In-depth Exploration of Engagement Patterns in MOOCs

First Author[1][0000-1111-2222-3333] and Second Author[2][1111-2222-3333-4444]

[1] [first address]
[2] [second address]
email@com

Abstract. With the advent of 'big data', various new methods have been proposed to explore data in several domains. In the domain of learning (and e-learning, in particular), the outcomes lag somewhat behind. This is not unexpected, as e-learning has the additional dimensions of learning, engagement, and other psychological aspects, to name but a few, beyond 'simple' data crunching. This means, the goals of data exploration for e-learning are somewhat different to the goals for practically all other domains: finding out what students do is not enough, it is only the means to the end of supporting student learning and increasing their *engagement*. This paper thus focuses specifically on student engagement, a crucial issue especially for MOOCs, by studying in much greater detail than previous work, the *engagement of students based on clustering students according to three fundamental* (and, arguably, comprehensive) *dimensions: learning, social, assessment*. The study's value lies also in the fact that it is among the few studies using *real-world longitudinal data* (6 runs of a course, over 3 years) from a *large number of students*.

Keywords: E-learning · Behavioral analysis · Clustering analysis · K-means · MOOCs.

1 Introduction

"Educators, theorists, and policymakers alike tout engagement as a key to addressing educational problems, such as low achievement and escalating dropout rates" [12]. However, it is a known fact that, in many online environments, and especially in MOOCs, due to "the absence of teacher supervision and opportunities to provide direct feedback, students may lack opportunities to control and interact with a learning environment" [10]. Thus, engagement of the students is a vital target for analysis as well as enhancement, especially for MOOCs. This paper proposes an *in-depth method for exploration of engagement patterns in MOOCs*, based on clustering of students according to *three fundamental* (and, arguably, comprehensive) *dimensions: learning, social and assessment*. This is, to our knowledge, the first study to propose such an in-depth engagement exploration based on clustering (here, the widely popular k-means clustering method is used). Additionally, the paper analyses *real-world data* from a *longitudinal study* of 6 runs of a course, between 2015-1017, with a *large number of students* (48,698), on a MOOC platform that has seen less exploration,

albeit rooted in pedagogical principles [13], unlike some of its competitors, namely, *FutureLearn* (www.futurelearn.com).

The remainder of the paper contains first related work, after which section 3 presents our methodological approach; next, section 4 reports the results of the in-depth analysis. We briefly discuss findings and conclude in section 5.

2 Related Work

Student and learner engagement has many definitions, depending on the perspective it is employed from. Similarly, there is no clear collective understanding on the methods of monitoring and measuring engagement [5, 12]. An interesting way of defining engagement is that based on the Flow theory [2]. Recent highly-cited work under this umbrella [11] analyses student engagement defined in terms of concentration, interest and enjoyment, which are measured based on self-reporting. The work however focusses on a relatively small number of US high-school students, unlike our study. Recent research on engagement in MOOCs [4] defines engagement in terms of length of time of video-watching, as well as existence of problem solving attempts, parameters which are related to our current study.

A special issue on the subject discusses also the variety in grain-size of the measurement tools, starting from microlevel (e.g., individual's engagement in the moment, task or learning activity) versus macrolevel (as in groups of learners), with measurements for the former such as brain imaging, eye tracking, etc., and for the latter, discourse analysis, observations, ratings, etc. [12] They also note that, whilst they categorize engagement as behavioral, cognitive, emotional and agentic, the motivational and self-regulations constructs run through each of these dimensions. Behavioral engagement, which is the one targeted also by the current study, has been shown in the past to be related to achievement in learning [9]. [12] cautions however that higher-order processing (such as exams or strategic thinking tasks) might not be well caught by behavioral engagement. However, we argue here that MOOCs don't provide normally exams per se, and that the type of tests at the end of a MOOC often emulate the level of the tasks given during the MOOC (including the inclusion or exclusion of higher-order processing tasks).

A relatively recent review of measurement methods for student engagement [5] concludes that, whilst most technology-mediated learning research uses self-report measures of engagement, in fact, physiological and systems data offer an alternative method to measuring engagement, and that more research is needed in this area, including determining the system data and values needed for the engagement evaluation. Our in-depth, longitudinal study builds upon these recommendations and proposes novel ways of analyzing and measuring the student engagement in MOOCs.

3 Method

3.1 MOOC Settings and the Dataset

Each MOOC on FutureLearn is hierarchically structured into *weeks*, *activities* and *steps*. A *week* may contain several *activities* and an *activity* may contain several *steps*. A *step* is a basic learning unit which may be an article, a video with or without a discussion (comment) list. A *step* may also be a quiz which consists of a set of questions.

The MOOC presented in this study consists of 4 *weeks*. Each *week* contains 4 *activities* and each *activity* contains between 2 and 8 *steps*. In total, there are 18 *steps* in Week 1, 22 *steps* in Week 2, 15 *steps* in Week 3 and 19 *steps* in Week 4. Thus, in total, there are 74 *steps* in the MOOC. The last *activity* of each *week* contains a 'quiz type' of *step*. Each quiz has 5 questions, so there are 20 questions in total in the MOOC. Each *step*, except the 'quiz type' of *step*, provides a discussion board where students can submit comments, and 'like' (as in social network apps e.g. Facebook) each other's comments. Each *step*, except the 'quiz type' of *step*, also provides a "Mark as complete" button for students to claim that they have learnt the *step*.

