Indian Research Journal of Extension Education
ISSN: 0972-2181 (Print), 0976-1071 (Online)
NAAS Rating 5.22 (w.e.f January 2021)
journal homepage: https://www.seea.org.in

Study of Learners' Behavior in Massive Open Online Course using Cluster Analysis

G.R.K. Murthy[1], Seema Kujur[2], S. Senthil Vinayagam[3], Yashavanth B.S.[4], Ch. Srinivasa Rao[5], P. S. Pandey[6], Vanita Jain[7] and Indradevi T.[8]

1&3. Pr. Scientist, 4. Scientist, 2&8. Young Professional-II, 5. Director, ICAR- NAARM, Rajendranagar, Hyderabad, Telangana, 6. ADG (Edu.), 7. Pr. Scientist (Edu.Divi.), ICAR, New Delhi

Corresponding author e-mail : lilyseema805@gmail.com

Paper Received on May 29, 2021, Accepted on August 20, 2021 and Published Online on October 01, 2021

ABSTRACT

The purpose of the study was to analysis of the learners' behavior during MOOC and determining the learners' course completion rates. The explorative research design was followed in the study. The sample of the study was 250 learners, which was selected using of sampling technique called stratified simple random sampling. Non-hierarchical cluster analysis was employed to determine learners' behavior in MOOCs and a t-test for unequal variances was used to find a significant difference between learners learning behavior among clusters. MOOCs were administered on the MOODLE platform, targeting learners from diverse domains and organizations. Results found two engagement patterns emerged from this study as participated learners visited the course periodically and certified learners completed the course thoroughly. The learning behavior and pattern of these two clusters were likely to prove an improvement in the quality of learning environment. Findings indicated that there was a significant difference in the fourth-week engagement pattern and quiz marks between two cluster learners. The study underlines the importance of internal and external drivers in higher education as distance, vocational and online education to support the quality of technology-enhanced learning in higher education.

Key words : *Learner's behavior; Massive open online courses; Login pattern; Connectivism theory.*

Massive Open Online Courses (MOOCs) enables the millions of learners worldwide to enhance expertise in a wide domain of subjects *(Ferguson et al.,2015; Chen et al.,2016)*. MOOC ecosystem opens up broader educational opportunities for traditional and non-traditional learning. MOOCs being free and open in nature allows anyone to enroll, access and leads to diversity in motivations and expectations among learners. MOOC platforms are playing a central role in open education. It has facilities to provide an openly shared curriculum and open-ended outcomes *(Clow 2013 and Littlejohn et al.,2016; Zhao et al., 2017)*. It attracts a large number of diverse learners and bids higher quality education provided by domain experts of various subjects to a massive number of participants with very different experiences *(Boroujeni et al., 2017)*. However, MOOC instructors need to understand how the learners interact with lecture videos as well as how they perceive them *(Dowell et al., 2018; Yang & Juan 2018; Li et al., 2015)*.

MOOCs enhances the teaching-learning process through video lessons, quizzes, assignments, and discussion forums. Under MOOCs, learners are required to learn independently and be self-disciplined.

(Dissanayake et al., 2018; Shimony et al., 2015). In the present scenario, online mode of education has been taught as potentially powerful enabling tool for educational change and reform. Using this mode of education helps to increase the access to education at its all level and it has been said that the using e-learning tools in educational settings suggests that the full realization of the potential educational benefits is not automatic, however integration of technologies into teaching-learning process is feasible to the teaching-learning environment *(Malik & Godara, 2020)*. Technology-based interactions in MOOC blogs and forums play a key role in facilitating discussions among learners, instructors, and teaching assistants. It plays a beneficial role in the area of knowledge development, communication, information management and its application in several fields. Also, it has made teaching learning more meaningful, interesting, and learners can construct their own understanding *(Khandave & Shaik, 2020)*. The opportunities offered by MOOCs for cost-effective massification of learning have generated significant interest from governments, higher education institutions (HEIs) and commercial organizations *(Yuan & Powell, 2013)*.

