based on which the tree splits into branches. The split with the best score (or highest score) was selected. In both courses, the variable that best classified students to better grade exams of the subject was forums.

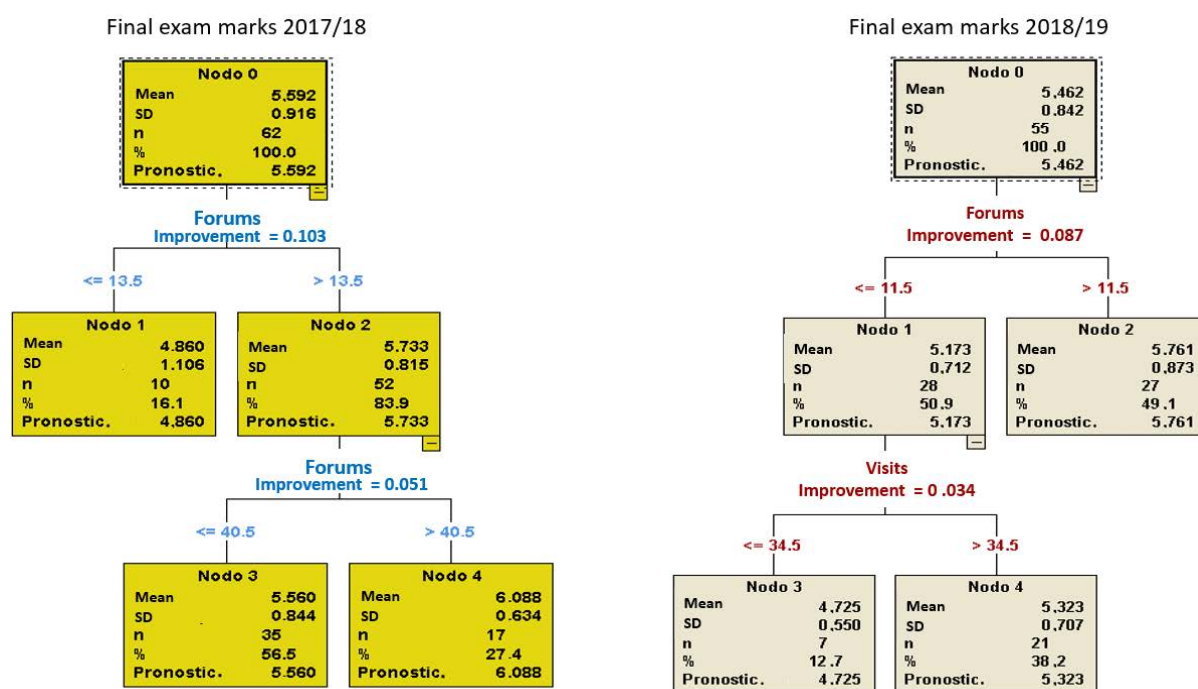


Figure 5. Decision tree for predicting continuous values of the students' marks and VS activities in the Occupational Health subject in the two academic years, 2017-18 (left) and 2018-19 (right) (Growing method: CRT).

4 DISCUSSION AND CONCLUSIONS

Research on virtual learning environments demands that both teachers and students deepen in the use and updating of online resources as well in the designed activities to improve students' learning. Virtual learning, with this Moodle educational technology, depends largely on the self-motivation of the student, as an undoubtedly necessary factor, to actively participate in their learning process and to achieve good results [10]. For all that, it is very important to upgrade interactive tools, so that teachers can design a clear framework and can stimulating the distance learning through the Moodle platform, otherwise, the VS would not positively influence the learning process of the students.

In this article, we presented an overview of the students' behaviour in the VS created for the Occupational Health, a compulsory subject in the degree of Podiatry. We analysed their interactions with the resources and activities offered and their possible correspondence with the academic performance.

Through the tools incorporated in the Moodle platform, the most widely used platform in open source learning management system in higher education, we analysed a set of activities such as homework delivery, forums, questionnaires, wikis, glossaries, URLs as well as the connection time to the VS. Students enrolled in the subject of Occupational Health were able to access all academic materials and work in practical activities, as well as using links to multimedia contents intended to help them to improve learning and complete their training. Furthermore, students could find all the information regarding their exam's dates, course grades, learning support bibliography, academic grades, and all the complementary information such as news, cultural events, instant notifications of messages, images, audio, video and other files, throughout the school semester. Our results showed that there was not a high variability in the students' behaviour on the VS. While lessons and seminars were the most common framework of the teaching-learning process in the traditional educational structure, in the e-learning and b-learning modalities, students should be able to maintain their own learning pace and to adapt the work according to their available time. Also teachers should be able to better adjust the works to the levels of knowledge of each student, improving the interaction between teachers and students, as well as with the didactic contents [11].