The MOOC ran 6 times between 2015 and 2017. The total number of students we analyzed is of 45,321, divided between the six runs as 12,628, 9,723, 7,755, 6,218, 8,432 and 3,942. However, 3,377 students unenrolled from the MOOC. Next, 20,532 did not visit any *step*, being thus passively. Therefore, after filtering these out, in total, 24,798 students are considered in this study.

The dataset collected on FutureLearn platform contains behavioral information including visiting a step, marking a step as completed, submiting a comment, 'like'ing a comment, and attempting to answer a question in a quiz. These 5 types of behavioral information are considered in this study. Besides, each time a student attempts to answer a question, FutureLearn records if the answer is correct or incorrect. This is the additional data used in this study.

3.2 Clustering and Fundamental Dimensions

With complex dataset, we empirically explore student engagement patterns without relying on predefined classes. Clustering, an unsupervised machine learning method, can uncover new relationships in a complex dataset, and has been used to develop profiles that are grounded in student behaviors [1]. In this study, k-means [8], a well-known non-hierarchical clustering method, is used to partition students into different clusters. This is essential, as it provides insights into engagement patterns caused by the diversity of students, as well as opportunities to compare these patterns and predict behaviors. K-means requires *k*, the number of clusters, as a parameter, but determining this parameter is known to be a challenging issue. One way to determine an optimal *k* is the "elbow method", which relays on visually identifying the "elbow point" of a curve drawn on a line chart, but the problem is that this "elbow" cannot always be unambiguously identified, and sometimes there is no elbow or several elbows [6]. In our case, indeed, we were not able to identify a conclusive *k*, but instead obtained several interesting clustering options. Besides, a k-means algorithm normally

favors higher values of *k*, but the latter is not necessarily desirable, as it is very important to consider a more sensible *k* for the nature of the dataset. It is common to run k-means clustering a few times with 3, 4 or 5 as *k*, and compare the results, to determine which is the "final optimal" k to use [1]. In this study, we start with *k=2*, with further increments of *1*.

The clustering is based on three fundamental dimensions, namely, *learning*, *social* and *assessment*. In terms of how we determine these three fundamental dimensions, firstly, the core of using FutureLearn (or any other e-learning platforms) is to learn. Thus, to explore engagement patterns, it is essential to investigate how students access learning content. On FutureLearn, basic learning units of a MOOC are *steps*. Therefore, we consider how students visit *steps* as the first dimension, and we label it as 'learning'. Secondly, FutureLearn employs a social constructivist approach inspired by Laurillard's Conversation Framework [7], which describes a general theory of effective learning through conversation (or social interaction). Therefore, we consider how students interact with each other as the second dimension, and label it as 'social'. Thirdly, FutureLearn, as an xMOOCs platform, consider both content and assessment as essential elements of the teaching and learning process [3]. Therefore, we consider how students attempt to answer questions in quizzes as the third dimension, and label it as 'assessment'. Regarding the parameters for the k-means algorithm, we use: 1) (the number of) *steps* (visited by students) to represent the first dimension – learning; 2) (the number of) comments (submitted by students) to represent the second dimension – social; and 3) (the number of) attempts (to answer questions in quizzes) to represent the third dimension – assessment.

4 Results

4.1 Clustering and Validation

Firstly, we tested $k = 2$ (two clusters) for the k-means analysis. The convergence was achieved in the 17th iteration. The final cluster centers (Fig. 1) show that on the standardized scale, all the three variables, i.e., $\mathrm{Zscore(steps)}$, $\mathrm{Zscore(comments)}$ and $\mathrm{Zscore(attempts)}$, of cluster I are higher than those of cluster II. This suggests that students who were allocated in cluster I may be more engaged in the learning, in terms of visiting steps, submitting comments (discussions) and attempting to answer questions in quizzes, than those who were allocated in cluster II.

A one-way ANOVA test was conducted to compare the relative weight of $\mathrm{Zscore(steps)}$, $\mathrm{Zscore(comments)}$ and $\mathrm{Zscore(attempts)}$ in the clustering process. The result shows very large *F* scores ($F_{steps} = 85{,}155.321$; $F_{comments} = 4{,}441.474$; $F_{attempts} = 90{,}965.767$), and very small p values ($p_{steps} < .001, p_{comments} < .001, p_{attempts} < .001$) indicating that all three variables have a statistically significant impact on determining the clustering, and that the variables of $\mathrm{Zscore(steps)}$ and $\mathrm{Zscore(attempts)}$ have stronger impact than the variable of $F_{comments}$ ($F_{steps} \approx F_{attemps} \gg F_{comments}$, $F_{steps} \approx F_{attemps} \approx 22 \times F_{comments}$).

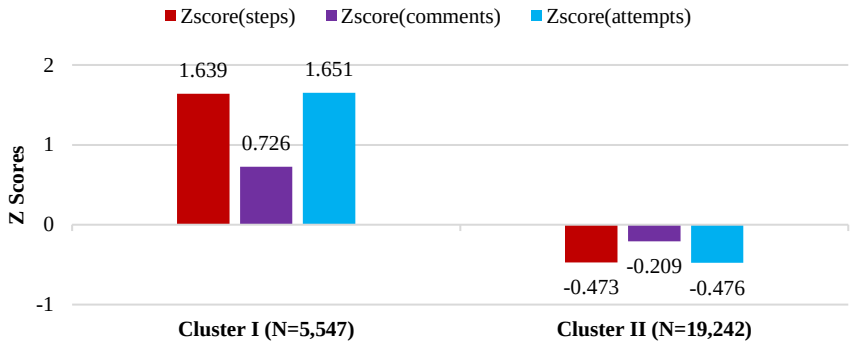

Fig. 1. Comparisons of standardized numbers of steps, comments and attempts between two clusters (k-means analysis, when k=2)

