Study of the academic behavior of participants in a MOOC course on Energy Sustainability using Principal Component Analysis (PCA)

Gioconda Riofrío-Calderón
Department of Education Sciences,
Private Technical University of Loja
Loja, Ecuador
gcriofrio@utpl.edu.ec, 0000-0001-5116-381X

María-Soledad Ramírez-Montoya
Institute for the Future of Education,
Technological of Monterrey
Monterrey, Mexico
solramirez@tec.mx, 0000-0002-1274-706X

***Abstract*— Massive Open Online Courses (MOOCs) have generated great expectations around the world, hence the need to continue researching this educational phenomenon. The purpose of this study is to perform an analysis of the data obtained in a MOOC course on Energy Sustainability, to determine the academic behavior of the participants, based on numerical variables and categorical variables. We analyzed 13 variables raised in the MOOC course to a population of 453 participants whose condition was the completion of an initial and final survey, that is, they remained in the course until the end of it, for data analysis was used the "Principal Component Analysis" (PCA), as for the software used is the PCAmixdata package, which allows working with numerical and categorical variables. The results reflect findings on variables that account for the influence and relationship that exists in certain activities with the approval of the participants. The study contributes for both teaching and technical teams to consider the results in future MOOC designs and also to incorporate valuable elements of data analytics.**

Keywords*— *MOOC, component analysis, educational innovation, energetic sustainability.

I. INTRODUCTION

It has been more than a decade since the Massive Open Online Courses, MOOCs, came to light and since then they have not ceased to be news, especially in the global academic community. The genesis of this movement was given since the first MOOC was launched in 2008 and in the following years would be where its dissemination and incorporation in different platforms of universities and companies around the world took force. So far a lot of water has flowed, and there have been countless writings generated by MOOCs. On the one hand, the enormous advantages of being massive and open are identified, however, on the other hand, it was found with a worrying reality such as high dropout rates. However, beyond the ups and downs encountered along the way, the innovative character of MOOCs is undeniable, as they incorporate various elements that were not present in a formal offer. MOOCs were presented with this innovative character as another way of reaching broad sectors with education.

As can be seen, there are important elements that contributed to the creation and rapid expansion of MOOCs. Evidently, the Figures attest to the true revolution caused by this phenomenon: millions of people around the world have enrolled in or followed a MOOC course [1]. It should be noted that one of the levels where this offer was most widely received is in higher education, because one of the main reasons for its evolution was to make higher education more accessible to people around the world [2]. However, the offer also contributed and motivated other sectors where it had great acceptance and interest such as professionals who take advantage of the flexible delivery of this offer and the lack of a cost to meet the learning needs [3]. Therefore, the MOOC phenomenon, regardless of the level or sector from which it is approached, in general, we can indicate that advantages have been attributed to it because it serves and responds to different sectors.

Focusing exclusively on the educational level, it is interesting to review the issue of learning and performance in order to diagnose what results have been presented in these areas based on the experiences carried out. In this way, it is pointed out that from an educational point of view, MOOCs are becoming a technology that is gaining ground and some scholars believe that it will transform higher education [4]. However, in addition to the transforming factor of MOOCs, it is worth reflecting that MOOCs are technological innovations to generate learning and have been the subject of study due to their potential and performance characteristics [5]. In this sense, according to [6], MOOCs should be considered as instruments that make it possible to significantly improve the learning of the participants; therefore, they are allies in the educational environment. Evidently, strengthening the educational process today is a challenge, since what is aspired is the development of society that comes from education and if MOOCs have done their bit, the activity has been worthwhile.

Regarding the evaluation processes, specifically how to evaluate in a MOOC, there are studies that address the issue since it is important to consider what criteria have been given with respect to learning in MOOCs. In this context, it is pointed out that a traditional form of student learning is through videos and the completion of online questionnaires, however, it has evolved to other forms of evaluation [7]. Starting from the typology of MOOCs, an interesting contribution is the analysis that points out that for the xMOOC, since the content is the main focus of the evaluation, then test-type tests are applied, while in a cMOOC it is based more on peer evaluation [8]. Finally, another important contribution is that which addresses in general the different possibilities of applying evaluative processes to measure learning in a MOOC in this way, for example, multiple choice tests, programming tasks, essays, exams, automatic evaluation of problems, among others [9]. Although in previous lines it was stated that the MOOC line has been widely analyzed in the last decade, the issue of how a MOOC is evaluated and which instruments are applied are aspects that should perhaps be strengthened.