Institutions and educators are rapidly growing their interest in designing and delivering MOOCs for higher education by using popular providers, such as Coursera, MiríadaX, edX, and Future Learn, or adapting open platforms to their infrastructure. This creates a new form of educational provision with occupying a space between formal online courses and informal learning. *(Fassbinder et al., 2017; Walji et al., 2016)*. Unlike regular courses in which learners engage with classroom materials in a structured and monitored manner, and instructors directly observe learners' learning behavior and obtain scheduled feedback, the distance and sheer size of an online course require new approaches to provide learners feedback and instructor guidance. In the context of MOOCs, which provide a low level of support and guidance, the lack of external pressure to make progress and explicit social norms around completion require that learners be highly self-directed to achieve their course goals.

Theoretical frame work : MOOCs provide opportunity for learners from all over the world to connect with those who have common interests. As the MOOCs are delivered on websites focuses on colleges/ institutions and are most often focused on conventional course materials, learning techniques and teaching method. They are usually organized around pre-recorded video lectures, that are posted on the platform, although interactive elements including quizzes and discussion forums are often included.

Siemens (2008) developed an extensive network of connected learners and resources that learners can access and use to design and direct their learning. These connections are established in a biological/neural, conceptual, and social/external context. Connectivism and related knowledge were claimed that educators had the role of facilitator or were totally absent from the learning process. All course material was accessible via feeds from the RSS (Really Simple Syndicate), and learners could participate in Moodle, blog posts, and synchronous online meetings with their choice of resources as threaded discussion. The learning process in connectivism occurs as the learner feeds their knowledge through making connections with the collective knowledge of the community *(Anderson & Dron, 2011)*. The theory of connectivism indicates that each individual is responsible for their own learning. In the present study, apart from all learning theories, connectivism theory was used to assess the learning behavior of learners in terms of login activities, watching a video lecture, interaction with co-learners in discussion forums, assignment submission, and attempting quizzes etc. which help in assessing the learners' behavior with their interest and seriousness to the course.

Present days, MOOCs are being used in very creative and effective way in education. Apart from making the teaching-learning process more interesting, the MOOCs offer opportunity to the teacher to teach how to gain information, which not only make learning attractive and interesting, but also increase retention of the learners *(Malik et al., 2021)*. Despite the high level of interest and the rapid development of MOOCs, there is still a lack of understanding of learners' behaviors and how learners engage in such courses *(Anderson et al., 2014)*. The high dropout rate for most MOOCs is the biggest challenge facing online education providers. This paper explores learners' learning behavior and their completion rates. Learning behavior in the context of MOOCs refers to how, when, and in what order learners watch videos and process other MOOC resources; and when and in what order they make quizzes and assignments. The present study attempts to classify MOOC learners into different groups based on their login patterns, their engagement in assignments, discussions, and quizzes. This paper aims to explain learners learning behavior, with the purpose of finding indicators for improving the quality of teaching-learning process. It attempts at understanding the learning behavior using a cluster approach by considering intrinsic parameters that define their seriousness about learning. In the above context, the present study was conducted to :

i. Analyze the completion rate behavior among the learners
ii. Segregate the learners based on their learning behavior using cluster analysis

METHODOLOGY

The sampling framework comprised representative learners whoever was participated in MOOCs from the various subject domain. The learners in the sampling framework included learners from various subject domains such as agriculture, veterinary, agricultural engineering, agribusiness management, and education. Stratified simple random sampling with proportional allocation was carried out for getting a specific sample size for the study. A stratified random sampling with proportional allocation consists of dividing the entire population into homogeneous groups which is called strata. A random sample from each stratum was taken in a number proportional to the stratum's size when compared to the population. These subsets of the strata were then pooled to form a random sample and was taken for further study.

Table 1. Sampling framework

Learners from the various subject domain	Sample size Total	Selected
Agriculture	798	167
Veterinary	218	46
Agribusiness Management &Agri. Engg.	90	19
Education	89	18
Total	1195	250

The study banks on the large volume of data generated during the offer of a MOOC on "Dynamics of Teaching-Learning" aimed at having skills of teaching for teachers and aspirant teachers. Explorative research design technique was employed to analyze the empirical data concerning learners' participation in course-related activities like discussion forums, assignments, quizzes, and accessing the networks like videos and documents. In each of the weekly assignments, learners were instructed to submit an assignment on the weekly delivered course content and participate in the discussion forum. Learners had the provision for obtaining a certificate of completion or participation based on the extent of their participation in course activities and meeting certain numerical criteria in quizzes and participation in the discussion forum and assignment submission. MOOCs brings a various cross-section of learners on to the same learning platform. They represent a wider spectrum of age, gender, region, cultural backdrop, working status, and educational background. The MOOC log files that reflect all the activities of learners represent a huge data mountain that hides various valuable nuggets of information that can provide a road map to design MOOCs in the future. The learners under the MOOC platforms are diverse in nature. Understanding the behavior of these learners by identifying different types of learners will help in designing better courses. In this study, the learners are grouped using a cluster analysis technique.