Secondly, we tested $k = 3$ (three clusters) for the k-means analysis. The convergence was achieved in the 16th iteration. We can see from Fig. 2, the final cluster centers, that similar to the k-means analysis result using $k = 2$, the majority of students (18,998, 69.52%) were allocated in cluster II, and they were less engaged in terms of visiting steps, submitting comments (discussions) and attempting to answer questions in quizzes, than those who were allocated in cluster I and cluster III. Interestingly, Fig. 2 also shows that despite $mean_{Zscore(steps)}$ and $mean_{Zscore(attempts)}$ of cluster I and cluster III are similar, $mean_{Zscore(comments)}$ of cluster I is much smaller than that of cluster III. This suggests that among the students who were more engaged, those who were allocated in cluster II might submit much more comments (discussions) than those who were allocated in cluster III.

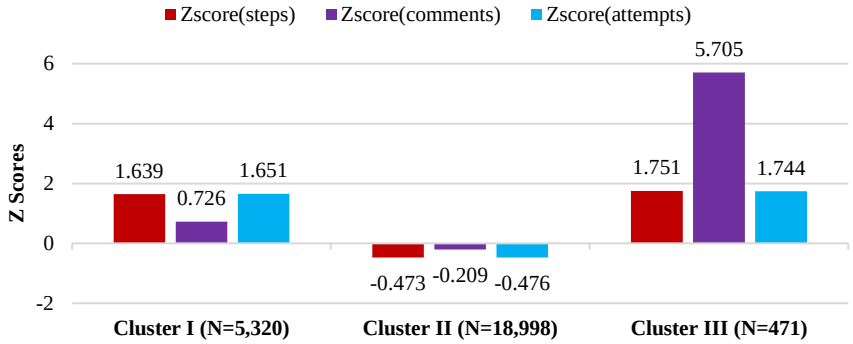

Fig. 2. Comparisons of standardized numbers of steps, comments and attempts between three clusters (k-means analysis, when k=3)

A one-way ANOVA test was performed to compare the relative weight given to Zscore(steps), Zscore(comments) and Zscore(attempts) in order to determine which cluster a student was allocated to. We find from the ANOVA test result that, similar to k-means analysis with *k=2*, the *F* scores of all three variables are very large

($F_{steps} = 47{,}865.306$; $F_{comments} = 24{,}194.624$; $F_{attempts} = 43{,}215.587$, and their p values are very small ($p_{steps} < .001, p_{comments} < .001, p_{attempts} < .001$) indicating all these three variables have a statistically significant impact on determining the clustering. Nevertheless, interestingly, with $k = 2$, F_{steps} and $F_{attemps}$ are about 22 times of $F_{comments}$; whilst with $k = 3$, F_{steps} (47,865.306) and $F_{attemps}$ (43,215.587) are only about 1.8 times of $F_{comments}$ (24,194.624). This means that, in comparison to the k-means analysis with $k = 2$, Zscore(steps), Zscore(comments) and Zscore(attempts) have relatively more even impact on student clustering.

Thirdly, we tested $k = 4$ (four clusters) for the k-means analysis. The convergence was achieved in the 18th iteration. Fig. 3 shows the final cluster centers. Similar to the clustering analysis result using $k = 2$ and $k = 3$, the largest cluster is the one with the less engaged students. For the three clusters with more engaged students, their $mean_{Zscore(steps)}$ and $mean_{Zscore(attempts)}$ are similar, whereas Zscore(comments) of these three clusters are very different. This means that, Zscore(comments) plays a major role of allocating engaged student to different clusters.

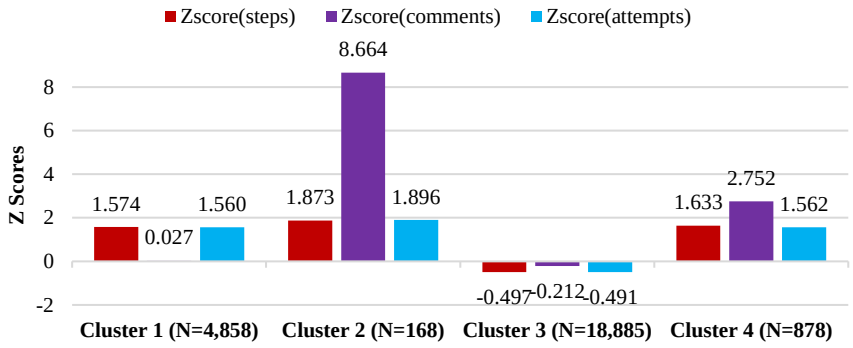

Fig. 3. Comparisons of standardized numbers of steps, comments and attempts between four clusters (k-means analysis, where k=4)

A one-way ANOVA test was conducted, indicating, indeed, in opposite of the k-means analyses results using $k = 2$ and $k = 3$, the Zscore(comments) has stronger impact on determining which cluster a student was allocated to than Zscore(steps) and Zscore(attemps), but not much stronger ($F_{steps} = 31{,}478.383$; $F_{comments} = 35{,}541.079$; $F_{attempts} = 27{,}902.858$; $F_{steps} \approx F_{comments} \approx F_{attempts}$), in comparison to the results from k-means analyses with $k = 2$ and $k = 3$. However, the impacts of all Zscore(steps), Zscore(comments) and Zscore(attempts) are significant on determining student clustering ($p_{steps} < .001, p_{comments} < .001, p_{attempts} < .001$).