978-1-6654-5106-2/22/$31.00 ©2022 IEEE

II. METHOD

A. Context

The course on Clean Conventional Energies and their technology is part of the results obtained from the creation of the Binational Laboratory for the Intelligent Management of Energy Sustainability and Technological Training, which corresponds to an initiative of the Ministry of Energy, the National Council of Science and Technology and the Tecnológico de Monterrey, whose interest is to generate technology and knowledge on energy in collaboration with other public and private, national and international higher education institutions [10]. In this way, within the Binational Laboratory, a subproject called Open, Interdisciplinary and Collaborative Innovation was generated to train in energy sustainability through MOOCs, which resulted in the creation of 10 MOOC courses on energy sustainability [11]. One of these courses is Clean Conventional Energy and its Technology (ECLT), mentioned above and whose data are analyzed in this study.

Fig. 1. MOOC: Conventional and clean energy and its technology [12]. Source: López (2016).

The ECLT course, which we will call it in this way from now on, was taught in a first phase from February 27 to April 15, 2017, had a duration of seven weeks and was taught through the Mexico X platform, which is an open platform and virtual space for its execution, achieving a terminal efficiency of 17%. Its design included six objectives as stated by [12], which are summarized below:

- To know where, how and when energy manifests itself, through the identification of all the variables and phenomena related to it.
- Identify conventional energies through the study of different energy resources and the technologies used to compare the advantages and disadvantages.

The achievement of these objectives was accompanied by a series of strategies and resources to motivate the participant so that week after week he/she complements his/her activity and the learning process is effective. In this way, various strategies were proposed during the six topics taught in the course. Figure 2 shows the key elements of the instructional model. In addition, in each of the six topics of the course a structure was presented, consisting of: introduction, where a brief explanation of the topic was given; video and self-evaluation, where the expert presented the topic and ended with a self-evaluation; explanation, through html, pdf or infographic format, the content was presented in greater detail, including tables, maps, readings and others. Another element of the structure was the Networking, which corresponded to the discussion generated with 3 discussion options for the participant to choose to participate; in addition, the section To know more was a space dedicated to references to deepen specific topics, review exercises and finally, a graded evaluation at the end of each topic.

Fig. 2. Four key elements of the Instructional Model of the Sub-project: Open, interdisciplinary and collaborative innovation to train in energy sustainability through MOOCs [13]. Source: Villegas et al. (2016).

Based on the above context, the present study focuses on the behavior and results obtained by the students in each of the graded or qualitatively assessed activities in which they participated.

B. Participants

The number of people enrolled in the course ECLT was 6024 people; however, for the present study we took as a sample those participants who completed both the initial and final survey, which were the instruments applied in the course, which is equivalent to 453 participants. The data with information on gender and age are shown in detail in Table 1.

			Gender		
			Female	***Male***	***Total***
		Count	156	297	453
		%	34.437	65.563	100
Age	< - 21	Count	43	56	99
		%	9.492	12.362	21.854
	22 - 24	Count	43	67	110
		%	9.492	14.79	24.282
	25 - 34	Count	37	80	117
		%	8.168	17.66	25.828
	35 - >	Count	33	94	127
		%	7.285	20.751	28.036

Table 1. Age and gender of mooc eclt course participants. Source: Self-made.

C. Objetives

The objective of this study was to analyze the results of the 13 variables proposed in the MOOC course on clean conventional energy and its technology.

The following specific objectives are also derived:

- Individually identify the performance of MOOC course participants.
- Determine the activities in which participants performed the best.
- To identify the variables that influence course failure.

D. Techniques

The data was analyzed by means of Principal Component Analysis (PCA), using the R package "PCAmixdata", which allows processing numerical and categorical data [14], in our case the numerical data corresponds to the grades obtained by the students in the different evaluations, while the categorical information corresponds to the activities that the students carried out in the course. In addition, the studies [15]; [16] regarding the technique used were of reference.

