

Creating Collaborative Groups in a MOOC: A Homogeneous Engagement Grouping Approach

Luisa Sanz-Martínez*, Erkan Er, Alejandra Martínez-Monés, Yannis Dimitriadis and Miguel L. Bote-Lorenzo

GSIC-EMIC Research Group, Universidad de Valladolid, Valladolid, Spain.

*Corresponding Author, email: luisa@gsic.uva.es

Creating Collaborative Groups in a MOOC: A Homogeneous Engagement Grouping Approach

Collaborative learning can improve the pedagogical effectiveness of MOOCs. Group formation, an essential step in the design of collaborative learning activities, can be challenging in MOOCs given the scale and the wide variety in such contexts. We discuss the need for considering the behaviours of the students in the course to form groups in MOOC contexts, and propose a grouping approach that employs homogeneity in terms of students' engagement in the course. Two grouping strategies with different degrees of homogeneity are derived from this approach, and their impact to form successful groups is examined in a real MOOC context. The grouping criteria were established using student activity logs (*e.g.*, page-views). The role of the timing of grouping was also examined by carrying out the intervention once in the first and once in the second half of the course. The results indicate that in both interventions, the groups formed with a greater degree of homogeneity had higher rates of task-completion and peer interactions. Additionally, students from these groups reported higher levels of satisfaction with their group experiences. On the other hand, a consistent improvement of all indicators was observed in the second intervention, since student engagement becomes more stable later in the course.

Keywords: automatic group formation; MOOCs; collaborative learning; engagement; homogeneous grouping

1. Introduction

Massive Open Online Courses (MOOCs) have arisen as an innovative medium to deliver free online education to millions of persons all around the world, thus promoting global and equal access to knowledge. Currently, most MOOCs follow an instructivist pedagogy and prioritise content acquisition (Daniel, 2012). In this type of MOOCs, known as xMOOCs, learners are required to individually study the course content that is typically organised in weekly modules, and their performance is assessed automatically (usually through quizzes at the end of each module). xMOOCs are often criticised for

lacking active learning strategies that are typically employed in constructivist pedagogies (Margaryan, Bianco, & Littlejohn, 2015). This pedagogical limitation is argued as one of the reasons for low student engagement and low-completion rates. Enhancing MOOC pedagogy with active learning strategies has been proposed as one approach to promote MOOC students' engagement in learning tasks and increase the completion rates (Dillenbourg, Fox, Kirchner & Wirsing, 2014; Manathunga & Hernández-Leo 2015).

One popular active learning strategy that could enhance MOOC pedagogy is collaborative learning (CL). CL has been proven to offer many learning benefits in various learning settings (e.g., developing higher-order thinking skills, learning from multiple perspectives, and improving teamwork skills) (Dillenbourg, 1999; Stahl, Koschmann, & Suthers, 2006). Several research attempts have been made to implement CL in MOOC contexts to enhance student learning. However, the true potential of CL in supporting student learning in MOOCs so far has not been realized due to the challenges in the design and implementation of CL activities induced by the massive and variable scale of MOOC contexts (Manathunga & Hernández-Leo 2015).

One major challenge in this regard is the formation of student groups. Group formation is a critical step in the design of a CL activity (Muehlenbrock, 2008; Ounnas, 2010) as the success of collaborative groups depends, to a large extent, on the effective matching among their members (Isotani, Inaba, Ikeda & Mizoguchi, 2009). Group formation in MOOCs faces distinct difficulties. The initial large number of participants and its variability due to new enrolments and dropouts throughout the course pose special challenges in the creation of groups. Moreover, MOOC participants' profiles are diverse, as they register for the course with different goals, and as they come from a variety of backgrounds, education levels, interests, and skills. This diversity often

creates notable differences in their interactions with the course content and engagement in the learning activities. Accordingly, in the MOOC literature, learner subpopulations unique to MOOC contexts have been identified, such as *no-shows*, who never visit a course page after their enrolment or *observers*, who view the course contents but have no interest in actively participating in the course (Hill, 2013; Kizilcec, 2013; Alario-Hoyos, Pérez-Sanagustín, Delgado-Kloos, Parada-G., & Muñoz-Organero, 2014). Such learners, with limited or no engagement in the course, may hamper the collaboration in their own groups, thus negatively affecting the learning process of their peers (Ounnas, 2010). Therefore, it is necessary to consider the engagement levels of MOOC learners when forming student groups for conducting collaborative learning activities in massive learning scales.

Due to an increased recent interest in applying CL in MOOCs, several approaches towards group formation have been proposed. Some grouping strategies (widely used in contexts other than MOOCs) have been tested to create random and criteria-based groups (Zheng, 2017), in which criteria such as knowledge background, personality or preferences (Spoelstra, 2014) have been applied to create homogeneous or heterogeneous groups (Wichmann et al., 2016). Furthermore, various technological approaches, such as social network analysis (Sinha, 2014) or natural language processing (Wen, 2016), have been applied. Using these grouping approaches, some MOOC-specific tools have been developed (Cheng, Yu, Park & Zhu, 2015) to facilitate group formation. However, these previous research efforts did not take into account the existence of some learner subpopulations in MOOCs with a very low engagement level (e.g. *no-shows*) that could impede productive peers' collaborative learning experiences in their groups. This gap in the literature calls for new group formation approaches

aiming to increase the likelihood of forming successful collaborative groups among the massive and varying MOOC learners.

The present research study proposes a novel group formation approach that incorporates knowledge about MOOC learners' engagement levels in order to create student teams in which effective collaboration is likely to occur. This approach requires establishing a certain degree of homogeneity within the teams in terms of engagement in the course. To do so, the proposed approach uses the analytics of students' activity in the course to gauge their engagement through various indicators, which then inform the grouping process to create homogeneous engagement teams. This grouping approach was implemented in a MOOC using two strategies with different levels of homogeneity, both of which were deployed in two distinct modules of the course. The success of the four types of resulting groups was assessed in terms of peer interactions, completion of the collaborative task and student satisfaction with group work.

The research question guiding the study was: *How does the application of homogeneous-engagement criteria into the group formation process influence the success of the resulting student groups in MOOCs?*

The success of the four types of resulting groups was assessed in terms of peer interactions, completion of the collaborative task, and student satisfaction with group work.

The rest of the paper first presents an analysis of the state of the group formation problem with a particular focus on the issues raised in MOOC contexts, followed by the description of the proposed group formation approach. Then, we present the study carried out in a MOOC deployed in the Canvas Network platform, and discuss the results of the experiments. The final section presents the conclusions and suggests several lines of future research.

2. Related Work in Automatic Collaborative Group Formation in MOOCs

The strategies employed for group formation strongly influence the learning experiences of the students during collaboration, and consequently, group performance and the individual learning gains (Isotani et al., 2009; Konert et al., 2014). Poorly formed groups can negatively influence the peer interactions, which may lead to detrimental effects such as isolation, conformity, anti-intellectualism, intimidation, and levelling-down of learning quality (Ounnas, 2010). Therefore, group formation is a very critical stage in CL.

In the CSCL field, several tools and systems have been proposed to support automatic group formation in face-to-face and blended learning scenarios using different techniques and algorithms (Magnisalis, Demetriadis & Karakostas, 2010). However, MOOCs have particular characteristics, which preclude a direct extrapolation of lessons learned from these studies to massive and open learning contexts. In MOOCs, the flexibility in the enrolment dates, the high dropout rate, and the presence of students with no activity in the course cause major variations in the target population throughout the course. Moreover, the diversity among MOOC participants results in a high variability in students' engagement levels and learning behaviour, thus hindering the process of group formation.