We also tested $k \in \{n | 5 \leq n \leq 8\}$. However, the convergence could not be achieved within 20 iterations. Because there are only three variables, i.e. Zscore(steps), Zscore(comments) and Zscore(attempts), we determined that the largest possible number of clusters should be $k \in \{n | n \leq 2^3 = 8\}$. Therefore, we decided to discard the options where $k \geq 5$.

Overall, with $k \in \{n|2 \leq n \leq 4\}$, all k-means analysis results suggest a division between more engaged students and less engaged students; whilst with $k \in \{3, 4\}$, the k-means analysis results reveal more information about how the students were engaged in learning. Additionally, we conducted two Tukey's honestly significant difference (HSD) post hoc tests (at 95% Confidence Interval): the first test was to compare the differences of Zscore(steps), Zscore(comments) and Zscore(attempts) between these three clusters under the condition of $k = 3$ (Table 1 shows the result); the second test was to compare these three variables between four clusters under the condition of $k = 4$ (Table 2 shows the result). We can see from Table 1 that, when $k = 3$, all the variables $\mathrm{Zscore(steps)}$, $\mathrm{Zscore(comments)}$ and $\mathrm{Zscore(attempts)}$ are significantly different ($p < .001$) between these three clusters. However, as shown in Table 2, when $k = 4$, whilst $\mathrm{Zscore(steps)}$ and $\mathrm{Zscore(attempts)}$ are significantly different ($p < .001$) between those four clusters, the $\mathrm{Zscore(comments)}$ does not significantly differ from cluster I and cluster IV. Therefore, we determine that only when $k = 3$, we can obtain strong and stable clusters.

Table 1. Tukey HSD test result - Multiple Comparisons (**k=3**)

Dependent Variable	Cluster Number of Case	Cluster Number of Case	Mean Difference	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Z score (steps)	I	II	2.094*	.007	.000	2.077	2.110
		III	-.149*	.022	.000	-0.200	-0.098
	II	I	-2.094*	.007	.000	-2.110	-2.077
		III	-2.243*	.021	.000	-2.292	-2.193
	III	I	.149*	.022	.000	0.098	0.200
		II	2.243*	.021	.000	2.193	2.292
Z score (comments)	I	II	.429*	.009	.000	0.408	0.450
		III	-5.480*	.028	.000	-5.546	-5.414
	II	I	-.429*	.009	.000	-0.450	-0.408
		III	-5.909*	.027	.000	-5.973	-5.846
	III	I	5.480*	.028	.000	5.414	5.546
		II	5.909*	.027	.000	5.846	5.973
Z score (attempts)	I	II	2.070*	.007	.000	2.052	2.087
		III	-.161*	.023	.000	-.214	-.108
	II	I	-2.070*	.007	.000	-2.087	-2.052
		III	-2.231*	.022	.000	-2.283	-2.179
	III	I	.1613*	.023	.000	.108	.214
		II	2.231*	.022	.000	2.179	2.283

*. The mean difference is significant at the 0.05 level.

Table 2. Tukey HSD test result - Multiple Comparisons (**k=4**)

Dependent Variable	Cluster Number of Case	Cluster Number of Case	Mean Differ-ence	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Z score (steps)	I	II	-.299*	.036	.000	-.391	-.207
		III	2.071*	.007	.000	2.052	2.090
		IV	-.059*	.017	.002	-.102	-.016
	II	I	.299*	.036	.000	.207	.391
		III	2.370*	.035	.000	2.279	2.461
		IV	.240*	.038	.000	.141	.338
	III	I	-2.071*	.007	.000	-2.090	-2.052
		II	-2.370*	.035	.000	-2.461	-2.279
		IV	-2.130*	.016	.000	-2.171	-2.090
	IV	I	.059*	.017	.002	.016	.102
		II	-.240*	.038	.000	-.338	-.141
		III	2.130*	.016	.000	2.090	2.171
Z score (comments)	I	II	-8.636*	.034	.000	-8.724	-8.549
		III	.239*	.007	.000	.221	.257
		IV	-2.725*	.016	.000	-2.766	-2.684
	II	I	8.636*	.034	.000	8.549	8.724
		III	8.876*	.034	.000	8.789	8.962
		IV	5.911*	.037	.000	5.817	6.005
	III	I	-.239*	.007	.000	-.257	-.221
		II	-8.876*	.033	.000	-8.962	-8.789
		IV	-2.964*	.015	.000	-3.003	-2.926
	IV	I	2.725*	.016	.000	2.684	2.766
		II	-5.911*	.037	.000	-6.005	-5.817
		III	2.964*	.015	.000	2.926	3.003
Z score (attempts)	I	II	-.336*	.038	.000	-.432	-.240
		III	2.051*	.008	.000	2.032	2.071
		IV	-.002	.018	**1.000**	-.047	.043
	II	I	.336*	.038	.000	.240	.432
		III	2.387*	.037	.000	2.292	2.483
		IV	.334*	.040	.000	.231	.438
	III	I	-2.051*	.008	.000	-2.071	-2.032
		II	-2.387*	.037	.000	-2.483	-2.292
		IV	-2.053*	.017	.000	-2.095	-2.011
	IV	I	.002	.018	**1.000**	-.043	.047
		II	-.334*	.040	.000	-.438	-.231
		III	2.053*	.017	.000	2.011	2.095

*. The mean difference is significant at the 0.05 level.