Cluster analysis : Cluster analysis is a multivariate statistical technique that groups subjects such that the subjects in the same group are more similar to each other than the subjects in other groups. It allows us to better understand how a sample might comprise distinct subgroups, given a set of variables. Cluster analysis has been successfully used to identify learner groups *(Douglas et al., 2016)* and to develop profiles that are grounded in learner activities that reflect learning behavior in the MOOC context *(Antonenko et al., 2012)*.

Cluster analysis is employed in the present study to decide the number of clusters, interpret the profile clusters and finally, assessing the significant difference among clusters attributes in terms of their activities in the course. It determines learners' learning behavior by grouping a set of MOOC data. The data consisted of weekly login, days login, total login, average login, assignment, discussion, and quiz marks. MOOC data was grouped using cluster analysis in such a way that some data in the same group (called a cluster) were more similar with some data to each other than to those in other groups or clusters. After getting newly formed clusters, learners' behavior was distinguished from a cluster and it was further used for the comparative analysis to find a significant difference among attributes of newly formed clusters.

A cluster analysis can be hierarchical or non-hierarchical. Hierarchical clustering involves creating clusters that have a predetermined ordering from top to bottom. *Tseng et al., (2016)* used the hierarchical clustering methods to determine the number of clusters and to classify different clusters of learners in MOOCs. The non-hierarchical clustering method aims to find groups of subjects that maximizes or minimizes some evaluating criterion. There are different types of clustering algorithms such as Centroid based clustering, Density-based clustering, Distribution clustering, and Hierarchical clustering. The present study employs the Non-hierarchical clustering technique using the Partitioning Around Medoids (PAM) algorithm to identify the learners' groups of similar subjects or attributes from the MOOC data set. The PAM algorithm takes the input parameter, k, and partitions a set of n objects into the k cluster so that the resulting intra-cluster similarity is the high but inter-cluster similarity is low.

Partitioning around medoids (PAM) : PAM algorithm is also known as k-medoids clustering. A medoid can be defined as the subject of a cluster whose average dissimilarity to all the objects in the cluster is minimal, i.e., it is the most centrally located point in the cluster. K-Medoid is a classical partitioning technique of clustering that clusters the data set of n subjects into k clusters known a priori. The PAM algorithm partitions the dataset of n subjects into k clusters, where both the dataset and the number k is an input of the algorithm. This algorithm works with a matrix of dissimilarity, whose goal is to minimize the overall dissimilarity between the represents of each cluster and its members. k-medoids has two advantages as it presents no constraints on attributes (data) types and the preference of medoids is dictated by the location of a predominant i.e. major fraction of points inside a cluster and, therefore, it is lesser sensitive to the presence of outliers and noise. The PAM algorithm was used following formula:

$$Z = \sum_{i=1}^{k} \sum \|x - m_i\|^2$$

Where, Z is the sum of absolute error for all items in the MOOC data set (weekly login, days login, total login, average login, assignment, discussion, and quiz marks); x is the data point in the space representing a given object or data item; m_i is the medoids of cluster Ci(both x and m_i are multidimensional); Ci is the centroid/mean of cluster Ck. Formula express that, for each object in each cluster, the distance from the object to its cluster center is squared, and the distance is summed. This criterion tries to make the subsequent k clusters as compact and separate as possible.