Currently, only a few MOOC platforms (*e.g.*, Canvas, NovoEd, edX) offer features to set up collaborative groups, while in the courses delivered in platforms (*e.g.*, Coursera, Udacity, FutureLearn) with no group formation support students have even formed external networks to meet and create study groups using services such as

MeetUp¹. Among the three-main group formation approaches i.e., random, self-selected and teachers' criteria-based groups (Ounnas, 2010), the aforementioned MOOC platforms only allow for the automatic formation of random groups. This is a simple yet convenient way of ensuring that every student is assigned to a group, however, it does not guarantee that groups will work productively. Some platforms (e.g., Canvas Network) allow teachers to manually assign students to groups. However, this solution is not always feasible in a course with a massive number of students. There exist also other features to allow students to browse through existing teams (created by the teacher) and select the team that they want to join (mostly by interest on the topic). However, it has been reported that, when this method is used, many students do not manage to join a team (Wen, 2016). Nevertheless, the criteria-based grouping approach, which is the preferred method at small-scale contexts due to its pedagogical affordances, is not currently supported by automatic means by the existing MOOC platforms.

There have been few research addressing the issue of group formation in MOOCs (Wen, 2016). Zheng (2017) used random and survey-based algorithms to compose the groups, proposing later a method for re-composing the groups that are incomplete in size (due to the students' dropout) when a new task begins. Spoelstra and colleagues (2014) analysed team formation in project based learning, using data gathered from surveys about the background knowledge, preferences, and personality of the students as grouping criteria. Sinha (2014) proposed a theoretical approach for dynamic group formation focusing on the use of Social Network Analysis and Machine

¹ <https://www.meetup.com/es/topics/coursera-org/>; <https://www.meetup.com/es-ES/topics/udacity/>; <https://about.futurelearn.com/meetups>

Learning techniques to find relations among students in order to configure the groups. Wen (2016) tested the effectiveness of giving the students the opportunity to interact meaningfully with the community before they are assigned to teams, in order to extract evidence of which students would work well together.

Although some of the aforementioned research studies have considered social interactions among students to create the groups (Sinha, 2014; Wen 2016) and others have taken into account the possible re-composition of damaged groups (Zheng, 2017) none of these studies has considered the students' engagement dynamics in MOOCs and their distinctive behavioural patterns (e.g., *no-shows*) as main factors to inform the group formation process. Thus, for implementing successful collaborative activities in open and massive contexts, there is a need for automatic group formation approaches that consider a variety of indicators of the learners' engagement in the course (Sanz-Martínez, Martínez-Monés, Bote-Lorenzo, Muñoz-Cristobal & Dimitriadis, 2017).

3. Proposed Approach

The proposed approach aims at creating potentially successful groups, in which interactions are more likely to occur, by establishing a basic level of homogeneity regarding students' engagement with the course. This basic homogeneity is implicit in non-open and in formal educational contexts, where less variance in students' behaviour is expected. Setting this homogeneity is essential to obtain a more reliable student sample by minimizing the harm introduced by absent students (i.e., *no-shows*).

To implement the homogeneous engagement grouping approach, learning analytics are employed to track MOOC learners' activities and to obtain the following three indicators of student engagement that were already proposed in (Ferguson & Clow, 2015):

- Number of `page_views` (coded as [`num_page_view`]), to measure the content engagement.
- Number of submitted assignments (coded as [`num_subm_assi`]), to estimate the assessment engagement.
- Number of posted messages in forums (coded as [`num_post_mess`]), to gauge the discussion engagement.

These indicators are used to inform the grouping method to establish some degree of homogeneity among the members of a group. An important issue is to detect the students with no-engagement at all, because they are not really in the course and consequently they do not leave any traces. In other words, it is necessary to identify those students who are enrolled in the course but shows no activity (i.e., no-shows). This type of students, which hardly exists in formal educational contexts, represents in MOOCs a considerable percentage of the total number of enrolled students, resulting in a handicap to form effective collaborative groups. In the proposed approach, we can identify no-show students as those who have viewed zero pages and therefore they have a zero value in the variable [`num_page_view`].

Section 5.2 describes how we applied this approach to the study reported in this paper, thus showing one possible implementation of this grouping method.

4. Method, Research Question and Data Sources

This section describes the methods and research question used to conduct the study, as well as the data sources used to carry out and to evaluate the experiments. This study is part of an overarching research project conducted through a Design Based Research Methodology (DBR) (McKenney & Reeves, 2012) beginning with explorative iterations and evolving towards more evaluative cycles. The research work reported in this paper

corresponds to the second iteration of this project, where we aimed to explore the effectiveness of the grouping approach described in the previous section to form successful student teams. This study was guided by the following research question:
RQ: How does the application of homogeneous-engagement criteria into the group formation process influence the success of the resulting student groups?

In order to study the degree to which the homogeneous engagement grouping approach leads to successful collaborative groups, we designed two strategies that implemented the approach with different levels of homogeneity. We also tested both strategies at different weeks of the course to assess the influence of the timing of the group formation on the effectiveness of the approach.

The success of the resulting groups was measured in terms of: (i) participation level in the collaborative activity (i.e., number of messages posted and number of active participants), (ii) submission status of the collaborative activity (i.e., submitted or not), (iii) satisfaction of students regarding the collaboration carried out in their group.

We performed an anticipatory data reduction process (Miles & Huberman, 1994) and identified two main issues that should be explored through different topics and informative questions shown in Figure 1. The first issue (I1) was related to the impact of implementing this approach using different degrees of homogeneity on the resulting student groups themselves and on their members, while the second issue (I2) was related to the influence of the timing of the group formation (i.e., earlier or later in the course) on the effectiveness of the implemented grouping approaches.

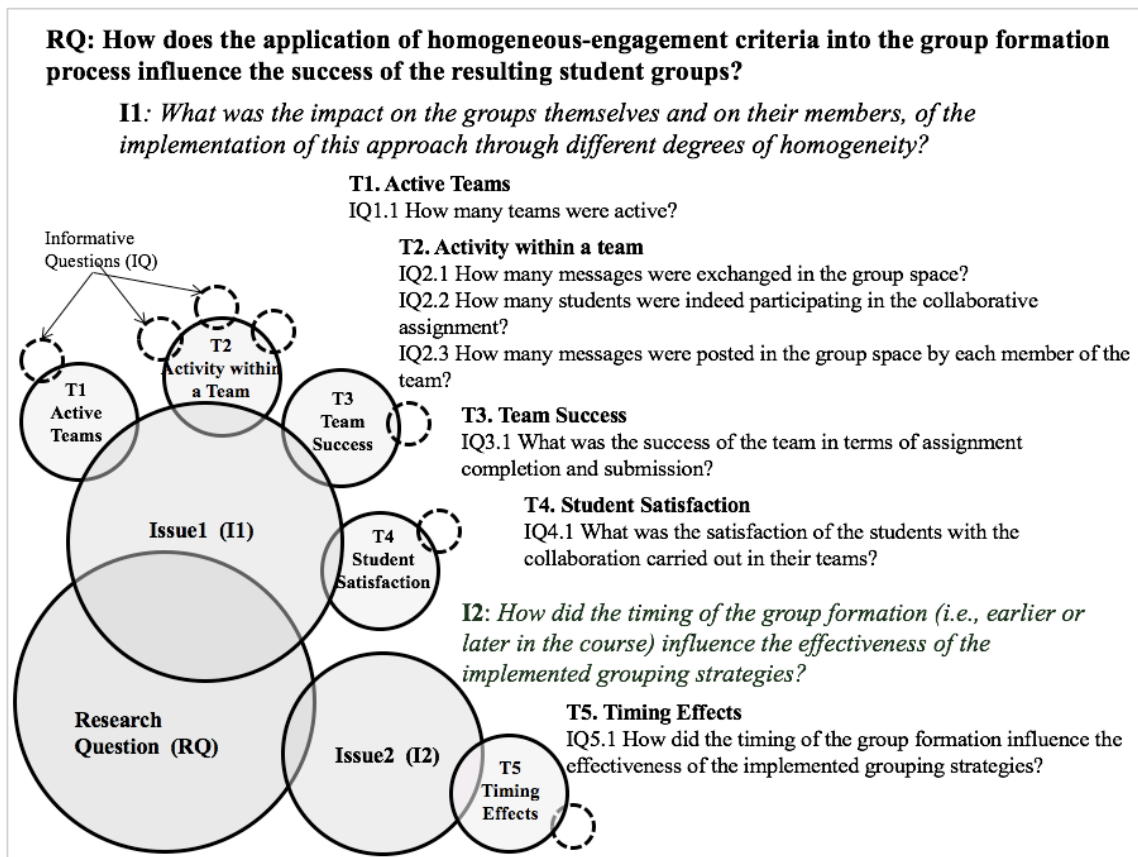


Figure 1. Anticipatory data reduction diagram showing research question (RQ), issues (I), topics (T) and informative questions (IQ).