4.2 Comparisons between Clusters

Table 3 summarizes descriptive statistics of steps, comments and attempts of these three clusters. Cluster II has the most students i.e. 18,998, followed by Cluster I with 5,320 students; whilst Cluster III is the least represented, with only 471 students.

Table 3. Descriptive Statistics of Steps, Comments and Attempts of three clusters

		N	Min	Max	Median	Mean	SD
Cluster I	Steps	5,320	6	6	74	64.95	12.241
	Comments	5,320	0	0	1	3.81	5.403
	Attempts	5,320	0	0	23	22.79	7.228
Cluster II	Steps	18,998	1	1	10	12.34	11.147
	Comments	18,998	0	0	0	.75	2.023
	Attempts	18,998	0	0	0	2.21	3.618
Cluster III	Steps	471	20	20	74	68.69	11.373
	Comments	471	24	24	36	42.78	20.198
	Attempts	471	5	5	25	24.39	6.446

The 1st Dimension – Learning: Step Visits and Completion Rate

In Cluster I (5,320 students), a very large amount of the students visited a large percentage of steps. In particular, more than half, i.e., 3,079 (57.88%) students visited more than 90% of the steps; 5,278 (99.21%) students visited more than half, i.e., 50% of the steps. Although very few, there were still some, i.e., 2 (0.04%) students visited less than 10% of the steps; and 5 (0.09%) students visited 10% ~ 20% of the steps. Nevertheless, the smallest percentage of steps visited by the students was 8.0% (6 steps), and there were 912 (17.14%) students who visited all, i.e., 100%, of the steps. In terms of completion rate, 4,920 (92.48%) students marked more than 90% of the steps they visited as 'complete', by clicking on the button “Mark as complete”. There wasn’t any student who did not mark any step that they visited.

In Cluster II (18,998 students), a very large amount of the students visited only a very small percentage of steps, and the more percentage of the steps, the fewer the students. In particular, 8,032 (42.28%) students visited less than 10% of the steps; 4,231 (22.27%) students visited 10% ~ 20% of the steps; 3,393 (17.86%) students visited 20% ~ 30%; 1,789 (9.42%) students visited 30% ~ 40%; and 1,087 (5.72%) students visited 40% ~ 50%. In total, 18,532 (97.55%) students visited less than 50% of the steps. However, there were 9 (0.05%) students visited more than 90% of the steps, and 31 (0.16%) students visited 80% ~ 90%. Nevertheless, the largest percentage of the steps visited was 94.67% (71 steps), i.e., there wasn’t any student who visited all the steps. Regarding the completion rate, 5,493 (28.91%) students marked less than 10% of the steps that they visited as 'complete', and 5,702 (30.01%) students marked more than 90% of the steps that they visited as 'complete'. 11,845 (62.35%) students marked more than 50% of the steps they visited as 'complete'. Comparing to 'visit rate', the completion rate was very high. Interestingly, albeit 5,339 (28.10%)

students did not mark any step that they visited as 'complete', there were still 1,541 (8.11%) students who marked all the steps that they visited as 'complete'.

In Cluster III (471 students), similar to Cluster I, a very large number of students visited a large percentage of steps. In particular, 332 (70.49%) students visited more than 90% of the steps; 53 (11.25%) students visited 80% ~ 90%; 30 (6.37%) students visited 70% ~ 80%. Interestingly, only 1 (0.21%) student visited 20% ~ 30% of the steps, and only 1 (0.21%) student visited 10% ~ 20%. Surprisingly, no student visited less than 10% of the steps. Regarding completion rate, surprisingly, 457 (97.03%) students marked more than 90% of the steps they visited as 'complete'; 295 (62.63%) students marked all of the steps that they visited as 'complete'; apart from only 1 (0.21%) student marking only 13.51% of the steps as 'complete', all the rest, 471 (99.79%) students, marked more than 70% steps as 'complete'. The completion rate in Cluster III was surprisingly very high.

Fig. 4 compares the percentage of steps visited by students for these three clusters. Cluster I and Cluster III are similar – the larger the percentage of the step being visited, the more the students; whereas Cluster II shows an opposite trend. A Kruskal-Wallis H test was conducted. The result shows that there is a statistically significant difference in the number of steps being visited between these three clusters, $\chi^2(2) = 1{,}318{,}881.844, p < .001$, with a mean rank correct answers rate of 21,773.63 for Cluster I, 9,520.72 for Cluster II, and 22,397.58 for Cluster III. Further Mann-Whitney U tests show that there are statistically significant differences between Cluster I and Cluster II ($Z = -111.106, U = 365{,}933.5{,}796, p < .001$), between Cluster I and Cluster III ($Z = -8.086, U = 978{,}445.5, p =< .001$), and between Cluster II and Cluster III ($Z = -36.97, U = 37{,}226, p < .001$).