Distance measure : Several distance measures are used to find the similarity of the subjects and consequently group them into clusters. Distance measures like Euclidean distance, Manhattan distance, and Minkowski distance are useful when the variable is continuous whereas the Dice coefficient can be used for categorical variables. Since this study consisted of both categorical as well as continuous variables, Gower's distance (*Gower, 1971*) is used. Gower's distance measures the distance between two entities whose attributes have a mix of categorical and numerical values. In the Gower distance method, for each variable type, a particular distance metric that works well for that type is used and scaled to fall between 0 and 1. Then, a linear combination using user-specified weights (most simply an average) is calculated to create the final distance matrix. The Gower distance method can use for each type of data such as quantitative, ordinal and nominal.

$$d(i,j) = \frac{1}{p} \sum_{i=1}^{p} d_{ij}^{(f)}$$

Where, $(d\_ij\,\hat{}\,f)$ = partial dissimilarities computation which depends on the type of the variable being evaluated for the learning behavior.

Number of clusters : The PAM algorithm requires the number of clusters to be identified before running the algorithm. For this purpose, the Silhouette width is used (*Rousseeuw, 1987*). The silhouette width compares the average distance to subjects in the same cluster with the average distance to subjects in other clusters. The silhouette width is calculated for the different numbers of clusters and the number of a cluster for which the silhouette width is maximum will be considered as the optimum number. For cluster analysis variables like gender, subject domain, weekly login patterns, assignment, discussion, the total number of logins, the number of not login, quiz marks, participated and certified learners based on provision for the certificate were used. For the present study, cluster analysis used an analysis of variance approach to evaluate the distances between clusters. Once identified the number of clusters, the method was further used to analyze learning behaviors in different clusters of MOOC learners.

Two sample t-test assuming unequal variances was used to determine the significant difference of listed variables among clusters and provided inference of the test. Hypothesis of the study are as follows:

Table 2. The hypothesis of the study

Variables	Weekly login pattern, Discussion forum, Quiz marks,
Null hypothesis (H_0)	There is no significant difference in weekly login pattern, discussion forum, and quiz marks among clusters
Alternative hypothesis (H_a)	There is a significant difference in weekly login pattern, discussion forum, and quiz marks among clusters

RESULTS AND DISCUSSION

The employed cluster analysis investigated to indicate aspects of low completion rates are influenced by the activity pattern of learners. These results quite a meaningful picture of what drives a learner to participate in MOOCs. This section explains about certificate earners, completion rate and learning behavior of the learners by using non-hierarchical clustering. Further explains about significant differences among clusters.

Certificate earners and completion rates : The completion rate was selected as the variable of success because it is one of the most common metrics used in MOOC research (*He et al., 2015*). In this study, the completion rate was considered based on learners who completed their course through MOOCs. Out of 1195 learners, 482 have got a certificate for completion and participation. 40.3 per cent was the completion rate of the MOOC on "Dynamics of Teaching-Learning". Learners got a certificate for participation because they participated in the course but could not fulfill all requirements of the MOOC. Learners who got a certificate for completion were those who have participated and completed all requirements of MOOC in all aspects.

Determined the number of K (Clusters) : The PAM algorithm requires that the number of clusters to be predetermined. Hence, different values from 2 to 10 for the number of clusters were tried and the silhouette width for each value is obtained (Fig.1). Since, the silhouette width for k=2 was found to be maximum, the optimum value for k is considered as 2. The highest and lowest silhouette width was approximately 0.38, at k = 2, and 0.27, at k = 6 respectively. The highest silhouette width was taken which infer the cluster size of the study.

Table 3. General description of newly formed clusters

Variable	Component	Cluster 1 (n=100)	Cluster2 (n=150)
Gender	Female	23	39
	Male	75	113
Type of Learner	Certified	2	151
	Participated	96	1
Education	Graduates	2	2
	Masters	38	60
	Doctorate	58	90
Domain	Agribusiness Management	1	7
	Agriculture	67	100
	Education	6	12
	Engineering	6	5
	Veterinary	18	28

Learners demographic information is presented in Table 3 which indicated that the majority of the learners under cluster one were participated learners, male, having doctorate degree, from the agricultural subject domain. The majority of the learners under cluster two were certified learners, male, having doctorate degree, from the agricultural subject domain. It may be concluded that the majority of the participated and certified learners (male) having a doctoral degree and from the agriculture domain in both clusters.

