THE ONLINE LEARNERS IN MOOC SOCIAL ENVIRONMENTS

Udi Shimony, Anat Cohen, Rafi Nachmias, Tal Soffer, Tel Aviv University, Israel

Background

Massive Open Online Courses (MOOCs) are the latest revolution in online teaching and learning. The world's leading universities, such as Stanford, Harvard, and MIT are offering MOOCs to the general public, worldwide, without any preconditions and free of charge (Johnson et al., 2013; Allen & Seaman, 2014; Adams & Williams, 2013; Stewart, 2013). These MOOCs are offered by a variety of initiatives, such as Coursera, Udemy, MITx, edX, Udacity, and are taught by professors around the world, in various fields. They allow for flexible learning at any time and any place, integrating a variety of tasks into the course structure. A major emphasis is put on selecting suitable course subjects, instructors, high quality video production, and a friendly interface. These MOOCs may alter the conception of education and create a culture of collaborative social learning (Brinton et al., 2013; Kissinger & Bennett, 2014) combined with peer assessment (Balfour, 2013; Piech et al., 2013). This type of course development enables the teaching of MOOCs.

The literature emphasizes that MOOCs are based on active student engagement in accordance with learning aims and objectives as well as their early knowledge and skills (McAuley, Stewart, Siemens & Cormier, 2010). In online learning, Self-Regulated Learning (SRL) that is characterized by the ability to initiate learning and adapt to new learning methods (Bidjerano & Dai, 2007) is crucial. The learner is required to have the ability to learn independently and be self-disciplined (Komarraju, Karau, Schmeck & Avdic, 2011; Mackness et al., 2013). In addition, many researchers and educators (those who support the theory of connectivism) claim that most of the significant learning in online courses takes place through the sharing of information, and interpersonal interaction (Cormier, 2008; Downes, 2007; Bell, 2011; Siemens, 2014), which may occur in the forums offered to learners. Thus, social learning is a central theme of online courses and especially of MOOCs. MOOCs are based on technologies that enable students to learn independently in cooperation with other students. The main channel for this interpersonal interaction is the course website forums, in which the learning community can manage interactions on issues related to the course (but not necessarily). Hence, this study will examine the extent of forum usage in MOOCs and the patterns of students' activity using Educational Data Mining (EDM) and learning analytics methods to analyze the data accumulated in Coursera log-files. The objective is to produce significant information regarding students' activity in MOOC forums.

The presented study is part of a larger research that aims to characterize and promote effective online teaching and learning processes as well as assessment. In light of the claim that meaningful learning takes place through interpersonal interaction in discussion groups (forums) offered to learners, the fact that the percentage of participants in these discussion groups is lower than expected is not fully explained (Brinton et al., 2013). This study focuses on large-scale discussion groups operated in connection with one MOOC offered by Tel Aviv University through Coursera. MOOC forums were chosen due to the wide range of activity of thousands of students who registered for this course, and the findings of this study will contribute to the developing body of knowledge on learner communities in other online learning environments as well as academic fully online courses.

The aim of this study is to explore the students' participation patterns in forums and characterize them as well as to understand the factors that correlate with the students' participation in these forums. Furthermore, the correlation between performance on the final test and participation level in the forums was tested, too. Accordingly, the study questions were:

1. What is the volume of students' activity?
2. What is the volume of students' weekly activity?
3. What are the student participation patterns reflected in the forums?
4. What are the factors which may correlate to the level of participation?
5. Is there a correlation between forum participation and passing the course?

Methodology

The presented study is based on an analysis of empirical data concerning students' activity in 652 forums, retrieved in 2014 from Coursera log-files. These forums were created in connection to one MOOC that was developed by Tel Aviv University. 27,322 students from all over the world were registered for this MOOC. Coursera automatically accumulates a vast amount of data regarding students' activity in its server web logs. At the end of the course, the course log-files were obtained through web mining techniques. This enabled the retrieval of data regarding hidden traces of students' activity, which is continuously left on the course website.