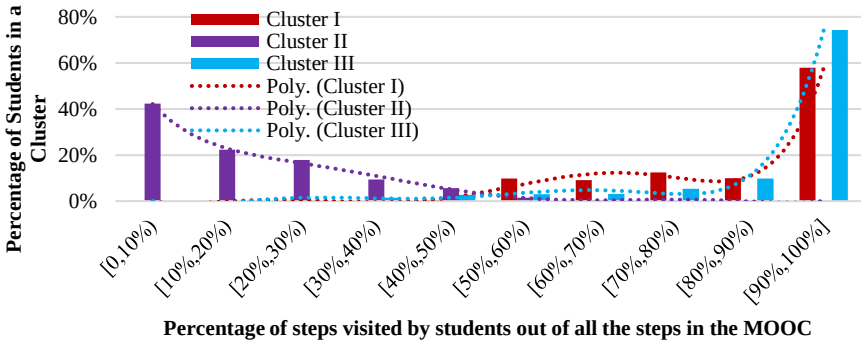

Fig. 4. Comparison of the number of steps visited by students between three clusters

Fig. 5 shows comparisons of the completion rate between clusters. Again, Cluster I and Cluster III share a similar trend, whilst Cluster II is very different from Cluster I and Cluster II. A Kruskal-Wallis H test was performed to examine these differences. The result shows that there is a statistically significant difference in completion rate between these three clusters, $\chi^2(2) = 8{,}461.923, p < .001$, with a mean rank correct answers rate of 19,830.14 for Cluster I, 10,106.02 for Cluster II, and 20,741.28 for Cluster III. Further Mann-Whitney U tests show statistically significant differences

between Cluster I and Cluster II ($Z = -88.502, U = 10{,}796{,}504, p < .001$), between Cluster I and Cluster III ($Z = -5.635, U = 1{,}069{,}606, p < .001$), and between Cluster II and Cluster III ($Z = -31.447, U = 726{,}183, p < .001$).

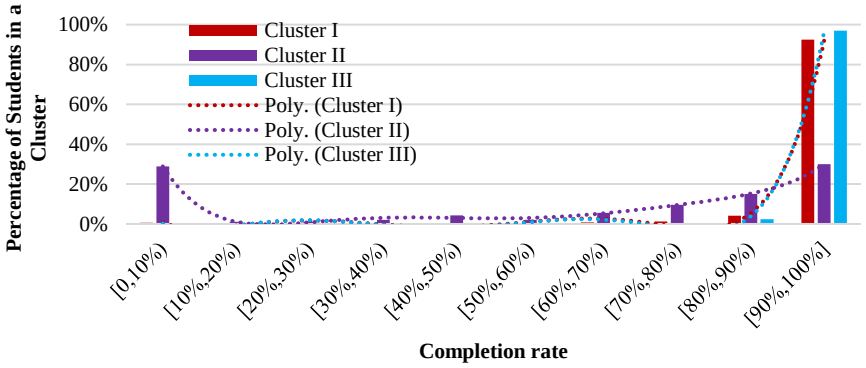

Fig. 5. Comparison of completion rate between three clusters

The 2nd Dimension – Social: Comments and 'Likes'.

In Cluster I (5,320 students), 2,959 (55.62%) students submitted at least one comment; they submitted 20,254 comments in total, with an average of 6.84, standard deviation of 5.63, and median of 5. Overall (all 5,320 students together), the average number of comments was 3.81 with standard deviation of 5.40. The median number of comments was 1. There were 26 students who submitted the largest number of comments, i.e., 23. Regarding 'likes', all those 2,959 students, who submitted at least one comment, received 30,044 'likes', in total. The average number of 'likes' was 10.15 with standard deviation of 12.34. The median number of 'likes' was 5. There were 409 students who submitted at least one comments but did not receive any 'likes'. The most popular student received the largest number, 81, of 'likes'.

In Cluster II (18,998 students), only 5,007 (26.36%) students submitted at least one comment; they submitted 14,318 comments in total, with an average of 2.86, standard deviation of 3.08, and median of 2. Overall (all 18,998 students together), the average number of comments was 0.75 with standard deviation of 2.02. The median number of comments was 0. There was 1 student who submitted the largest number of comments, i.e., 26. Regarding 'likes', all those 5,007 students, who submitted at least one comment, received 14,979 'likes', in total. The average number of 'likes' was 2.99 with standard deviation of 5.43. The median number of 'likes' was 1. There were 1,773 students who submitted at least one comments but did not receive any 'likes'. The most popular student received the largest number, 86, of 'likes'.

In Cluster III (471 students), all (100%) students submitted comments: 20,151 in total, with an average of 42.78, standard deviation of 20.20, and median of 36. 24 students submitted the smallest number, 24, of comments. One student submitted the largest number, 179, of comments. Regarding 'likes', in total, they received 30,164 'likes'. The average number of 'likes' was 64.04 with standard deviation of 50.38. The

median number of 'likes' was 52. Only 1 student did not receive any 'likes'. The most popular student received the largest number, 441, of 'likes'.

Fig. 6 (left) compares the percentage of students submitting comments in the three clusters. Cluster I and Cluster II are similar– a very large percentage of students submitted very few comments; although a larger percentage of students did not submit any comments in Custer II. As for Cluster III, the peak appears between 20 and 30, and no student submitted less than 24 comments, which is very different from Cluster I and Cluster II. A Kruskal-Wallis H test shows a statistically significant difference in the number of comments submitted by the students between these three clusters, $\chi^2(2) = 4{,}168.348, p < .001$, with a mean rank correct answers rate of 15,605 for Cluster I, 11,194.54 for Cluster II, and 24,553.76 for Cluster III. Mann-Whitney U test results show statistically significant differences between Cluster I and Cluster II ($\mathrm{Z} = -48.616, \mathrm{U} = 322{,}02{,}216, \mathrm{p} < .001$), between Cluster I and Cluster III ($\mathrm{Z} = -37.337, \mathrm{U} = 0, \mathrm{p} < .001$), and between Cluster II and Cluster III ($\mathrm{Z} = -46.886, \mathrm{U} = 112, \mathrm{p} < .001$). Interestingly, the Mann-Whitney U test for Cluster I and Cluster II results is $U = 0$. Thus, all the students in Cluster III submitted more comments than any students in Cluster I.