The analysis of the students' behavior revealed that the main use of the VS was to download materials such as text and slide files. Participation was good in those tasks that provided points for the final exam, such as the forums. On the contrary, low participation was found in relation to non-qualifying

tasks. This seems to support the idea that our students used the VS primarily as a repository of study materials and information and that they focused on tasks that provided bonus points to the final grade.

In the academic year 2018-19, we introduced new tools such as a glossary, a quiz and a wiki. Quiz had the greatest participation (96.5%), followed by the creation of a glossary of terms (87.93%) related to occupational health. However less than 50% of the students worked on the wiki. Although the wikis have a great potentially value in the collaborative teaching-learning procedure, enhancing collaborative creativity, probably one handicap could had been the fact of being a resource rarely used for teaching or maybe the low participation could be due to the students' lack of motivation because of the time consuming in creating and developing the wiki. In our experience, students usually do not participate much in the activities if there is no grading "retribution" for that participation, as it was our case. Our study allowed us to realize that the students were not very trained in how to use wikis. In addition, we also noted that was difficult to rate their activity on wiki, coinciding with what other authors have already published [12,13].

Although we did not find significantly relationships between the final grades and resources, URLs, quiz, glossary and wiki, three statistically significant correlations at a level 5% were revealed: to the use of forums in the 2017-18 course, and to the number of visits to the VS and to the participation in tasks in the 2018-19 course. Results showed that students were highly trained in discussion forums use, given that they are good tools for developing the ability to think, and in our study, their participation and debate in the forum contributed up to 10% to their final grade. It is also worth noting that the number of posts and discussions in the forums were substantially larger with respect to the debates in class, so that forums in VS were perceived as useful and they improved effectively foster students' questioning and reason abilities.

Tree-like graphs identified three clusters of students related with participation in VS forums in 2017-18. In 2018-19, three clusters of students related firstly with participation in forums and secondly with the total number of entries in the VS. Through the decision trees, one tool among all the Moodle tools studied emerges as one that best classified our students to obtain the best academic performance in both academic courses: the participation in the forums.

Although designing this Occupational Health subject in a Moodle platform and creating resources and materials had required enormous demands from teachers, our findings suggested that material and activities may have a direct positive impact on the outcomes of the students' learning, agreeing with others authors [14,15]. We detected some weaknesses in the use of some Moodle tools that will be important to improve to better adapt teaching-learning process to a new Information and Communication Technologies (ICT) models in education. Furthermore, we should be able to identify undergraduate students who are less trained with some ICT tools, what would entail some future disadvantages in the era of knowledge society, like preventing the acquisition bibliographic searching skills, worse ability to analyze and discuss the scientific production, as well as a poor critical ability to face the new technological changes. All of these abilities are currently very necessary for full participation in the digital society of the future and changing jobs, aspects that are becoming an important consideration for curricular frameworks [16-18].

In the future, we would like to further research the use of different Moodle activities and the students' preferences, in order to manage and improve their academic performance in the e-learning process. In this sense, data mining analysis for students' tracking within Moodle platform, could provide a collection of technical information with a textual representation of the activity within the subject, besides its pedagogical deeply meaning.

ACKNOWLEDGEMENTS

This work is part of the Innovation and Improvement of the Quality Teaching project (pimcd2018-20) of the Universidad Complutense de Madrid. We are very grateful for the support of the Vice-Chancellor of Technologies, the Office of the virtual campus and the collaboration of our students.

REFERENCES

- [1] F. J. García-Peñalvo, A Hernández-García, M. A. Conde, A Fidalgo-Blanco, M. L. Sein-Echaluce, M. Alier, et al. "Looking into the future: Learning services-based technological ecosystems", III Congreso Internacional sobre Aprendizaje, Innovación y Competitividad (CINAIC 2015), 2015.