The study was conducted in three stages:

1. **Data organization** – Using data mining techniques, hundreds of thousands of web log-file records regarding the 27,322 students' activity were extracted from Coursera web log-files and organized into two data-files. The first data-file contained variables regarding each student's activity in a specific forum (each row represented a different action). In this file, each student action was represented by an identification code (unique to each student), which allowed for simultaneous individual and **anonymous** global monitoring (due to privacy and ethical issues); the date and time of the activity; type (e.g., posting in forum, viewing posts/comments); and content. The second data-

file contained summative variables and nominal variables that characterized the learner (each row represented one student). The summative variables calculated activity intensity and views, and their values addressed the timespan of the entire course as well as each week separately. For example, data might have included the number of student posts, number of student comments (reply messages), number of posts/comments in relation to other students, number of evaluation points, and number of tags he/she used as well as the device he/she used, geographic location, etc. Notably, we were committed to protecting student privacy. Student names and e-mail addresses are regarded as personally identifiable information (PII); they were handled very carefully and were not included in research data exports.

2. **Identifying patterns of students' activity** – The second stage focused on analyzing the web log-file records regarding the students' activity in 652 forums. Learning analytics of these log-file records were conducted to identify patterns in students' activity. The raw data was analyzed in stages. In the first stage the posts and the messages were processed in order to perform an initial qualitative analysis. Then, the course meta-data and relevant key student profile parameters were incorporated to generate insights based on unseen correlations. Several variables were calculated in order to characterize the students' activity in the forums such as type of participation (e.g. active: post and/or response, passive: just watching, lack of participation); scope and extent/intensity of participation (e.g. number of posts, depth of discourse – responses for discussion); giving/receiving points for messages; number of participating days; using tags; and characteristics of the messages/texts. In addition, types of devices and operating systems used by students (mobility characteristics) were included variables that were calculated.

3. **Correlation analyses** were conducted among variables that characterized the activity in the forum. In addition, the correlation between these variables of participation, finishing the course successfully, and students' location were tested.

The field and the population

The field of the study included 652 forums which were offered to 27,322 MOOC learners from around the world. These forums were operated throughout the semester. The learners' distribution by geographic location is presented in Table 1.

Table 1: Learner distribution by geographic location

Countries	Students
United States	23,427
Europe	2,223
Asia	1,287
Africa	166
Australia	143
Pacific	34
UTC	30
Atlantic	5
India	4
Antarctica	3
Total	**27,322**

Research tools and data analysis methods

The following data mining tools (quantitative and textual) were used in this study: a log-file information retrieving tool; two data-files for presenting the data in a relational database to be cultivated; and statistical tools.

Quantitative and Qualitative methods were used to achieve the goals of the study. Data regarding the students' activity in the forums was organized in a relational database. This initial step helped in structuring the data and identifying relevant relationships among all the key parameters. Also, the structural data enabled us to generate a matrix of variables to use in examining the correlation between the various student activity variables. Analysis was used for clustering and identifying the patterns of student activity/participation based on defined criteria. In addition, qualitative text analysis was generated on the posts and comments.

Preliminary Findings

Volume of students' activity

652 forums were initiated on a variety of topics. Of the 27,322 students registered for the course, approximately 4,500 registered to use the course's forums during the course, and only 1,257 of them were active in the forums throughout the entire course. The following analyses were conducted on these 1,257 learners. Table 2 presents preliminary data regarding the volume of activity in the forums with respect to: number of forum views; number of posts (messages written on the first level), number of comments (responses to posts – second level, and up); number of evaluation points given by the students; and number of tags used.