We used a mixed-methods approach in order to better capture the effects of the grouping strategies examined in the study. The goal of mixing was complementarity (Greene, Caracelli & Graham, 1989) by using several data sources to collect information about group performance, students' participation in group activities and their satisfaction with group experience. Mixed methods allowed us to triangulate and complement results (Greene, Caracelli & Graham, 1989). This approach is a consequence of our underpinning pragmatic worldview, focusing on the problem to be solved and to real world practice (Creswell, 2014).

Accordingly, we gathered both quantitative and qualitative data from multiple sources, shown in Table 1.

Table 1. Data sources used (codes indicated within brackets) to create the groups and to measure the effects of the grouping strategies employed.

Source	Description
Surveys [SurX]	<p>Course surveys composed of open-ended and closed questions in a 4-point Likert scale (1 = strongly disagree, 2 = disagree, 3= agree, 4= strongly agree, + don't know/no answer) were administered:</p> <ul style="list-style-type: none"> • [Sur1]. - At the beginning of the course (optional) to get ethnographic data and preferences of the students. • [Sur2]. - At the end of the course (mandatory) to obtain students' satisfaction with the course.
Platform use Analytics [AnaX]	<p>GET functions of the Canvas LMS REST API were used to collect indicators about:</p> <ul style="list-style-type: none"> • [Ana1], [Ana3]. - Students' engagement variables (i.e., <code>[num_page_view]</code>, <code>[num_subm_assi]</code> and <code>[num_post_mess]</code>) used to inform the group formation process. • [Ana2], [Ana4]. - Activity carried out during the group assignments (active teams, activity carried out within a team, effectiveness of the teams), used to evaluate the impact of the strategies approaches implemented.
Communication from students to teachers [Com]	<p>Emails and personal messages sent in the Canvas platform from the students to the teachers during the collaborative assignments (4th and 6th weeks).</p>

During the enactment of the collaborative activities, data from the platform analytics was collected in order to check the teams' performance and the students' participation. This information allowed us to find out the teams that were active and the students of each team that were indeed participating in the activity. We also measured the teams' effectiveness regarding the task completion. We use the term *active team* to refer to those groups that exchanged messages in the group space. The term *active student* is used to refer to the students who participated in the collaborative activity by posting messages and announcements in the group space. Similarly, the term *team size* refers to the number of *active students* within an *active team*. We used the term *small size* when the team had one or two active students, *medium size* for those teams with

three or four active students and *large size* for teams that registered interactions among five or six active students.

To measure the satisfaction of the students with the collaboration carried out in their group we gathered the communications sent from students to teachers during the enactment of the collaborative assignments as well as their responses in the final satisfaction survey.

The diagram depicted in Figure 2 shows the timeline of data collection and other main events (i.e., end of enrolment, creation of the collaborative groups) as related to the course schedule.

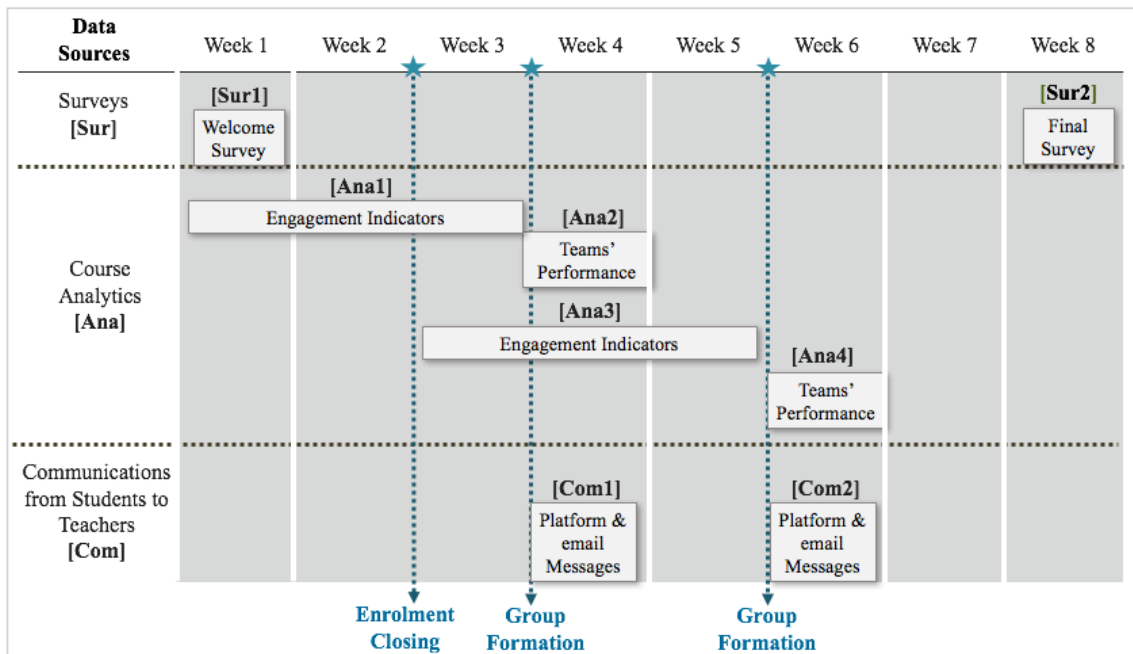


Figure 2. Timeline of main events as related to the course schedule.

5. Description of the Study

5.1 Context

This study was carried out in an introductory-level MOOC that teaches how to translate economy and finance related texts from Spanish to English. This course had been

initially envisioned by instructors of the Faculty of Translation at University of Valladolid, Spain. We formed a co-design team composed of these instructors and researchers to review the learning design of the course and improve its instructional quality in several ways including active learning pedagogies (Ortega-Arranz et al., 2017). The team decided to design a collaborative activity that was deployed as two identical compulsory assignments on the fourth and the sixth week of the course.

The collaborative activity consisted in extracting terminology from a set of financial texts in teams of six members. Each team was required to create a glossary of 20 economic or financial English terms and their corresponding Spanish translations. Students used group discussion forums and announcements features of Canvas Network MOOC platform to communicate with peers in their groups and to carry out the collaborative task. Also, each team was required to choose a spokesperson who would be in charge of the assignment submission (i.e., the glossary with the final list of terms). Finally, each group member had to perform a peer review on the glossary produced by another group. A student was considered to have successfully completed the assignment only if the group assignment and individual peer review were submitted.

The course was deployed in the Canvas Network platform between February the 6th and April the 2nd, 2017, i.e., a total of eight weeks: seven weeks (one per module) plus an additional week that allowed students to complete any pending activity (e.g., submitting the last assignments, completing peer reviews and answering the final satisfaction survey). The enrolment was closed at the end of the second week. The total number of students enrolled was 1031 (which dropped to 875 until the end of the course) and 132 students achieved the certificate (15.09% of the students remained enrolled till the end of the course).

5.2 Experimental Design

Two distinct strategies were developed to test the impact of homogeneity in the success of the resulting teams (see I1 in section 4). The first one, called Random Strategy (RN-S), consisted in randomly selecting the six members of each team. However, to avoid the foreseen negative effects of including no-shows in the groups, the students that had shown no activity at all in the course were taken apart in a previous step. This led to two slightly homogeneous clusters: students with no activity at all (i.e., $[\text{num_page_view}] = 0$), and students with some activity (i.e. $[\text{num_page_view}] > 0$). Within these two clusters the RN-S strategy was applied to create random groups. Thus, with this strategy, a very coarse level of homogeneity within teams was established.

The second grouping strategy, called Homogeneous Strategy (HM-S), aimed to achieve a higher level of homogeneity within the teams by forming groups based on the similarity in students' levels of engagement in the course. Three variables (described in section 3) were computed based on the data collected from the course analytics (data sources [Ana1] and [Ana3] of Figure 1) to measure student engagement:

$[\text{num_page_view}]$, $[\text{num_subm_assi}]$ and $[\text{num_post_mess}]$.