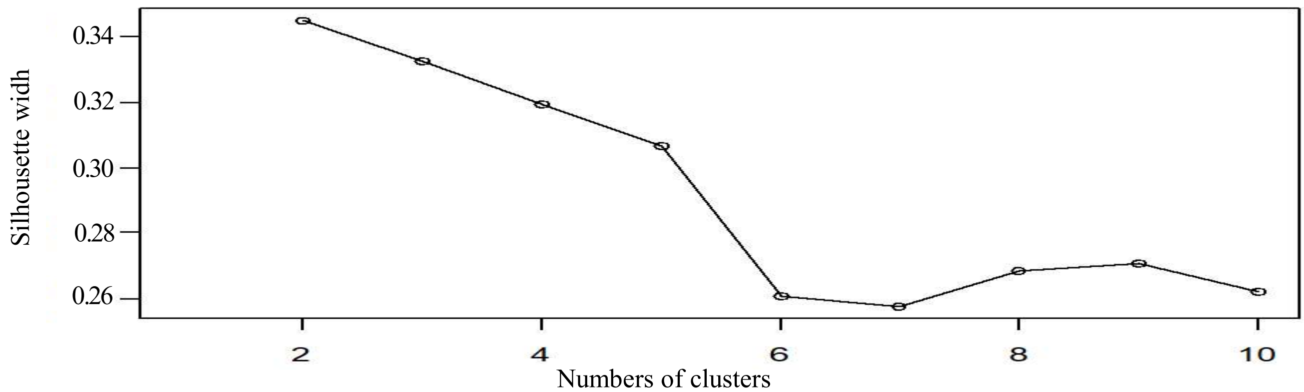

Figure 1. Silhouette widths for different values of k

Table 4. Description of clusters

Variables	Cluster 1 (n=100)				Cluster 2 (n=150)			
	Mean	Median	Max	Min	Mean	Median	Max	Min
Logged in week 1	19.42	4.5	110	0	54.84	37.5	491	0
Logged in week 2	16.19	0	125	0	63.93	51	279	0
Logged in week 3	20.88	0	163	0	74.7	55	389	0
Logged in week 4	76.55	59	340	0	113.6	93.5	580	14
Days Login	6.01	5	24	1	13.42	13	27	4
Total Login	133	118.5	482	3	300.9	269	927	46
Average Login	24.66	21.55	89.7	3	22.48	22.35	46.9	8.7
Assignment	1.306	1	4	0	3.645	4	4	1
Discussion	1.551	0	29	0	7.553	4	49	0
Quiz marks	62.26	64.5	86	0	68.05	71	90	0

As given in Table 4, clusters characteristics of various considered variables, showed that the majority of the participated learners were falling under cluster one. The participated learners were those who participated in the course but did not fulfill all requirements of MOOCs and got a certificate for participation. As per the weekly mean value, it calls that the participated learners are not active in the 1st, 2nd & 3rd weeks of the course. In the last week of the course, learners were more active as compared to previous weeks. Participated learners showed the lower mean value of their total days of login, total login, assignment, discussion, and quiz marks. However, their average login was shown that these learners were also interested to get engaged in the course.

Results also asserted about certified learners' characteristics who have fallen under cluster two. The certified learners' were those who have completed MOOC requirements in all aspects and got a certificate for completion. As shown in the Table 4, the high mean values of various weekly logins revealed that certified learners were very interested and active in the course as they were frequently logged in as per weekly basis. The result also revealed that certified learners' daily login was very high as per the mean value, highly total login. They were interested in the forum activities such as assignment and discussion which was mandatory to have certified from the course. Certified learners were occupied with high marks in the quiz.

Fig. 2 reveals about scatter plot distribution of two clusters and it shows the newly formed clusters. Two of the groups which are blue and dark peach color are somewhat overlapping. A rough description is given for each cluster that describes learners' way of engagement as: cluster (1) or "participated learners", with the dark peach dots contains 100 learners (40%). This learners' group shown the less engagement or activity as compared to the other learners' group in the MOOC. The attrition rate of this group was high as they did not complete the course properly. The cluster