III. RESULTS AND ANALYSIS

As could be evidenced, the MOOC course participants have different profiles in terms of age and gender, according to the information generated for this study, in this way of the 453 participants an analysis of the numerical and categorical variables was performed. The numerical variables correspond to "grade", "Evaluation Topic 1", "Evaluation Topic 2", "Evaluation Topic 3", Evaluation Topic 4", Evaluation Topic 5", Evaluation Topic 6" and "Final Exam", on the other hand, the categorical variables are given by "Exercise", "Practice", "Challenge", "Status", "gender"; total 13 variables used for the study were worked with.

As mentioned above, the technique used is Principal Component Analysis (PCA), this technique is used to reduce the dimensionality of the data, in our case we are working with 13 variables (8 numerical and 5 categorical). Firstly, Figure 3 shows the map of individual components based on the first 2 dimensions obtained from the analysis, it is important to indicate that these dimensions explain almost 63% of the total variance (52.4% in dimension 1 and 9.27% in dimension 2).

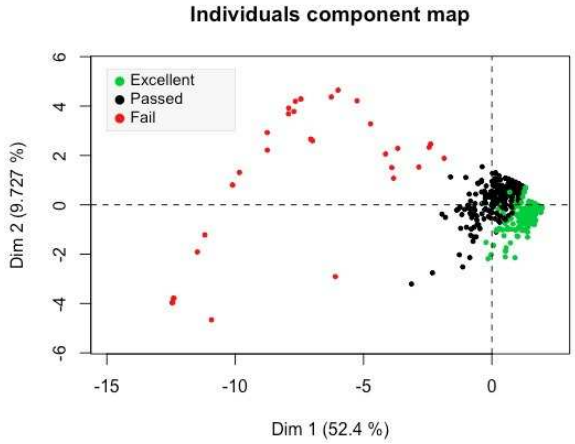

Fig. 3. Map of participant components. Source: Self-made.

Now, from the analysis of academic performance it can be observed that in Figure 3 the 453 participants are presented and with the 2 dimensions three groups are clearly visualized, the green color represents those participants with an excellent grade, those in black are those who passed the course and those in red did not pass the course. The location of the participants who failed the course in the graph, is due to the fact that there are several conditions that influenced them not to pass the course.

Figure 4 below shows the information related to the numerical and categorical variables, in this way distinguishing the former with black color and the latter with red color. The most important variables that have the greatest influence on the success or failure of the course are aligned with dimension 1 and they have a very small angle between them, which indicates that together they are influencing the results of the analysis, additionally we can see by the length of the vector (very close to 1) that their importance is ratified. In this way we can see that the last evaluations (from 4 to 6) and the final evaluation are key in the approval of the course.

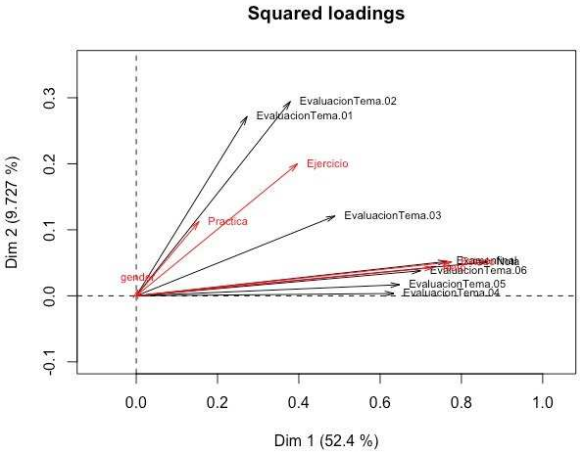

Fig. 4. Importance of the variables used in the análisis. Source: Self-made.

In the case of the other type of activities located on the vertical axis (dimension 2), the vector size is smaller (0.3), so it can be seen that these activities did not have a greater impact on the approval of the participants. In this case, a review of these activities should be made to evaluate the impact, the approach of these activities and to determine what other factors influence these results. A reflection that emerges is that these evaluations are carried out by all participants, but those who pass the course are those who do the activities aligned with dimension 1.