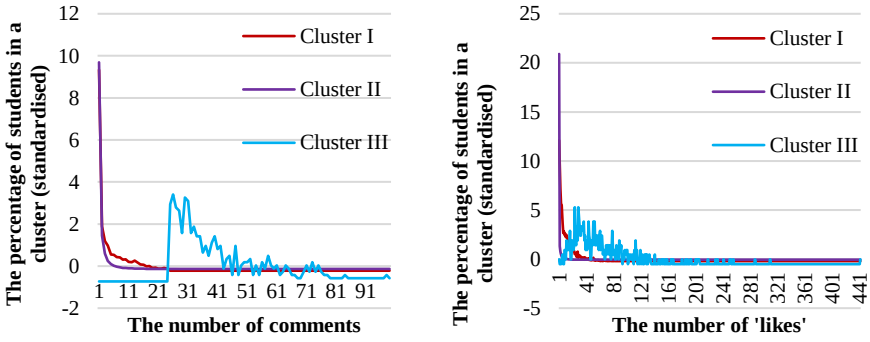

Fig. 6. Comparison of the number of comments (on the left) and 'likes' (on the right) vs the percentage of students between three clusters

Fig. 6 (right) shows the comparison of the number of 'likes' received by certain percentage of students between three clusters. Similar to the comparison of the number of comments, Cluster I and Cluster II share similar trend, but Cluster III has a very different trend: the peak of Cluster I and Cluster II appear in the very left end of the horizontal axis i.e. the majority of students in Cluster I and Cluster II received very few, if not zero, comments; whereas Cluster III has a peak at around 30 'likes'.

The 3rd Dimension – Assessment: Attempts and Correct Answers Rate.

In Cluster I (5,320 students), only 51 (0.96%) students did not attempt to answer any question; the majority of 5,269 (99.04%) students attempted between 2 and 88 times answering questions. Overall (all 5,320 students together), the average number of

attempts was 22.79 with standard deviation of 7.23. The median number of attempts was 23. Regarding the correct answers rate, all those 5,269 students who attempted answering questions correctly answered at least one question. The average correct answers rate was 72.48% with standard deviation of 11.15%. The median correct answers rate was 71.42%, and the lowest correct answers rate was 9.52%. Gratifyingly, 43 (0.82%) students' correct answers rate was 100%.

In Cluster II (18,998 students), 13,477 (70.94%) of them did not attempt to answer any question, and the rest 5,521 (29.06%) attempted to answer at least one question. Overall (all 18,998 students together), the average number of attempts was 2.21 with standard deviation of 3.62, but the median number of attempts was 0. In terms of the correct answers rate, among those 5,521 (29.06%) students who attempted at least once to answer a question, only 14 (0.25%) of them did not correctly answered any question, even though they attempted between 1 and 6 (mean=1.86, SD=1.61) times to answer a question. Excluding those 13,477 students who did not attempted to answer any questions, the overall average correct answers rate was 67.84% with standard deviation of 14.52%; and the median correct answers rate was 62.50%. Surprisingly, there were 372 (6.74%) students whose answer was 100% correct.

In Cluster III (471 students), every student attempted at least 5 times to answer a question. The maximum number of attempts was 45, the median was 25. The average number of attempts was 24.39 with standard deviation of 6.45. With regards to correct answers rate, the lowest correct answers rate was 43.75%, the median was 74.07%. Overall, the average correct answers rate was 73.67% with standard deviation of 10.06%. However, only 2 (0.42%) students' correct answers rate was 100%.

The 100% stacked column chart (Fig. 7 left) suggests that the pattern of attempting answering questions between three clusters are very different: the majority of students (13,477; 70.94%) in Cluster II did not attempted to answer any question; whilst almost all the students in Cluster I (5,269; 99.04%) and Cluster III (471; 100%) had attempted answering questions. A Kruskal-Wallis H test shows a statistically significant difference in the number of attempts between these three clusters, $\chi^2(2) = 15{,}326, p < .001$, with a mean rank attempts of 21,678.18 for Cluster I, 9,552.93 for Cluster II, and 22,176.74 for Cluster III. Further Mann-Whitney U tests show statistically significant differences between Cluster I and Cluster II ($Z = -120.445, U = 949{,}826.5, p < .001$), between Cluster I and Cluster III ($Z = -5.709, U = 1{,}054{,}516, p < .001$), between Cluster II and Cluster III ($Z = -44.784, U = 965{,}172.5, p < .001$).

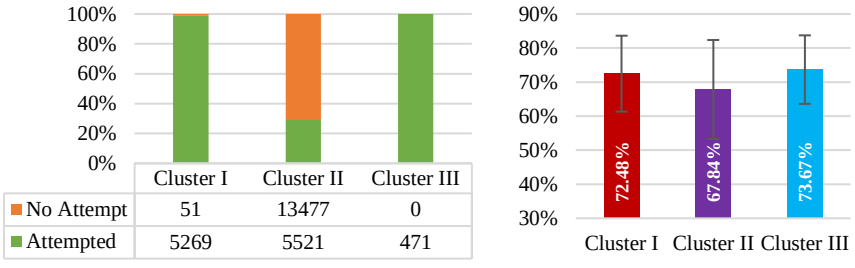

Fig. 7. Comparisons of numbers of attempts (left); correct answers rates (right)

Although the mean correct answers rates of these three clusters are similar, as shown in **Error! Reference source not found.** on the right, a Kruskal-Wallis H test shows that there is a statistically significant difference in correct answers rate between these three clusters, $\chi^2(2) = 499.995, p < .001$, with a mean rank correct answers rate of 6,268.20 for Cluster I, 4,939.29 for Cluster II, and 6,610.90 for Cluster III. Further Mann-Whitney U tests show statistically significant differences between Cluster I and Cluster II ($Z = -21.323, U = 11{,}113{,}796, p < .001$), between Cluster I and Cluster III ($Z = -2.148, U = 1{,}166{,}974.5, p = .032 < .05$), and between Cluster II and Cluster III ($Z = -10.894, U = 912{,}535.5, p < .001$).