Table 2: Volume of forum activity

Activity Type	Volume
Number of forum views	103,848
Number of posts (first level messages)	4,949
Number of comments (second level messages, and up)	4,598
Number of evaluation points given to the messages	4,537
Number of tags used	491

Volume of participation in relation to student's geographic location

The MOOC was attractive to a large number of participants from different countries, especially from the United States (86%), Europe (8.1%) and Asia (4.7%). Table 3 shows the percentage of participants from each country, and it is interesting to note that a similar percentage of students from each country participated in the forums.

Table 3: Participants' geographical location

Country	Participants	Percentages
United States	1011	4.32
Europe	139	6.25
Asia	85	6.60
Australia	13	9.09
Africa	6	3.61
Pacific	2	5.88
UTC	1	3.33
India	0	0
Grand Total	**1,257**	**39.09**

The overall forum activity conducted by those countries, who "participated" in it, was interesting. The ratio between the volumes of views versus posts/comments is very prominent. Australia is unique in that respect; the students played a balanced role in both writing and viewing during the course (Figure 1).
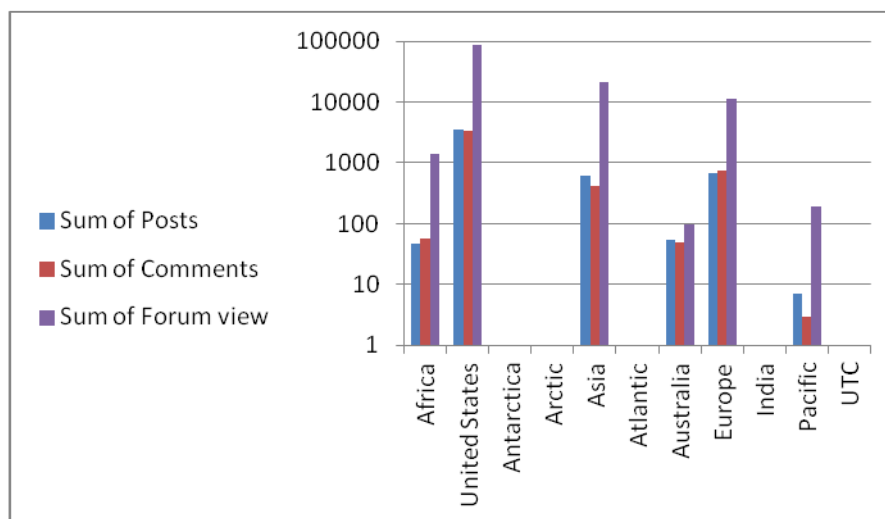

Figure 1. Students' activity by country

Volume of students' weekly activity

The analysis of the volume of students' weekly activity was only conducted on the 1,257 learners who were active in the forums throughout the entire course. Students' activity was explored using three key variables: number of posts, number of comments, and number of views. During the course, the same downward trends appeared regarding comments, posts, and views; though the volume of views throughout the course was, by far, much higher than the number of posts and comments. In the first week, the amount of views was 22.5 times higher than the amount of posts. On a weekly basis, posts and comments shared the same

average. Interestingly, weeks 13 and 17 were unique in that their number of comments was fairly higher than their number of posts.
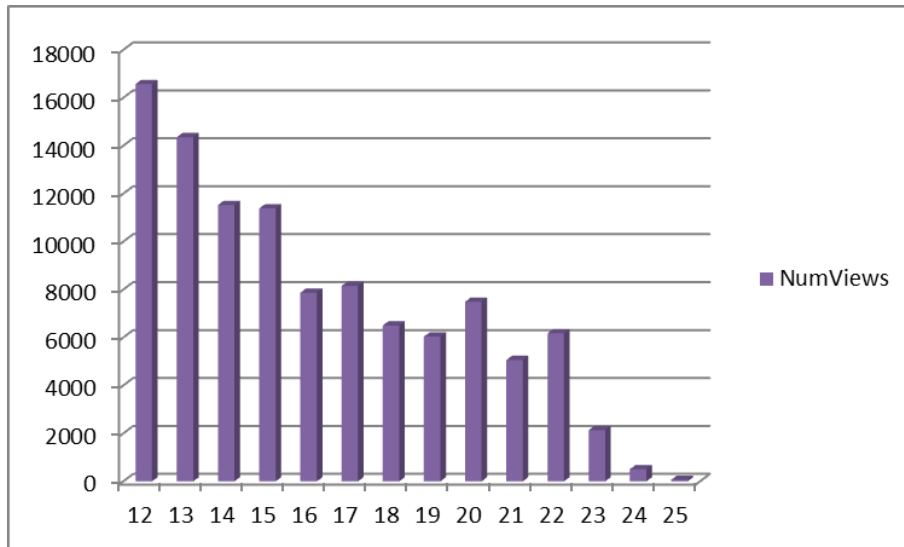