The algorithm selected for implementing the homogeneous grouping was k-means clustering as it has shown to be effective with large datasets (Wen, 2016). Since the k-means algorithm does not necessarily result in clusters (i.e., groups) with the same size, we slightly modified it to ensure that the resulting clusters had the same size (same size k-means variation²). Prior to the clustering process, the three engagement indicators were standardized in order to ensure that they had the same weight in the calculations of the grouping algorithm, as recommended in (Mohamad & Usman, 2013).

² https://elki-project.github.io/tutorial/same-size_k_means

Both strategies were applied to the group formation process in two collaborative assignments (see Section 4.1) planned for two different weeks of the course (see I2 in section 4), i.e., at the week four (4W) and six (6W) respectively. It is noteworthy to mention that in both assignments, a window of 21 days was used to trace data about the students' activity in the platform. For the first collaborative activity, this length was the distance between the course start and the beginning of the activity. The same window length (i.e., 21) was also applied when obtaining the trace data in the second assignment.

Before applying the grouping strategies, we divided the global cohort of students into two subsets (one for RN-S and another for HM-S). We ensured that the resulting subsets did not statistically differ from each other in terms of the variables selected as grouping criteria (i.e., *[num_page_view]*, *[num_subm_assi]* and *[num_post_mess]*). This was an essential step to avoid any bias that could have resulted from unbalanced distribution of students in terms of their engagement levels across two grouping strategies. Because the three variables followed a non-Gaussian distribution, the Wilcoxon test (Blair & Higgins, 1980) was used to test the statistical differences. The set formed by all the students of the course was first shuffled and then split in two equally-sized subsets. This process was repeated until the Wilcoxon test returned a satisfactory p-value for the three variables used as grouping criteria, which allowed us to reject the hypotheses that subsets were different.

In summary, the steps followed to carry out the experiment were:

- (1) Finding out the statistical distribution of the selected variables:
[num_page_view], *[num_subm_assi]* and *[num_post_mess]*).

- (2) Standardising the data, prior to the clustering, in order to assign the same weight to the three selected variables as recommended in (Mohamad & Usman, 2013).
Initially the variable [num_page_view] had a larger range than the other two.
- (3) Splitting the whole cohort of students into two subsets (where each grouping approach would be applied). Checking that they are not statistically different from each other regarding the variables used as grouping criteria.
- (4) Using one of the aforementioned subsets to create the six-member teams according to RN-S using the data source [Ana1] in 4W and the data source [Ana3] in 6W.
 - Identifying the no-show students and segregating them from the rest, thus producing two clusters (i.e., no-show students and the rest).
 - Creating six-member random teams in both clusters.
- (5) Using the other aforementioned subset to create the six-member teams according to HM-S applying as clustering criteria the variables [num_page_view], [num_subm_assi] and [num_post_mess] gathered from the data source [Ana1] in 4W and the data source [Ana3] in 6W.

6. Results

In this section, we describe the data analysis and the findings of implementing RN-S and HM-S in the experiments carried out in 4W and 6W.

6.1 Data Analysis

The first execution of the experiment, carried out in 4W, produced 162 teams (81 per each grouping strategy), while the second one, performed in 6W, produced 150 teams (75 per each grouping strategy). This decrease in the number of teams was due to

dropouts. Activity data of each group (e.g., number of messages exchanged in the group space) were collected from the data sources [Ana2] and [Ana4] (according to Figure 2) in order to respond to the informative questions (IQ) posed in Section 4 and depicted in Figure 1. Table 2 provides a summary of the collected data that were related with IQ1.1, IQ2.1, IQ2.2, IQ2.3 and IQ3.1.

Table 2. Data gathered from the Canvas LMS API at the end of each collaborative assignment.

	4W		6W	
	RN-S	HM-S	RN-S	HM-S
Total number of teams	81	81	75	75
Number of active teams [1,2]	47 (52.02%)	25 (30.86%)	32 (42.67%)	16 (21.33%)
Number of teams that submitted the assignment	46 (56.79%)	26 (32.1%)	30 (40%)	16 (21.33%)
Number of teams that were active but did not submit the assignment	4 (4.94%)	1 (1.23%)	2 (2.67%)	0 (0%)
Number of teams that were inactive but submitted the assignment	3 (3.70%)	2 (2.47%)	0 (0%)	0 (0%)
Number of messages exchanged [1,2,3]	300	372	338	349
Number of active students	76	78	76	71
Number of messages per active student	3.95 (mean) 2.69 (sd)	4.77 (mean) 3.67 (sd)	4.45 (mean) 3.42 (sd)	4.92 (mean) 3.95 (sd)
Median of number of messages per active team	3	10	8.5	15
Number of messages per active team [1,2,3]	6.38 (mean) 5.87 (sd)	14.88 (mean) 14.94 (sd)	10.56 (mean) 9.23 (sd)	21.8 (mean) 16.93 (sd)

Due to the non-Gaussian distribution of the data and the dependence between the measurements in each category, we carried out Wilcoxon tests in order to find out the statistical significance of the differences between the reported data. We coded the categories of significant differences and stated the appropriate codes in the first column of Table 2 between brackets. The codes assigned were:

- [1] - Significant difference between RN-S and the HM-S in 4W.
- [2] - Significant difference between RN-S and the HM-S in 6W.
- [3] - Significant difference between 4W and 6W in RN-S.

[4] - Significant difference between 4W and 6W in HM-S.

According to Table 2, the number of active teams (IQ1.1) and the number of teams that submitted the assignment (IQ3.1) were higher in RN-S than those in HM-S. However, the total number of active students was nearly the same in both strategies (76 with RN-S vs. 78 with HM-S in 4W, and 76 with RN-S vs. 71 with HM-S in 6W), which suggests that both cohorts of students were similar in their engagement levels, as intended in the group creation process. Nevertheless, the Wilcoxon test indicated that the distribution of these active students across the teams created with RN-S differed significantly from the distribution of those created with HM-S. The higher number of active teams achieved with RN-S can be attributed to the fact that active students were randomly spread across different groups. This strategy led to many groups with low activity. On the contrary, the homogeneity achieved using HM-S resulted in a concentration of these active students in fewer teams. Additionally, the total number of messages exchanged per active team in HM-S was more than twice of that in RN-S. Furthermore, the number of messages per active student was also higher in HM-S. These results suggest that HM-S teams showed a more intense activity. Furthermore, the number of teams that had some activity but did not complete the task (and therefore, could not obtain the course certificate) was higher in the RN-S (IQ3.1).

Figure 3 depicts the distribution of active students across the active teams in both weeks. This distribution of the two grouping approaches was significantly different, as shown by the Wilcoxon test. The analysis of this distribution allowed us to delve in IQ2.2.

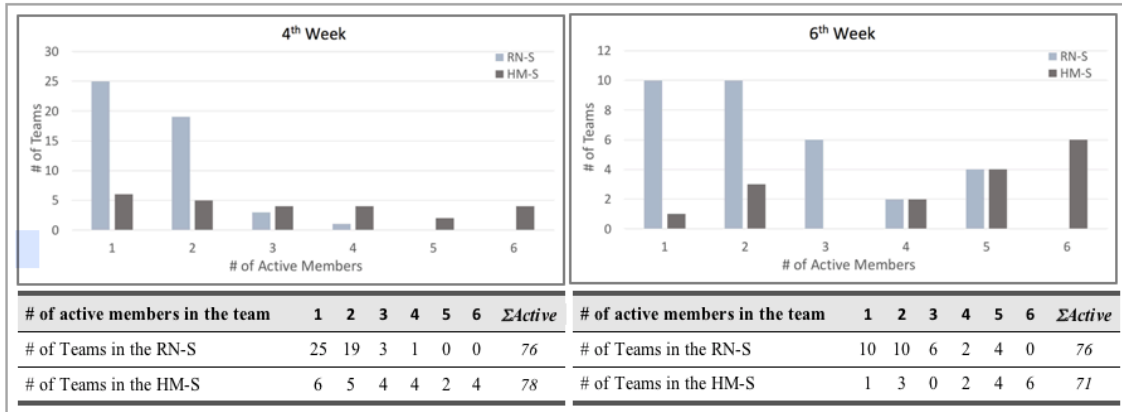


Figure 3. Number of teams with a concrete number of active members in each week.