- [2] G. A. Maldonado Berea, and E. Vega Gea. "Actitud de los estudiantes universitarios ante la plataforma moodle. ATTITUDE OF UNIVERSITY STUDENTS TOWARDS THE MOODLE PLATFORM ". *Píxel-Bit, Rev Medios y Educ*, 2015, no. 47, pp. 105–17, 2015. Retrieved from: <https://idus.us.es/xmlui/handle/11441/45296>
- [3] J. Cabero-Almenara, M. L. Arancibia, and A. del Prete. "Technical and Didactic Knowledge of the Moodle LMS in Higher Education. Beyond Functional Use", *J New Approaches in Educational Research*, vol. 8, no. 1, pp. 25–33, 2019. Retrieved from: <https://naerjournal.ua.es/article/view/v8n1-4>
- [4] M. F. Paulsen. "Experiences with Learning Management Systems in 113 European Institutions", *J Educ Technol Soc [Internet]*, vol. 6, no.4, pp. 134–148. 2003. Retrieved from: <http://www.jstor.org/stable/jeductechsoci.6.4.134>
- [5] C. C. Aydin, and G. Tirkes. "Open source learning management systems in e-learning and Moodle. In: IEEE EDUCON 2010 Conference. 2010. p. 593–600. 2010.
- [6] C. Romero, S. Ventura, and E. García. "Data mining in course management systems: Moodle case study and tutorial. *Comput Educ*, vol. 51, no. 1, pp. 368–84, 2008. Retrieved from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.331.667&rep=rep1&type=pdf>
- [7] W. Villegas-Ch, and S. Luján-Mora. "Analysis of data mining techniques applied to LMS for personalized education". In: 2017 IEEE World Engineering Education Conference (EDUNINE).. pp. 85–9, 2017. Retrieved from: <https://ieeexplore.ieee.org/document/7918188?denied=>
- [8] D. Watson. "Understanding the relationship between ICT and education means exploring innovation and change". *Educ Inf Technol [Internet]*. vol. 11, no. 3, pp. 199–216, 2006. Retrieved from: <https://doi.org/10.1007/s10639-006-9016-2>
- [9] C. Machuca, M. V. Vettore, M. Krasuska, S. R. Baker, and P. G. Robinson. "Using classification and regression tree modelling to investigate response shift patterns in dentine hypersensitivity". *BMC Med Res Methodol [Internet]*. Accessed Aug 14; vol. 17 no. 1, pp. 120, 2017. Retrieved from: <https://www.ncbi.nlm.nih.gov/pubmed/28806921>
- [10] S. A. García Flores. "Participación activa y valoración de los materiales didácticos en el éxito/fracaso académico en Educación a Distancia (Active participation and evaluation of the didactic materials in success / failure academic in Distance Education). *Hamut'ay*. vol. 5, no. 2, pp.32-45, 2018. <http://dx.doi.org/10.21503/hamu.v5i2.1619>
- [11] J. Bierne. "Actualizing Moodle Interactive Tools Usage within Distance Learning: Need for Multilevel Approach". *Int J Inf Educ Technol*.vol. 3,no. 1, pp.44–7, 2013. Retrieved from: <https://pdfs.semanticscholar.org/6c14/cdaacb12a13c7ff86740492710266165493e.pdf>
- [12] M. D. Lytras, P. Ordonez De Pablos, A. Ziderman, A. Roulstone, H. Maurer, and J. B. Imber (Eds). "Knowledge management, Information Systems, E-Learning; and Sustainability Research". vol. Part I. Berlin Heidelberg: Springer-Verlag. pp. 22–24, 2010. Retrieved from: <https://link.springer.com/content/pdf/10.1007%2F978-3-642-16318-0.pdf>
- [13] J. Lindsay. "Wikis that work. Practical and Pedagogical Applications of Wikis in the Classroom". [Internet]. 2007. Retrieved from: <http://julielindsaylinks.pbworks.com/w/page/12444656/Wikis%20that%20Work>
- [14] T. Yu, and IL. Jo. "Educational technology approach toward learning analytics: Relationship between student online behavior and learning performance in higher education". In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK '14)*. Association for Computing Machinery. pp. 269–270, 2014. DOI: <https://doi.org/10.1145/2567574.2567594>
- [15] S. Kotsiantis, N. Tselios, A. Filippidi, and V. Komis. "Using learning analytics to identify successful learners in a blended learning course". *Int J Technol Enhanc Learn*.vol. 5, no. 2, pp. 133–50, 2013.
- [16] M. Area-Moreira, V. Hernandez-Rivero, and J. J. Sosa-Alonso. "Models of educational integration of ICTs in the classroom". *Comunicar. Media Educ Res J [Internet]*. vol. 47, no. XXIV, pp. 79–87, 2016. Retrieved from: https://www.scipedia.com/public/Area-Moreira_et_al_2016a#article_es
- [17] S. T. Kerr. "Why we all want it to work: towards a culturally based model for technology and educational change", *British J Educ Technol*, vol. 36, no. 6, pp. 1005-1016, 2005.

- [18] H. Drenoyianni. "Reconsidering change and ICT: Perspectives of a human and democratic education". *Educ Inf Technol*, vol. 11, pp. 401–13, 2006.