Figure 2. Number of views by week
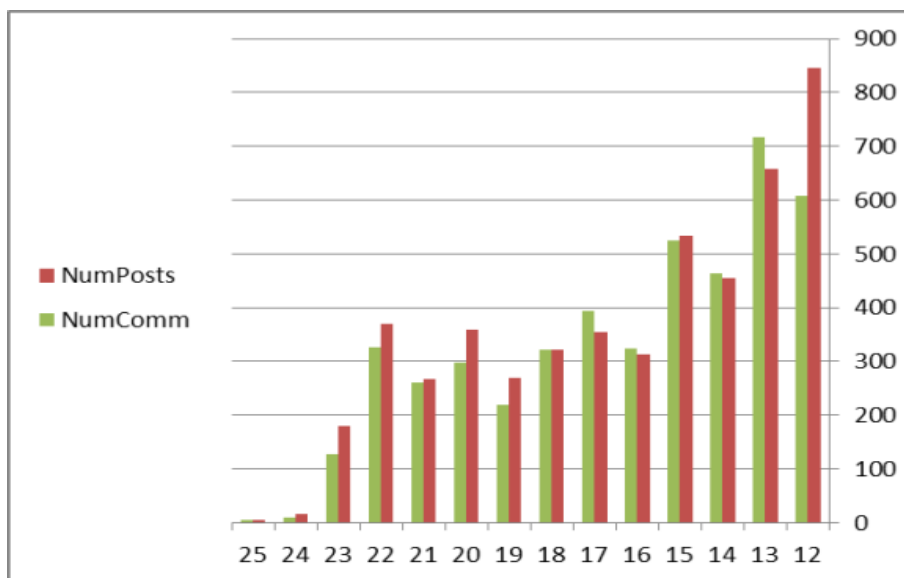
Figure 3. Number of posts and comments by week

Participation patterns of students as reflected in the forums

27,322 students were registered for the MOOC, 4,500 were registered to use the forums, and only 1,257 were active in the forums. Furthermore, when the data was reviewed at the student level, it was surprising to see that three quarters of students were hardly active. Most of the activity came from one quarter of the students who wrote 3 – 266 posts, wrote 2 – 279 comments, and had 18 – 7,013 views in the forums.

Taking a deeper look at student activity, we identified those who chose only to "view" and not to interact with their peers. Especially during the first couple of weeks, most students just "watched" and were "off stage". The number of students who just watched during the first week was 5 times higher than the number of students who had written posts that week.