As observed in Figure 3, RN-S resulted in many *small size* teams with only one or two active members (44 in 4W and 20 in 6W) whereas HM-S achieved to minimize this type of teams (11 in 4W and 4 in 6W). On the other hand, the number of *large size* teams was higher in HM-S as compared with RN-S (16 vs. 4), and only HM-S resulted in teams with six active members in both weeks. There was a significant positive correlation between the average number of messages per active user and the size of the team (0.79 in 4W and 0.66 in 6W). That is, students who were members of a team with many active students were likely to post more messages in their group space.

To address IQ4.1, we examined the student responses to the final survey (identified as data source [Sur2]) regarding their satisfaction with the collaborative assignments. Students' responses to the closed-ended questions are summarized in Table 3. In this table, the responses of "agree" and "strongly agree" are merged into a single category "agree", and similarly the responses of "disagree" and "strongly disagree" have been merged into a single category of "disagree". "The Don't Know/No Answer" responses are not included in this table. To indicate the significant differences with Wilcoxon tests we used the same categories and codes used in Table 2. These codes are shown in the first column of Table 3 between brackets. The distribution of

student responses is depicted as bar charts in Figure 4, in pairs of RN-S and HM-S, as well as 4W and 6W.

Table 3. Summary of the aggregated responses to the closed-ended questions of the satisfaction survey.

	4W				6W			
	RN-S		HM-S		RN-S		HM-S	
	Agree	Disagree	Agree	Disagree	Agree	Disagree	Agree	Disagree
Q1: Satisfaction with the collaboration in my team [1,2,3,4]	35.3%	59.1%	55%	36.6%	61.1%	34.8%	70%	20%
Q2: Inactive students in my team hindered collaboration [1,3,4]	78.9%	12.7%	52.1%	32.4%	59.8%	33.3%	31.4%	51.4%
Q3: Inactive students in my team affected negatively my satisfaction [1,2,3,4]	57.7%	31%	39.5%	43.7%	45.8%	47.2%	28.6%	55.7%
Q4: Collaboration in this activity enhanced my motivation [2,3]	42.3%	42.3%	40.9%	38%	45.8%	43%	54.3%	28.5%
Q5: Collaboration in this activity enhanced my participation	60.5%	26.7%	67.6%	19.7%	62.5%	27.8%	54.3%	22.9%

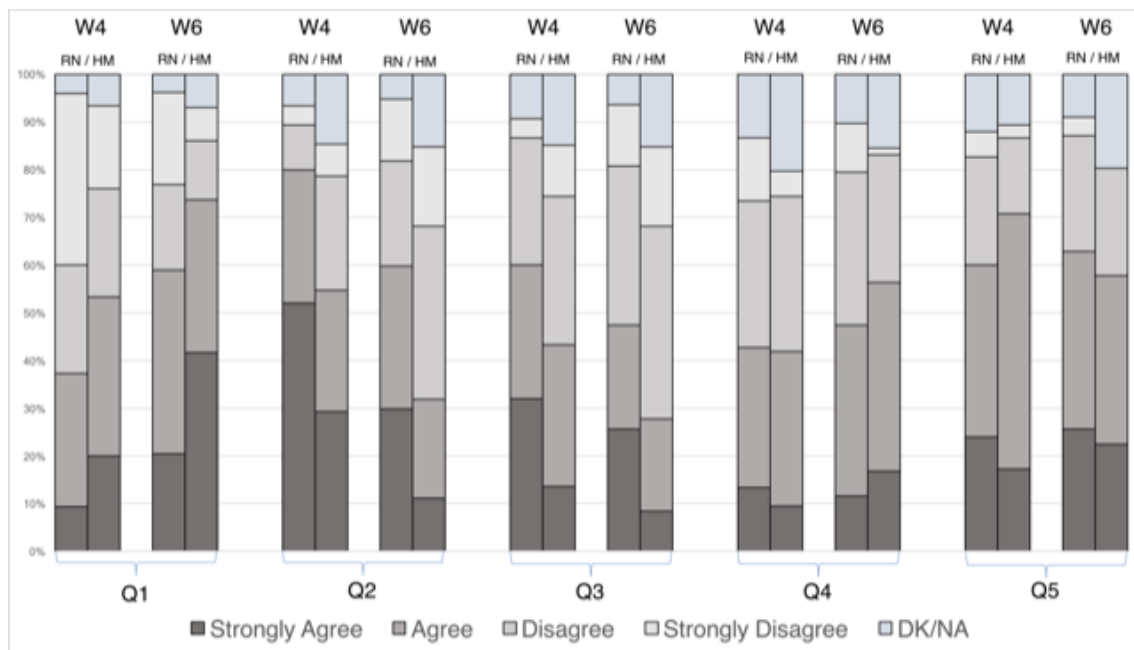


Figure 4. Distribution of the close-ended responses in the satisfaction survey.

According to the results in Table 3, the students who worked in teams built with HM-S were more satisfied with their collaborative work experiences, showing a higher percentage of positive responses for Q1 (55% in 4W and 70% in 6W). On the contrary, those who worked in teams built with RN-S were more frustrated with the presence of

inactive students in their teams, although this frustration decreased in 6W (from 78.9% to 59,8%) as shown in Q2. Furthermore, these inactive students affected negatively the satisfaction of their teammates during 4W in teams built with RN-S, as they stated in their responses to Q3 (57.7%). These observations can be triangulated with the data obtained from the communication logs (data sources [Com1] and [Com2]) between teachers and students. These communications revealed the negative effect of inactive students in the team, which was less prominent in homogenous teams (e.g., *“I sent a message to the group forum in order to distribute the work and I have not received any answer. I have been waiting but finally I have decided to complete this activity only by myself”*, *“I am very interested in completing this assignment, but in my team, there is not a lot of activity and only a girl has sent her proposal of terms for the glossary. Can I add my own terms and send to you our common glossary”*?)

On the other hand, for both strategies, the collaborative activity was perceived to have a positive effect on the participation of students, based on their responses for Q5 in both weeks. However, the responses given to Q4 showed that the collaborative activity had a neutral effect on the students’ motivation.

The data source [Sur2] (final satisfaction survey) also included open-ended questions to ask students about the aspects of the collaborative activity that they liked or disliked. We processed this information together with the email messages that students sent to the instructors and the (private) messages sent within the Canvas Network platform. Thus, we complemented the data obtained by means of the closed questions to get a deeper understanding of the students’ perceptions regarding the collaborative assignments. Table 4 shows a characteristic set of comments expressed by the students.

The majority of complaints came from the students who were the only active member (i.e., only member who posted messages in the group space) in their team. In

many cases, the students in teams with one or two active members expressed frustration due to the lack of participation of their teammates as well as perceptions of losing the opportunity of an enriching activity (e.g., the comments corresponding to students coded as *[std_1]*, *[std_2]*, *[std_3]*, *[std_8]* and *[std_9]* in Table 4). On the other hand, the most positive comments came from the students that belonged to teams with five or six active members. These students expressed their satisfaction with the opportunity of meeting their mates, helping each other and learning from different points of view (e.g., *[std_5]*, *[std_6]* and *[std_7]*, in Table 4). Teams with three or four active members provided both positive and negative comments. On the positive side, the students of these teams show their satisfaction in similar terms as the students of large size teams (e.g., *[std_4]* on Table 4), but on the negative side they express some frustration regarding the absence of some teammates (e.g., *[std_10]* on Table 4).

Table 4. Sample of comments expressed by the students in open-ended questions of the final satisfaction survey.