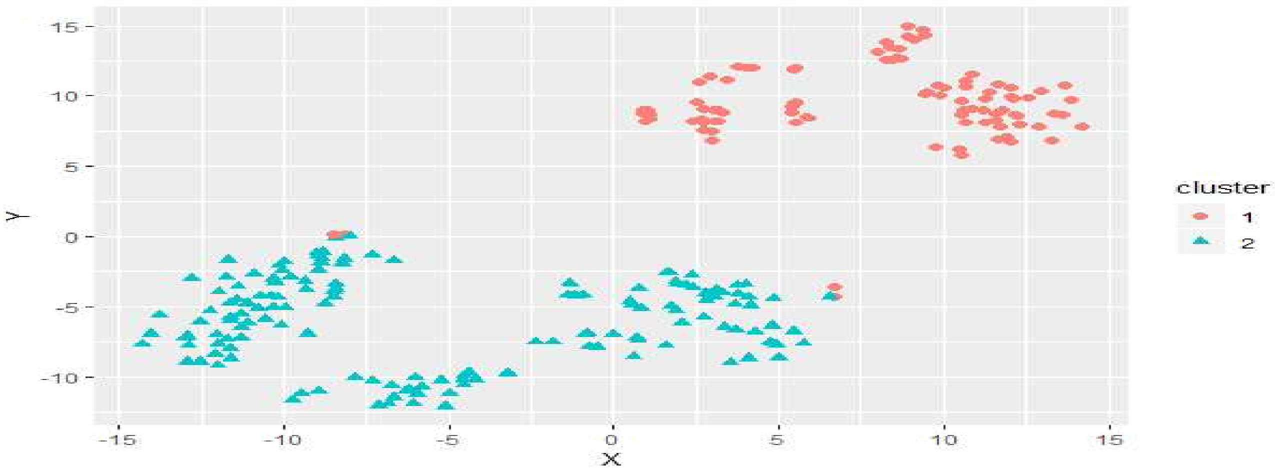

Figure 2. Distribution of subjects into two clusters

Table 5. Statistics of two-sample t-test for unequal variances for learner's weekly login pattern, assignments, discussion and quiz marks among clusters

Variables	Cluster 1M(V)	Cluster 2M(V)	t -value	ρ-value
Logged in Week 1	19.97(732.39)	54.94(4391.36)	2.6115	2.611^{NS}
Logged in Week 2	16.9(866.17)	64.09(3159.44)	-8.656	7.593^{NS}
Logged in Week 3	23.74(1495.10)	73.50(4017.59)	-7.705	3.241^{NS}
Logged in Week 4	79.88(4704.24)	111.86(6050.69)	-3.421	0.0007**
Average login	24.98(189.83)	22.22(72.09)	1.791	0.075^{NS}
Assignment	1.39(1.23)	3.62(0.38)	-18.288	4.828^{NS}
Discussion	1.88(31.58)	7.41(96.95)	-5.641	4.686^{NS}
Quiz Marks	62.53(293.96)	67.94(202.38)	-2.615	0.009**

**1% level of significance, NS- Non- Significant, M-Mean, V-Variance; Sample Size- Cluster 1 (n= 100), Cluster 2 (n= 150)

(2) or "certified learners", with the blue dots contains 150 learners (60%). Learners in this group were fully engaged and completed the requirements of the course successfully. The certification ratio of this group was high. This group was highly engaged in the discussion forums, accessing video lectures, submitting assignments and participated in the reading.

The second cluster basically comprises those who were completed MOOC requirements in all aspects. The result inferred that they are very active in online discussion and better in quiz performance in comparison with cluster one. It has seen from the result that the discussion contributed by cluster two is rich in quality and aided in collaborative learning this is a consonance with the connectivism that MOOCs provide learners from all over the world with the opportunity to connect with others who share similar interests. All educators should be familiar with the various learning theories and understand how they can enhance the learning experience of learners. A MOOC no matter how cleverly designed will only empower learners who are motivated to engage and learn, similar to traditional teaching environments. Connectivists state that learning is not merely the transfer of knowledge from the teacher to the learner and does not take place in a single environment, instead they state that knowledge is transformed and transferred through the interactions of people, especially in a web environment *(Kop, 2011)* and it support the findings of the present study. Connectivism presented a model of learning that acknowledges the tectonic shifts in society where learning is no longer an internal, individualistic activity.