5 Discussions and Conclusions

In this study, we firstly defined three-dimensional metrics to measure engagement patterns. The *fundamental* (and, arguably, comprehensive) *dimensions include learning*, *social* and *assessment*. We then employed *k-means analysis* to cluster students based on the metrics. Our clustering and validation approach resulted in very strong and stable three clusters. We further applied statistical models to explore the differences of engagement patterns between these three clusters, in three dimensions, from the point of view of six aspects:

1. *Learning (the 1st dimension)*
   a. the number of steps visited by students (determining aspect)
   b. completion rate (performance aspect)
2. *Social (the 2nd dimension)*
   a. the number of comments submitted by students (determining aspect)
   b. the number of 'likes' received by the students (performance aspect)
3. *Assessment (the 3rd dimension)*
   a. the number of attempts to answer questions in quizzes (determining aspect)
   b. correct answers rate (performance aspect)

The statistical analysis further supported that our clustering results were very strong and stable, as all three dimensions defined in the metrics (determining facts) had statistically significant impact on determining clusters, with very large *F* values, at a *p<.001* level. Moreover, all the above three performance aspects were statistically significantly different between those three clusters. This allowed exploring in depth how students were engaged in learning in the MOOC.

Students in Cluster II were the least engaged: they visited the least *steps*, attempted the least questions, submitted the least comments. The disengagement could predict poor performance, i.e., their completion rate, 'likes', and correct answers rate were the lowest in comparison with students in the other two clusters. Unfortunately, the least engaged students represented the largest share of the cohort – one of the issues in MOOCs in general. Cluster I and Cluster III represented engaged students yet in different ways. Students in Cluster I and Cluster III shared similar trend when comparing the above six aspects, but we did find statistically significant differences at a *p<.001* level. However, *U* values from Mann-Whitney U tests suggested that social dimen-

sion (comments) was the most differentiating aspect (only $U_{comments} = 0$, meaning all students in Cluster III submitted more comments than all students in Cluster I). Importantly, Cluster III received higher scores than Cluster I in terms of for all performance aspects. This interesting result suggests that socially engaged students would also be more engaged in the various learning activities, and their performance could be better than of those who are less socially engaged. Thus, recommender systems could support students in the social exchange, in order to enhance the learning – unlike it was considered in the past, that social exchange could only be distracting from the mainstream learning activity. This, we believe, is an important characteristic of MOOCs in particular, which may not be shared with other type of learning environments. Further work is necessary to refine the recommendations.

References

1. Antonenko, P.D. et al.: Using Cluster Analysis for Data Mining in Educational Technology Research. Educational Technology Research and Development. 60, 3, 383–398 (2012).
2. Csiksczentmihalyi, M. et al.: Flow: The psychology of optimal experience. Australian Occupational Therapy Journal. 51, 1, 3–12 (2004).
3. Ferguson, R., Clow, D.: Examining Engagement: Analysing Learner Subpopulations in Massive Open Online Courses (MOOCs). Presented at the (2015).
4. Guo, P.J. et al.: How video production affects student engagement: an empirical study of MOOC videos. Presented at the (2014).
5. Henrie, C.R. et al.: Measuring student engagement in technology-mediated learning: A review. Computers & Education. 90, 36–53 (2015).
6. Kodinariya, T.M., Makwana, P.R.: Review on Determining Number of Cluster in K-Means Clustering. International Journal. 1, 6, 90–95 (2013).
7. Laurillard, D.: Rethinking University Teaching: a Conversational Framework for the Effective Use of Learning Technologies. RoutledgeFalmer, London ; New York (2002).
8. MacQueen, J., others: Some Methods for Classification and Analysis of Multivariate Observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. pp. 281–297 Oakland, CA, USA (1967).
9. Marks, H.M.: Student Engagement in Instructional Activity: Patterns in the Elementary, Middle, and High School Years. American Educational Research Journal. 37, 1, 153–184 (2000).
10. Md Abdullah Al Mamun et al.: Factors affecting student engagement in self-directed online learning module. In: The Australian Conference on Science and Mathematics Education (formerly UniServe Science Conference). p. 15 (2017).
11. Shernoff, D.J. et al.: Student Engagement in High School Classrooms from the Perspective of Flow Theory. In: Csikszentmihalyi, M. (ed.) Applications of Flow in Human Development and Education: The Collected Works of Mihaly Csikszentmihalyi. pp. 475–494 Springer Netherlands, Dordrecht (2014).
12. Sinatra, G.M. et al.: The Challenges of Defining and Measuring Student Engagement in Science. Educational Psychologist. 50, 1, 1–13 (2015).
13. The Pedagogy of FutureLearn: How our learners learn, https://about.futurelearn.com/research-insights/pedagogy-futurelearn-learners-learn.