Figure 5 below shows the categorical information, that is, the variables related to gender, practice, challenge and exercise. These activities, according to the data extracted from the course, are qualified as categorical because they are assigned a criterion of "YES performs or DOES NOT perform", obviously except for gender. Now, the reasons why students fail the course are clearly stated here, in this way in the extreme left part are those who do not perform the activity and in the right part are those who perform the activity and pass the course. In fact, this information is related to Figure 3 where we observed the behavior of participants grouped in 3 colors with Excellent, Passed and Failed.

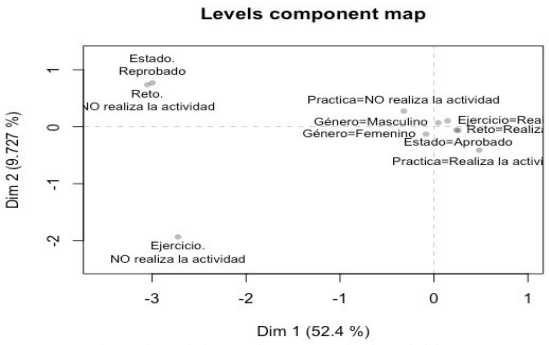

Fig. 5. Grouping of participants by categorical variable. Source: Self-made.

As can be seen within the variables of not performing the activity, the Exercise and the Challenge are the variables that are not performed by the participants who fail the MOOC course, but not activities related to the categorical variables such as practice, whose non-performance does not necessarily imply a behavior that leads to the participant's failure. Undoubtedly, there are still aspects to be delineated since the course, as can be seen, presented several activities, some with similar ratings, others with different weights, but all of them were valuable and corresponded to the objectives of the course and, obviously, to the contents developed.

IV. DISCUSSION

The study of the numerical and categorical variables that are part of the development of a MOOC course on energy sustainability is an area of study that can be analyzed using the Principal Component Analysis (PCA) technique. According to the results in Figure 3, it is shown how the individuals who are the object of study are grouped, in this way it is possible to clearly observe the different groups of participants who are in a condition of approval or reprobation of the course, but it is also highlighted that the technique allows simplifying the variables for a better analysis. In this case, since numerical and categorical information is available, one of the advantages of the technique used is the possibility of combining variables within a single data Table [14]. Based on the fact that we are interested in analyzing in detail the underlying behaviors in the activities developed by the students throughout the course and which have repercussions for the participants' passing or failing status, the technique used constitutes an excellent alternative for this purpose.

A factor of great interest in the analysis of the results of the 13 variables analyzed was to verify those with the greatest impact on the performance of the participants. Figure 4 shows the importance of each of the variables, detecting those with the greatest impact on performance, in this way, without underestimating the importance of some activities over others, the results tend to show that in at least 7 of the 13 variables the participants had better results. These results can be analyzed from different perspectives, for example, it should be considered that in MOOC courses there are different types of students, in this way, for example, there are passive and active participants, the latter being those fully committed to the course [4], this is in line with the study conducted by [5] when in the study conducted identifies that the final modules are those that are overcome more quickly by the participants, as opposed to the initial ones that in terms of time factor demand more from them. Being the main interest to analyze the results by activity, it is important to further analyze variables such as practice and exercise in order to verify how they were designed.

V. CONCLUSION

This article shows the importance of the results of the 13 variables proposed in the MOOC course on clean conventional energy and its technology with the participants who completed both the initial and final survey, since this population is the one we are working with for a broader research. It was evident that three types of participants were clearly distinguished (Excellent, Pass and Fail), taking into account that these are participants who practically remained until the end of the course. The technique used by means of principal component analysis (PCA) made it possible to determine in detail the information by both numerical and categorical variables, which made it possible to have a specific behavior for each activity proposed in the course.

The results are important to the extent that they contribute to consider the findings in the design of future courses where valuable elements arising from data analysis are incorporated, since it was possible to determine those activities that were not of greater influence or interest for the participants at the time of passing the course, therefore, it is important to review the proposed activities and assess the quantity and quality of the same. To conclude, this type of study should be encouraged in research, that is, to promote the use of data techniques to reach a deeper analysis of the information that in many cases is not visible to the naked eye.