It can be concluded from Table 5 that the calculated *P*-value (2.611) is more than the alpha value ($\leq$0.05) at 5% level of significance. Therefore, it can be concluded that there was no significant difference between the mean level of login pattern in the first week among cluster learners. Likewise, it can be concluded that there was no significant difference between the mean level of login pattern in the second week among cluster learners since the calculated r-value (7.593) is more than the alpha value ($\leq$0.05) at 5% level of significance. The result indicated that there was no significant difference between the mean level of login pattern in the third week among cluster learners because the calculated r-value (3.241) is more than the alpha value ($\leq$0.05) at 5% level of significance. Test for unequal variances revealed that the calculated r-value (0.007) was very less than the alpha value ($\leq$0.01) at 1% level of significance, which suggest to reject the null hypothesis. Therefore, it can be concluded that there was a significant difference between the means level of login pattern in the fourth week among cluster learners. Learners from cluster two were performing better in terms of login patterns during the course.

The average login of the learners has been compared by using a t-test for unequal variances which indicated that the calculated r-value (0.075) is more than the alpha value ($\leq$0.05) at 5 per cent level of significance. Therefore, it could be concluded that there was no significant difference between the mean level of the average login pattern of clusters one and two learners. The calculated r-value (4.828) is more than the alpha value ($\leq$0.05) at 5 per cent level of significance, which reveals that, there was no significant difference between the mean level of assignment submission of cluster one and two learners which means learners were submitting assignments during the course because it was mandatory for the certification.

There was no significant difference between the mean level of participation in the discussion forum among cluster learners which has mentioned in above Table 5. The calculated r-value (4.686) is more than the alpha value ($\leq$0.05) at 5 per cent level of significance. So, it can be concluded that cluster one learners were non-significant with cluster two learners in case of participation in the discussion forum during the course. Hypothesis test results revealed that the calculated r-value (0.009) is very low than the alpha value ($\leq$ 0.01) at 1 per cent level of significance, which suggest to reject the null hypothesis. So, there was a significant difference between the mean level of quiz marks of cluster one and two learners. So, it can be concluded that the learners' performance level on the course affects quiz marks as cluster two learners were got better quiz marks as compared to the cluster one learners.

CONCLUSION

This study described an exploratory analysis of MOOCs in the Moodle platform. The experiment result shows that there are several existing units of learners in the MOOC environment. Each unit of learners forms a cluster and has similar characteristics in the study. Two engagement patterns emerged from this study as participated learners visited the course periodically and certified learners completed the course thoroughly. The learning behavior and pattern of these two clusters are likely to prove an improvement in the quality of learning environment. The majority of the learners were male in both clusters, with a doctorate degree, from the agriculture domain. Findings indicated that there is a significant difference in fourth-week engagement pattern and quiz marks between two cluster learners. Those learners who frequently communicated, discussed, shared, and collaborated with others in the forum expressed a better learning outcome. Those learners who posted often in the discussion forums and assignments would have a higher rate of passing the course.

The results of the study demonstrate the utility of the framework for learning behavior patterns and the connectivism theory of learning is useful for all subject domain learners. Connectivism provides insight into learning skills and tasks needed for learners to flourish in a digital era. Analyzing learners' behavior and pattern of engagement help to suggest research and design directions for future online courses.

REFERENCES

Anderson, A.; Huttenlocher, D.; Kleinberg, J. and Leskovec, J. (2014). Engaging with massive online courses. In www'14, proceedings of the 23rd international conference on world wide web (pp. 687–698).

Anderson, T. and Dron, J. (2011). Three generations of distance education pedagogy. *Intl. Review of Res. in Open and Distributed Learning*, **12** (3) : 80-97.

Antonenko, P.; Toy, S., and Niederhauser, D.S. (2012). Using cluster analysis for data mining in educational technology research. *Edu. Tech. Res. and Devel.*, **60** : 383–398.

Boroujeni, M. S.; Hecking, T.; Hoppe, H. U. and Dillenbourg, P. (2017). Dynamics of MOOC discussion forums. In proceedings of the seventh international learning analytics & knowledge conference (pp. 128-137). ACM.

Chen, G.; Davis, D.; Lin, J.; Hauff, C. and Houben, G. J. (2016). Beyond the MOOC platform: gaining insights about learners from the social web. In proceedings of the 8th ACM conference on web science (pp. 15-24). ACM.

Clow, D. (2013). MOOCs and the funnel of participation. In proceedings of the third international conference on learning analytics and knowledge (pp. 185-189). ACM.

Dissanayake, D.; Perera, T.; Elladeniya, C.; Dissanayake, K.; Herath, S. and Perera, I. (2018). Identifying the learning style of students in MOOCs using video interactions. *Intl. J. Info. and Edu. Tech.*, **8** (3).