During the last two weeks of the course, activity reduced to zero (weeks 24, 25). Figure 4 shows that more students wrote posts than comments (responses to posts); on average, 1.76% more. Each student wrote an average of 2.2 posts compared to 3.7 comments during the course. We can assume that some posts created chain reactions of follow-up comments.
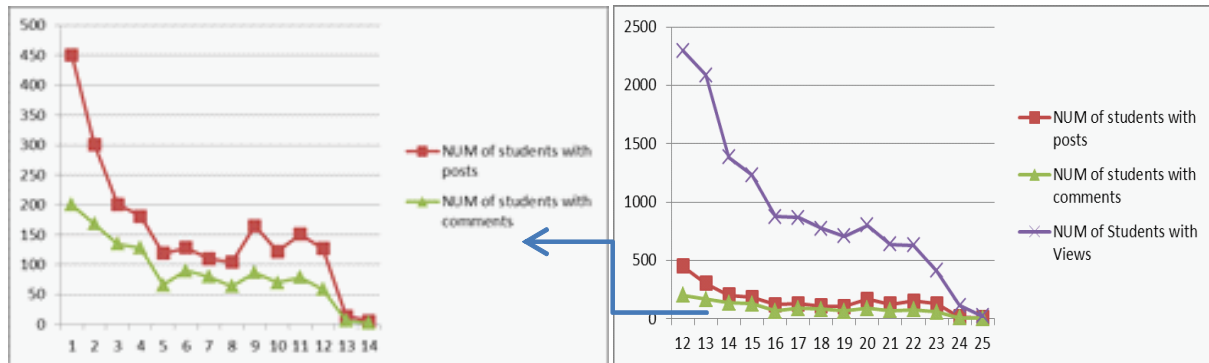
Figure 4. Number of student posts, comments, and views

In order to locate groups of learners with the same patterns of activity in the forums, Two-Step Cluster Analysis was conducted. The included variables were: sending posts, sending comments, viewing messages (posts/ comments), giving evaluation points to a message, tagging, and the device used while participating in the forums. Four clusters resulted, as shown in Figure 5. The second cluster (the activist-viewer) was the only group characterized as mobile users. The participants in clusters 1 and 2 were active, with the first cluster (24.9%) having a high level of each activity and the second cluster (25.5%) being characterized more as viewers. Clusters 3 and 4 had low activity, although cluster 3 (9%) was more reactive to others by writing comments and giving points of evaluation and cluster 4 (the largest cluster, 40%) did more posting, viewing, and tagging (Table 4).

Figure 5. Cluster sizes

Table 4:   Cluster descriptions (the colours represent the input (predictor) importance)

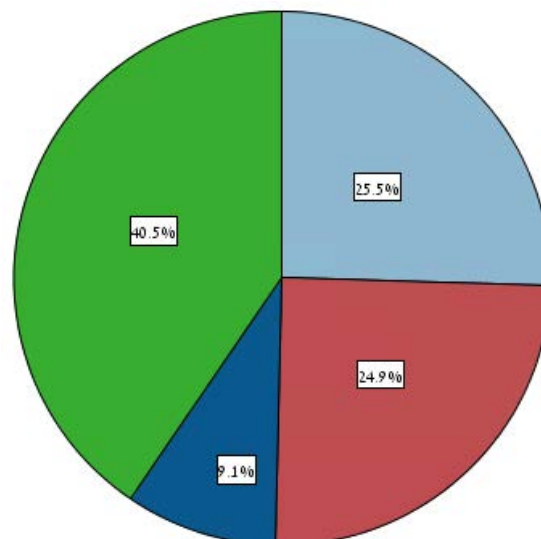| Cluster	1 "Activist – on the stage"	2 "Activist – viewer"	3 Low and reactive with comments and points	4 Low with views and tags
Size	24.9%	25.5%	9%	40%
Mobile/Not mobile	Not mobile	Mobile	Not mobile	Not mobile
Comments	2.75	1.67	2.15	1.00
Posts	2.32	1.78	1.0	1.45
Views	1.80	1.57	1.0	1.48
Points	10.33	2.19	1.05	0.95
Tags	0.78	0.38	0.01	0.24

Variables which may correlate with level of participation

Table 5 describes the correlation between key variables that characterize the activity in the forums. Significant correlations among all activity types (posts, comments, points, and tags) were found. Active students in the forum performed similarly in each channel by writing posts, providing comments, and tagging. However, there are no correlations among these variables and the students' location or achievements (successfully finishing the course).