Student Code	Week	Team type	# Active students	Response
<i>[std_1]</i>	4	RN	2	My colleagues were absent. At least they could have introduced themselves and said that they would not participate instead of keeping us waiting to see if they appeared
<i>[std_2]</i>	4	RN	1	No teammates showed up, although I sent them messages in the forum asking for their availability. I should say that it was an especially unpleasant experience.
<i>[std_3]</i>	4	HM	6	I disliked the lack of participation of many partners They don't answer and it was a handicap to reach agreements about the terms, the spokesman, etc.
<i>[std_4]</i>	4	RN	3	We were able to coordinate the work and we observed the way of working of others. We learnt from each other.
<i>[std_5]</i>	4	HM	6	We have been able to learn from each other and to correct the mistakes committed by our colleagues, a process that leads to a higher level of learning.
<i>[std_6]</i>	4	HM	6	What I liked most was the possibility of having real contact with the classmates. I loved reading many of the translations and the points of view provided by colleagues! There were frankly good translations.
<i>[std_7]</i>	4	HM	4	Although we are partners from all over the world, we managed to finish the activity and maintain a good communication.
<i>[std_8]</i>	6	RN	2	Nobody in the group showed signs of life until the last day. On Sunday afternoon, a girl answered and contributed her terms. She and I done all the assignment. We had no news of the rest of the team.
<i>[std_9]</i>	6	RN	1	I didn't receive any response from my teammates, so I had to do the assignment individually.
<i>[std_10]</i>	6	HM	4	There were some mates that waited till the end of the activity to make something – we didn't know till the last minute if they were still active in the course or if they planned to do something.

[std_11]	6	HM	4	This time we were more teammates resulting in an easier work. Very happy.
[std_12]	6	HM	6	The group assignment of the 6th week was more efficient for our team, although not all the members were able to contribute on time. In my case, I had no time to contribute during the week and I was only able to add my tasks at the weekend.
[std_13]	6	HM	6	We submit the assignment in the limit because two teammates did not answer until the end. The teammates were fabulous.
[std_14]	6	HM	6	This time I were in a more active team and this makes the experience more pleasant.
[std_15]	6	HM	5	In this occasion, I was luckier and almost all my teammates participated. There were some mistakes caused by teammates that appeared in the last moment and tried to participate in the activity.
[std_16]	6	HM	5	After the assignment of the fourth week where nobody in my team participated, it has been very pleasant to find some partners willing to work and participate to complete the assignment of the sixth week.

In order to measure the change in the satisfaction of the students from 4W to 6W, we divided the respondents into four categories regarding the type of strategy according to which their team was formed in each week, and we compared the responses to Q1 (i.e., *Satisfaction with the collaboration in my team*) of these four categories. We coded the responses by assigning the following values to the available options: (i) strongly disagree was assigned to the value 1, (ii) disagree was assigned to the value 2, (iii) agree was assigned to the value 3 and, (iv) strongly agree was assigned to the value 4. Table 5 shows three central tendency statistics (i.e., median, mode and interquartile range) about the responses provided by each category of students. The distribution of student responses is depicted as bar charts in Figure 5.

Table 5. Descriptive statistics of satisfaction for the four categories of students.

	# Students	Satisfaction in 4W			Satisfaction in 6W			Satisfaction Difference	
		Median	Mode	IQR	Median	Mode	IQR	Median	(p-value)
RN (4W), RN (6W)	33	2	1	2	3	4	1	1	0.0010602
RN (4W), HM (6W)	43	2	1	2	3.5	4	1.5	1.5	0.0000013
HM (4W), HM (6W)	35	3	3	1.25	3	4	0	0	0.0685
HM (4W), RN (6W)	42	3	3	2	3	3	0	0	0.2097
All respondents	153	2	3	2	3	3	1	1	0.0000001

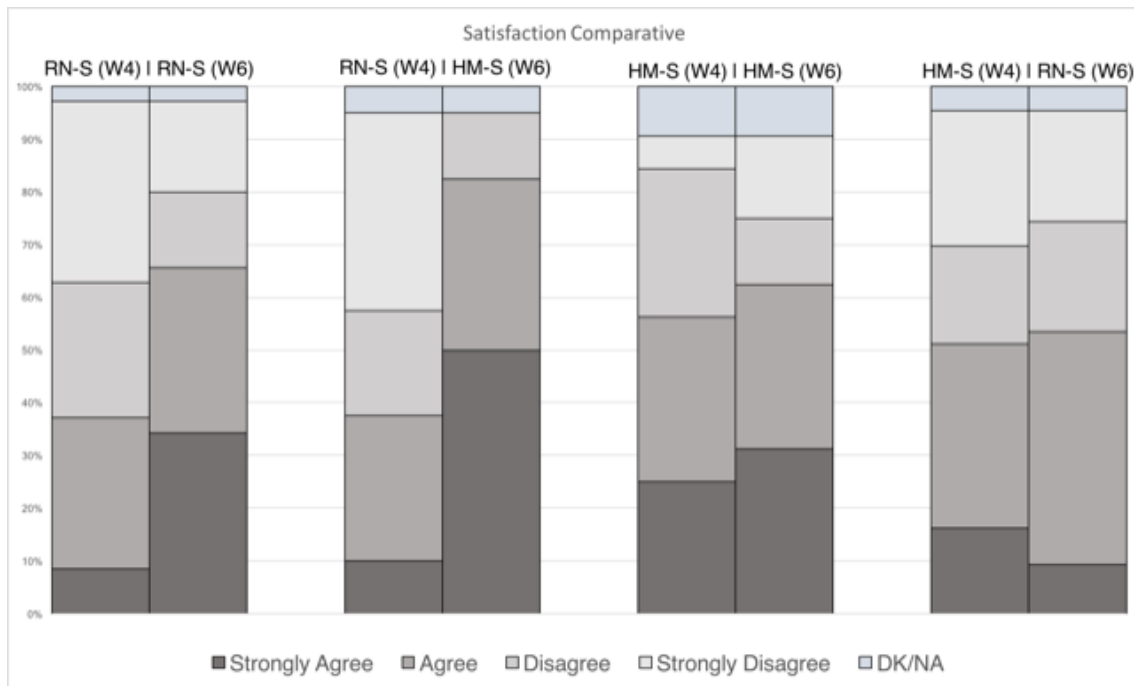


Figure 5. Distribution of the Q1 responses for the four categories of students

In Table 5, the number of students in each category as well as the median, mode and interquartile range of the responses are provided. Wilcoxon tests were used to check the statistical significance of the changes in satisfaction of students from 4W to 6W. The results of the Wilcoxon tests are provided in the last column of the table (bold text is used to indicate significant differences, i.e., $p < 0.05$).

According to the results, there was a significant increase in the satisfaction of students *who* first worked in a team created with RN-S in 4W and who later worked in a team created with HM-S in 6W (from median=2 and mode=1, to median=3.5 and mode=4). This result was complemented and triangulated with data coming from the Survey [Sur2]. For example, in Table 4 [*std_11*], [*std_12*], [*std_14*], [*std_15*] and [*std_16*] expressed an improvement of their satisfaction during 6W due to their membership in a larger size team. Similarly, students who worked in a team created using a RN-S in both 4W and 6W reported an increasing level of satisfaction with their experiences in the group work. On the other hand, students who worked first (i.e., in

4W) in a team that was created with HM-S did not experience a significantly higher level of satisfaction in 6W. In global terms, the satisfaction of all students (independent from the categories) increased significantly from 4W (median=2) to 6W (median=3).

6.2 Findings

In this subsection, we present the findings derived from the data analysis in terms of the issues of the RQ and their corresponding topics.

The first issue (I1) involved testing the influence of the homogenous engagement grouping approach with two different strategies at varying levels of homogeneity (i.e., RN-S and HM-S). To do so, the two grouping strategies were analysed in relation to the following topics:

T1: Active teams. - The strategy that required a lower degree of homogeneity within teams, i.e., RN-S, produced a high number of teams with small size (with only one or two active students), whereas the strategy with strong requirements of homogeneity, i.e. HM-S, produced a lower number of teams, but they had medium (three or four active students) and large (five or six active students) sizes.

T2: Activity within a team. As mentioned above, the activity carried out in teams formed using HM-S was more intense than that in RN-S. This finding is based on several indicators such as the number of active students per team, the number of messages exchanged in the group space per team, and the number of messages exchanged per student, which were higher both in 4W and 6W in teams formed with HM-S.