Douglas, K. A.; Bermel, P.; Alam, M. M. and Madhavan, K. (2016). Big data characterization of learner behaviour in a highly technical MOOC engineering course. *J. of Learning Analytics*, **3** (3) : 170-192.

Fassbinder, A.; Barbosa, E. F. and Magoulas, G. (2017). Developing and educational design pattern language for MOOCs. In Brazilian symposium on computers in education (Simpósio Brasileiro de Informáticana Educação-SBIE) (Vol. 28, No. 1, p. 456).

Ferguson, R.; Clow, D.; Beale, R.; Cooper, A. J.; Morris, N.; Bayne, S. and Woodgate, A. (2015). Moving through MOOCS: pedagogy, learning design and patterns of engagement. In design for teaching and learning in a networked world (pp. 70-84). Springer, Cham.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, **27**(4) : 859.

He, J.; Bailey, J.; Rubinstein, B.I.P. and Zhang, R. (2015). Identifying at-risk students in massive open online courses. In *AAAI 2015* (pp. 1749–1755). Austin, Texas: AAAI Press.

Khandave, S. and Shaik, M.S.R. (2020). Knowledge and awareness of post graduate students of college of agriculture, Pune about information technology tools. *J. of Global Commu.*, **13** (1) : 37-40.

Kop, R. (2011). The challenges to connectivism learning on open online networks: Learning experiences during a massive open online course. *Intl. Review of Res. in Open and Distributed Learning*, **12** (3) : 19-38.

Li, N.; Kidziñski, L.; Jermann, P. and Dillenbourg, P. (2015). MOOC video interaction patterns: What do they tell us? In *design for teaching and learning in a networked world* (pp. 197-210). Springer, Cham.

Littlejohn, A.; Hood, N.; Milligan, C. and Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Edu.*, **29** : 40-48.

Malik, A.K. and Godara, A.K. (2020). Information and communication technologies (ICTs) Use by the Students of CCSHAU, Hisar. *Indian Res. J. Ext. Edu.*, **20** (4) : 14-19.

Malik, A. K., Godara, A. K., and Yadav, V. P. S. (2021). Awareness, access and purpose of using information and communication technologies (ICTs) by the students of CCSHAU, Hisar, *Indian Res. J. of Ext. Edu.*, **21** (2&3) : 112-115.

O'Donnell, E.; Lawless, S.; Sharp, M. and O'Donnell, L. (2015). Learning theories: e-Pedagogical strategies for Massive Open Online Courses (MOOCs) in higher education. In *Macro-level learning through massive open online courses (MOOCs): Strategies and predictions for the future* (pp. 92-118).

Rousseeuw, Peter. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics*, **20** : 53–65.

Shimony, U.; Cohen, A.; Nachmias, R. and Soffer, T. (2015). The online learners in MOOC social environments. In *proceedings of the European distance and e-learning network 2015 annual conference*.

Siemens, G. (2008). What is the unique idea in connectivism? http://www.connectivism.ca/? p=116 [viewed 31-10-2009]

Tseng, S.F.; Tsao, Y.W.; Yu, L.C.; Chan, C.L. and Lai, K. R. (2016). Who will pass? Analysing learner behaviours in MOOCs. *Res. and Practice in Tech. Enhanced Learn.*, **11**(1) : 8.

Walji, S.; Deacon, A.; Small, J. and Czerniewicz, L. (2016). Learning through engagement: MOOCs as an emergent form of provision. *Distance Edu.*, **37** (2) : 208-223.

Yang, S. and Juan, L. (2018). Research on data mining of learning behaviours of college students on MOOC platform. In MATEC web of conferences (Vol. 246, p. 03025). EDP Sciences.

Yuan, L. and Powell, S. (2013). MOOCs and disruptive innovation: Implications for higher education. *e-Learning Papers, In-depth*, **33** (2) : 1-7.

Zhao, Y.; Davis, D.; Chen, G.; Lofi, C.; Hauff, C. and Houben, G. J. (2017). Certificate achievement unlocked: How does MOOC learners' behaviour change? In adjunct publication of the 25th conference on user modelling, adaptation and personalization (pp. 83-88). ACM.

• • • • •