Table 5:   Correlation among variables

	Comments	Forum Views	Points	Tags	Location	Achievement
Posts	.817**	.601**	.659**	.470**	.044	.197**
Comments	1	.641**	.475**	.381**	.045	.160**
Forum Views		1	.213**	.311**	.007	.064*
Points			1	.365**	.049	.176**
Tags				1	.034	.131**
Location					1	.138**

**P<0.01 *P<0.05

Analysis of forum activity using keywords

Observation of words used in discussion tags/posts/comments can be visualized using "cloud words", as shown in Figure 6 (the number of word instances is reflected in the font size of each word). The cloud shows that intensive dialogue revolves around topics covered in the forums and related to the course subject (Modern Middle East). However, words related to learning management were triggered as well, including assessment (exercises, assignments, tests) and other components available on the course website, such as video lecturers and course pages. Students also discussed different tracks of study.

Figure 6. Keyword cloud

This paper presents preliminary findings regarding students' activity in MOOC forums. These findings may help in understanding social processes and supporting the adoption of discussion groups, both in online courses and other virtual spaces. The results of the study may clearly identify the characterization of the online learner (viewers versus posters), the definition of effective forms of online learning (relevant topics), and support the discussion of "Whether this is a new instructional method".

References

1. Allen, E. and Seaman, J. (2014). *Grade Change: Tracking Online Education in the United States.* Babson Survey Research Group and Quahog Research Group, LLC.

2. Balfour, S.P. (2013). Assessing writing in MOOCs: Automated essay scoring and calibrated peer review. In *Research & Practice in Assessment, 8(1),* (pp. 40-48.)

3. Bell, F. (2011). Connectivism: Its place in theory-informed research and innovation in technology-enabled learning. In *The International Review of Research in Open and Distance Learning, 12(3),* (pp. 98-118.)

4. Bidjerano, T. and Dai, D.Y. (2007). The relationship between the big-five model of personality and self-regulated learning strategies. In *Learning and Individual Differences, 17(1),* (pp. 69-81).

5. Brinton, C.G.; Chiang, M.; Jain, S.; Lam, H.; Liu, Z. and Wong, F.M.F. (2013). *Learning about social learning in MOOCs: From statistical analysis to generative model.* ArXiv Preprint arXiv:1312.2159,

6. Cormier, D. (2008). Rhizomatic education: Community as curriculum. In *Innovate, 4(5).*

7. Downes, S. (2007). What connectivism is. In *Half an hour, blog, February 3, 2007.* Retrieved from http://halfanhour.blogspot.com/2007/02/what-connectivism-is.html

8. Johnson, L.; Adams Becker, S.; Cummins, M.; Estrada, V.; Freeman, A. and Ludgate, H. (2013). *NMC Horizon Report: 2013 Higher Education Edition.* Austin, Texas: The New Media Consortium.

9. Kissinger, J. and Bennett, D. (2014). Using connectivism as a framework for redesigning courses. In the *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education,* (pp. 1015-1019).

10. Komarraju, M.; Karau, S.J.; Schmeck, R.R. and Avdic, A. (2011). The Big Five personality traits, learning styles, and academic achievement. In *Personality and Individual Differences, 51(4),* (pp. 472-477).

11. Mackness, J.; Waite, M.; Roberts, G. and Lovegrove, E. (2013). Learning in a small, task–oriented, connectivist MOOC: Pedagogical issues and implications for higher education. In *The International Review of Research in Open and Distance Learning, 14(4).*

12. McAuley, A.; Stewart, B.; Siemens, G. and Cormier, D. (2010). *The MOOC model for digital practice.* Retrieved from https://oerknowledgecloud.org/content/mooc-model-digital-practice

13. Piech, C.; Huang, J.; Chen, Z.; Do, C.; Ng, A. and Koller, D. (2013). *Tuned models of peer assessment in MOOCs.* arXiv preprint arXiv:1307.2579.

14. Siemens, G. (2014). *Connectivism: A learning theory for the digital age.*