T3: Team Success. The number of teams that did not manage to complete and submit the collaborative assignments was higher in RN-S in both experiments (6 in RN-S vs. 2 in HM-S). It is noteworthy that all these dropped teams had a single active member.

T4: Student Satisfaction. The satisfaction with their collaborative group work was higher in both experiments (4W and 6W) for those students who worked in a team formed using HM-S. On the other hand, the students in teams formed using RN-S expressed a higher number of complaints about the presence of inactive students in their teams. Regarding the students who worked in groups created with different strategies (HM-S and RN-S) in 4W and 6W, the highest increase in satisfaction was observed among those who were in a team created with RN-S in the first experiment and in a team created with HM-S in the second one.

All these results indicate that the number of active students in a team was a key element that explains the level of interaction among team members (the number of messages exchanged) and the satisfaction of the students with the collaborative activity. The presence of various inactive students in a team may negatively affect students' satisfaction. With HM-S, which employs a greater degree of homogeneity in terms of students' level of engagement, we were able to minimize the number of teams with a single active student and obtain many large size teams. Therefore, we may conclude that stronger degrees of engagement homogeneity have a positive impact on group performance, group interactions and student satisfaction in MOOC contexts.

The second issue (I2) was related to the influence of the timing of the CL activity and the group formation on the effectiveness of the implemented grouping approaches. To that aim, the following topic was analysed:

T5: Timing Effects. - The results showed that the timing of the group formation had a strong influence on the effectiveness of the strategies: the collaborative groups functioned more successfully (higher number of messages exchanged, higher numbers of active members, and higher satisfaction with group work) in the second experiment carried out in 6W. This improvement was higher in the RN-S approach due to the

increase in the accuracy of the segregation process (i.e., having more teams created from no-show students) that caused a higher concentration of active students per team.

In both experiments, we gathered data from the platform analytics that were accumulated during the last 21 days just before the collaborative activity and used these data to feed the grouping strategies (e.g., the variable `[num_page_view]` was used to segregate students). In the first experiment (4W) this interval corresponded to the beginning of the course, while the enrolment was still open till the 14th day of the course. During this period, it was likely that student participation in the course was not stable (see Figure 6). On the other hand, the second experiment in 6W used data accumulated during the middle of the course; therefore, these data were from students with more consistent behaviour, which helped us to better identify the no-show students and the rest. This was an expected result because in MOOCs, at the beginning of the course, there exist many students browsing the course content and learning resources with no clear goals. Some of these students, although they have been active during the first weeks, may drop out. However, around the middle of the course the behaviour of the students tends to stabilize and the ratios of each behavioural pattern remain rather constant till the end of the course (Hill, 2013).

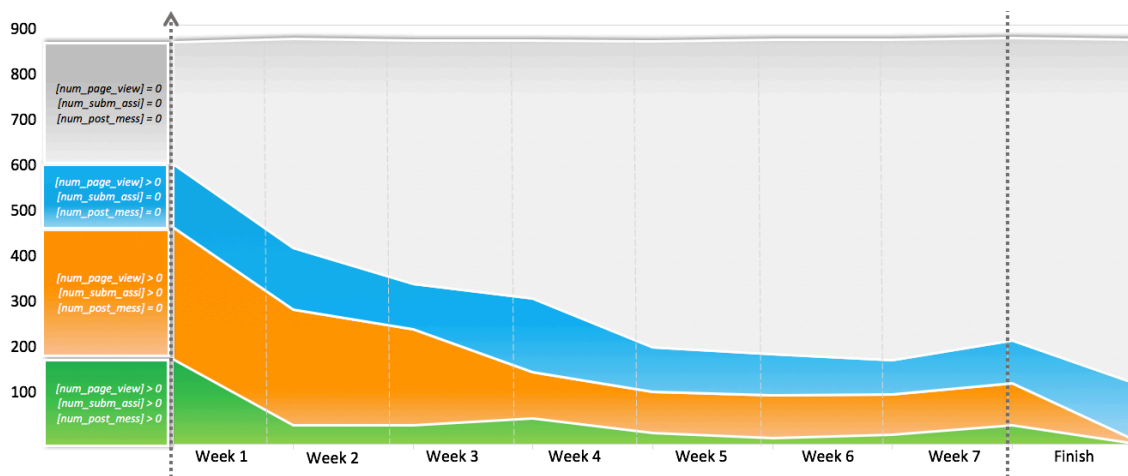


Figure 6. Patterns of engagement identified in the Translation MOOC

Considering the aforementioned analysis of the two main issues related to the RQ, we may state that HM-S was shown to be more effective in terms of team size, team performance, team interactions and student satisfaction. This strategy, which implemented a higher degree of homogeneity, was more effective when applied in the second half of the course.

7. Discussion

The study has revealed that setting homogeneity based on students' engagement led to an improvement in the performance and satisfaction of the groups. Furthermore, the number of group members that show activity and interact in the group space (i.e., what we called *team size*) seems to be a key aspect regarding the density of messages exchanged among the team members and the opinion of the students about the collaborative activity. Moreover, the presence of various inactive students in a team affects negatively, in a significant manner, the students' satisfaction.

The aforementioned findings of this study suggest that it is necessary to overcome the difficulties introduced by the variability of the open context (Dillenbourg et al., 2014) in order to create groups with the potential to interact and carry out CL in a MOOC context. To do so, the grouping strategy should aim to reduce the number of inactive students within a team by identifying and segregating no-show students (Hill, 2013; Alario-Hoyos et al., 2014). In this regard, requiring homogeneity based on students' engagement was found to be effective. Here, it must be noticed that this homogeneity in students' engagement does not impose any restriction regarding the degree of heterogeneity in the background of the students participating in a group. Therefore, with the homogeneous engagement grouping approach it is still possible to take advantage of the diversity offered by MOOCs to enrich the group interactions.

However, most previous research on group formation in MOOCs (Sinha, 2014; Spoelstra, 2015; Wen, 2016; Zheng, 2017) does not acknowledge the fact that high variability in MOOC learners' engagement is a critical issue to address in the creation of successful collaborative groups. Sinha's (2014) proposal aims to automatically group students with peers who have prior social connections. The model proposed by Spoelstra et al. (2015) applies criteria based on knowledge, preferences and personality. These proposals do not provide a solution for the students with no social connections or for those that do not answer the surveys like the case of no-shows, respectively. Furthermore, they did not present any experimental studies in order to evaluate the success of the teams created with the proposed approaches. Zheng's (2017) method addresses the problem of re-composing the groups due to dropouts. However, this method neither considered the dynamics of the course for the group formation, nor took into account other main issues in MOOC contexts such as the varying level of students' engagement and its impact in the satisfaction of the students with the collaborative experience. Wen (2016) tested her approach using a crowdsourcing service (Amazon Mechanical Turk) and acknowledged that "*Crowdsourced experiments may not represent how MOOC students will adopt or enjoy the designs*". Therefore, this study does not contribute to the solution of the issue about the varying students' engagement level that indeed exists in MOOCs.

Finally, Wichmann and colleagues (2016) compared the performance of heterogeneous and homogeneous created based on the engagement of students with forums. Their results showed that, overall, heterogeneous groups were either similarly or a bit more productive than homogeneous groups. They also found that homogeneous groups classified as high-engagement level were as or more productive than heterogeneous groups, and that students classified as low-engagement level were more

productive in homogeneous groups, suggesting that grouping less active students together makes social loafing more difficult and students participate more. However, it must be noted that this work did not take into account the engagement neither with content nor with assessment. Furthermore, it does not deal with the problems that no-show students introduce in groups. Moreover, the subjects of their study were students of two universities that would obtain credit for participating in the MOOC, so the patterns of engagement of these learners differed significantly from conventional MOOCs.

In summary, our approach is novel in considering the variability of engagement MOOC learners, and this study is the first one to provide initial evidence of the impact of different grouping approaches in group performance in a real MOOC context.