ACKNOWLEDGMENT

This research work has been developed within the Doctoral Program in Training in the Knowledge Society of the University of Salamanca, which the doctoral student carries out under the auspices of the Private Technical University of Loja, who is thanked for the support provided. This work is the result of the financing of the project by CONACYT-SENER México) through the project "Binational Laboratory of Intelligent Management of Energy Sustainability and Technological Training" (Ref.266632). The authors acknowledge the financial support of Writing Lab, TecLabs, Tecnológico de Monterrey, México, in the production of this work.

REFERENCES

[1] L. García Aretio, "MOOC: ¿tsunami, revolución o moda pasajera?", RIED Rev. Iberoam. Educ. Distancia, vol. 18, núm. 1, pp. 9–21, 2014.

[2] J. Loizzo, P. A. Ertmer, W. R. Watson, and S. L. Watson, "Adult MOOC learners as self-directed: Perceptions of motivation, success, and completion", Online learn., vol. 21, núm. 2, 2017.

[3] C. Milligan and A. Littlejohn, "Why study on a MOOC? The motives of students and professionals", Int. Rev. Res. Open Distrib. Learn., vol. 18, núm. 2, 2017.

[4] J. Cabero Almenara, "Visiones educativas sobre los MOOC", RIED Rev. Iberoam. Educ. Distancia, vol. 18, núm. 2, pp. 39–60, 2015.

[5] S. Ricart-Casadevall, R. Villar-Navascués, M. Hernández-Hernández, A. Rico-Amorós, and J. Olcina-Cantos, "¿Quién, cómo y cuánto se aprende?: análisis del perfil y el comportamiento de los estudiantes inscritos en un curso mooc sobre riesgos naturales" dentro de Congreso Iberoamericano de Docencia Universitaria. La transformación digital de la universidad Congreso Iberoamericano de Docencia Universitaria. Universidad de La Laguna (España), 2021.

[6] J. C. Aguado Franco, "¿Pueden los MOOC favorecer el aprendizaje y hacer disminuir las tasas de abandono universitario?", *RIED Rev. Iberoam. Educ. Distancia*, vol. 20, núm. 1, p. 125, 2017.

[7] Y. Feng, D. Chen, Z. Zhao, H. Chen, and P. Xi, "The impact of students and TAs' participation on students' academic performance in MOOC", en Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM '15, 2015.

[8] M. del M. Sánchez-Vera and M. P. Prendes-Espinosa, "Beyond objective testing and peer assessment: alternative ways of assessment in MOOCs", RUSC Univ. Knowl. Soc. J., vol. 12, núm. 1, p. 119, 2015.

[9] E. Sánchez Acosta and J. J. Escribano Otero, "Clasificación de medios de evaluación en los MOOC", Edutec Rev. Electrón. Tecnol. Educ., núm. 48, p. a279, 2014.

[10] "Laboratorio Binacional para la Gestión Inteligente de la Sustentabilidad Energética y la Formación Tecnológica", 2017. [En línea]. Disponible en: http://energialab.com/cursos/.

[11] M. S. Ramirez-Montoya et al., "Modelo instruccional de los MOOC de Energía", 2016.

[12] J. López, "Energías convencionales, limpias y su tecnología", [Laboratorio Binacional para la Gestión Inteligente de la Sustentabilidad Energética y la Formación Tecnológica], con financiamiento del Fondo de Sustentabilidad Energética CONACYT-SENER, 2016.

[13] A. Villegas et al., "Modelo Instruccional para los xMOOC de energía Sub-proyecto: Innovación abierta, interdisciplinaria y colaborativa para formar en sustentabilidad energética a través de MOOC", Laboratorio Binacional para la Gestión Inteligente de Sustentabilidad Energética y la Formación Tecnológica. [No publicado], 2016.

[14] M. Chavent, V. Kuentz-Simonet, A. Labenne, and J. Saracco, "Multivariate analysis of mixed data: The R package PCAmixdata", *arXiv [stat.CO]*, 2017.

[15] M. Chavent, V. Kuentz, and J. Saracco, "Rotation orthogonale dans PCAMIX", en 18èmes Rencontres de la Société Francophone de Classification, 2011, p. 4.

[16] J. Saracco and M. Chavent, "Clustering of variables for mixed data", EAS publ. ser., vol. 77, pp. 121–169, 2016.