7. Conclusions and future work

In this paper, we proposed and tested a grouping approach that applies homogeneous-engagement criteria in order to create successful teams in MOOCs. Informed by this approach, two grouping strategies were developed and used in a collaborative activity deployed in a real MOOC context in different points of the course timeline. The results showed that HM-S, the strategy with higher degree of homogeneity, that grouped students with similar levels of engagement, achieved the best results in terms of group performance, group interactions and student satisfaction. Therefore, higher degrees of homogeneity on students' engagement produced more successful teams regarding the terms analysed in this study. The success of the teams further improved when the collaborative activity was in a later phase of the course because the grouping criteria used logs from the middle of the course, when the student engagement was more stable. These results contribute to the MOOC literature by highlighting the importance of establishing a homogeneous engagement base in group formation and the influence of

the timing of the collaborative activity.

This study has several limitations. First, the data used for establishing homogenous engagement base was limited because we only considered three variables regarding the engagement, and all of them have been used with the same weight in the clustering process. Therefore, other students' digital traces from the platform analytics (e.g., video logs, private messages to teachers) should be further explored to form a more rigorous approach for setting the homogeneous engagement.

Moreover, we used quantitative data (e.g., number of messages shared in the group space) when assessing the activity level of groups. Along with the quantitative indicators, future research should also look into quality of the messages exchanged among team members through qualitative data analysis methods. Furthermore, the proposed homogeneous-engagement approach was tested in a specific type of collaborative activity in this study. To further support its relevance and effectiveness, in future research this approach should be tested in other types of collaborative activities that uses different Collaborative Learning Flow Patterns (e.g., pyramid, jigsaw).

In this study, the homogeneous-engagement has been the only grouping criteria applied to form the groups. However, it also could be the first step of the whole group formation process, prior to applying other possible grouping criteria. Future work plans include the application of two levels of criteria. First level would set the homogeneous-engagement ground needed to build successful teams. Once this homogeneity has been established, we will be able to apply a second level of criteria in order to implement the pedagogical objectives of the collaborative activity, taking advantage of the massive scale and enriching heterogeneity of MOOCs. Future research should also explore other possible solutions to avoid small *team sizes* (of one or two active members) such as re-organizing teams when isolated students are detected.

Acknowledgements

This research has been partially supported by the *Junta de Castilla y León*, Spain (VA082U16), *Ministerio de Economía y Competitividad*, Spain (TIN2014-53199-C3-2-R) and *Universidad de Valladolid*. The authors thank the rest of the GSIC/EMIC research group, Roberto Castellanos, Susana Álvarez as well as the Canvas team, for their valuable ideas and support.

References

- Alario-Hoyos, C., Pérez-Sanagustín, M., Delgado-Kloos, C., Parada-G., H. A., & Muñoz-Organero, M. (2014). Delving into participants' profiles and use of social tools in MOOCs. *IEEE Transactions on Learning Technologies*, 7(3), 260–266.
- Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. *In Proceedings of the 23rd international conference on World wide web - WWW '14*, Seoul (Korea), April 7-11, 2014 (pp. 687–698).
- Blair, R. C., & Higgins, J. J. (1980). A comparison of the power of Wilcoxon's rank-sum statistic to that of student's t statistic under various non-normal distributions. *Journal of Educational Statistics*, 5(4), 309–335.
- Cheng, H. F., Yu, B., Park, Y. H., & Zhu, H. (2017). ProjectLens: Supporting Project-based Collaborative Learning on MOOCs. *In Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale* (pp. 253–256).
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.
- Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education*, 2012(3), 18.
- Dillenbourg, P. (1999). *Collaborative Learning: Cognitive and Computational Approaches*. New York, NY: Elsevier Science, Inc.
- Dillenbourg, P., Fox, A., Kirchner, C., Mitchell, J., and Wirsing, M. (2014). Massively Open Online Courses, Current State and Perspectives, *Dagstuhl Reports*, 4(3), pp. 47–61, 2014.
- Ferguson, R., & Clow, D. (2015). Examining engagement: Analysing learner subpopulations in Massive Open Online Courses (MOOCs). *In Proceedings of the 5th International Conference on Learning Analytics and Knowledge - LAK*

- '15, 16-20 March, 2015, New York (USA) (pp. 51–58). New York, New York, USA: ACM Press.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255–274.
- Hill, P. (2013). Emerging student patterns in MOOCs: A graphical view. *E-Literate, 6*. Retrieved from <http://goo.gl/X032NT>
- Isotani, S., Inaba, A., Ikeda, M., & Mizoguchi, R. (2009). An ontology engineering approach to the realization of theory-driven group formation. *International Journal of Computer-Supported Collaborative Learning, 4*(4), 445–478.
- Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third Conference on Learning Analytics and Knowledge, LAK'13*, Leuven, (Belgium), 8-12 April, 2013. (pp. 170–179).
- Konert, J., Burlak, D., & Steinmetz, R. (2014). The group formation problem: An algorithmic approach to learning group formation. In *Proceedings of the 9th European Conference on Technology Enhanced Learning (EC-TEL-2014)*, Graz (Austria), 16-19 September 2014 (pp. 221–234).
- Magnisalis, I., Demetriadis, S., & Karakostas, A. (2011). Adaptive and intelligent systems for collaborative learning support: A review of the field. *IEEE Transactions on Learning Technologies, 4*(1), 5–20.
- Manathunga, K., & Hernández-Leo, D. (2015). Has research on collaborative learning technologies addressed massiveness? A literature review. *Educational Technology & Society, 4522*, 1–14.
- McKenney, S. E., & Reeves, T. C. (2012). *Conducting Educational Design Research*. New York, NY: Routledge.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Sage.
- Mohamad, I. Bin, & Usman, D. (2013). Standardization and its effects on K-means clustering algorithm. *Research Journal of Applied Sciences, Engineering and Technology, 6*(17), 3299–3303.
- Muehlenbrock, M. (2008). Learning group formation based on learner profile and context. In E. Duval, S. Ternier, & F. Van Assche (Eds.), *Learning Objects in Context* (pp. 19–25). AACE.

- Ortega-Arranz, A., Sanz-Martínez, L., Alvarez-Alvarez, S., Muñoz-Cristóbal, J. A., Bote-Lorenzo, M. L., Martínez-Monés, A., & Dimitriadis, Y. (2017). From Low-Scale to Collaborative, Gamified and Massive-Scale Courses: Redesigning a MOOC. In Proceedings of the 5th European MOOCs Stakeholders Summit (eMOOCs 2017).
- Ounnas, A. (2010). Enhancing the Automation of Forming Groups for Education with Semantics. PhD Thesis. University of Southampton.
- Sanz-Martínez, L., Martínez-Monés, A., Bote-Lorenzo, M. L., Muñoz-Cristóbal, J. A., & Dimitriadis, Y. (2017). Automatic group formation in a MOOC based on students' activity criteria. In Proceedings of the 12th European Conference on Technology Enhanced Learning (EC-TEL-2017), Tallinn (Estonia), 12-15 September 2017 (Vol. 10474 LNCS, pp. 179–193).
- Sinha, T. (2014). Together we stand, together we fall, together we win: Dynamic team formation in massive open online courses. *In Proceedings of the 5th International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014)* (pp. 107–112).
- Spoelstra, H., Van Rosmalen, P., & Sloep, P. (2014). Toward project-based learning and team formation in open learning environments. *Journal of Universal Computer Science*, 20(1), 57–76.
- Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge: Cambridge University Press. <https://doi.org/10.1145/1124772.1124855>
- Wen, M. (2016). Investigating Virtual Teams in Massive Open Online Courses: Deliberation-based Virtual Team Formation, Discussion Mining and Support. PhD Thesis. Carnegie Mellon University.
- Wichmann, A., Hecking, T., Elson, M., Christmann, N., Herrmann, T., & Hoppe, H. U. (2016). Group formation for small-group learning: Are heterogeneous groups more productive? *In Proceedings of the 12th International Symposium on Open Collaboration*, 14:1-14:4.
- Zheng, Z. (2017). Learning Group Composition and Re-composition in Large-scale Online Learning Contexts. PhD Thesis. Humboldt-Universität zu Berlin.