Automatic group formation in a MOOC based on students' activity criteria

Luisa Sanz-Martínez, Alejandra Martínez-Monés, Miguel L. Bote-Lorenzo,
Juan A. Muñoz-Cristóbal, and Yannis Dimitriadis

GSIC-EMIC Research Group, Universidad de Valladolid, Spain.
`luisa@gsic.uva.es`

Abstract. Although there is significant evidence regarding benefits of small group collaboration in small-scale contexts, several challenges have been detected about the use of collaborative learning in MOOCs. Group formation, which is a crucial activity in order to achieve effective collaboration, is scarcely covered in MOOC platforms, which do not allow the formation of teams using criteria defined by the instructors. This paper presents an exploratory study conducted in a seven-week MOOC, comparing our group formation proposal, based on students' activity criteria, to a baseline grouping function provided by the platform. We analyse the impact of each grouping approach over group performance, group activity, and student satisfaction. The results show initial evidence about the advantages of using the criteria-based group formation approach regarding student satisfaction and group interactions.

Keywords: MOOC, collaborative learning, automatic group formation, criteria-based group formation.

1 Introduction

The increasing popularity of MOOCs (Massive Open Online Courses), as a new and powerful medium to access knowledge, has fostered many discussions within the higher education domain. Several authors are concerned about their low instructional quality [18] or their high dropout rate [21], while others highlight the variety of research challenges triggered by the massive scale feature [8]. Some of these challenges are related to the promotion of social interactions that can generate knowledge [17] or the development of new pedagogical approaches which take advantage of the benefits of large scale [28].

Over the last decades, active pedagogies, such as Collaborative Learning (CL), have been largely studied at small-scale educational contexts. These studies have shown positive effects, *e.g.*, that collaboration enriches learning with social and cognitive dimensions that maintain student motivation and elicit verbal communication [26].

Currently, most MOOCs follow a behaviorist pedagogical approach where the instructors add the educational content to the course stream and the students self-assess their learning with questionnaires [7], limiting the interaction between

participants and instructors to discussion forums. However, since the appearance of the first MOOC in 2008 (Connectivism and Connective Knowledge - CCK08), many authors have explored the benefits of using active pedagogies in this type of courses claiming that these pedagogies have a positive influence in various facets such as students' engagement [10]. The analysis of collaboration among students shows that social participation has a positive influence into student performance [1]. Some studies have focused on the students' preferences [11] finding that learners demand more opportunities for discussing in groups. Nevertheless, the inclusion of effective collaboration in MOOCs is still a challenge [15] due to the specific characteristics of the MOOC context. At the moment, collaboration and social interactions are mostly pragmatically limited to peer reviews, forum interactions [4] or external social tools [2]. The massive scale and its variability, caused by latecomers and dropouts, the heterogeneity of the enrolled students or their low engagement level [3] hinder effective implementation and uptake of CL strategies.

Several studies on CL have showed that group formation is a crucial factor to put in practice collaboration [20,23] because successful collaboration depends, to a large extent, on the suitability of the peers included in the group [14,13]. However, group formation presents particular difficulties at massive scale that deserve a deeper analysis. Thus, we decided to address this question by investigating the issues involved in the group formation problem at massive and variable scale. To that aim, we deem it necessary to further study the criteria that can be used in group formation in MOOC contexts and analyze the impact of these group formation strategies on the groups themselves and their members. Based on the outcome of this study, we aim to provide support to teachers interested in introducing collaborative activities performed in groups in MOOCs. In previous studies [27], we have proposed a framework that considers the factors that could be taken into account in group formation, when the scale is large and suffers significant variations during the course enactment. Based on this framework, appropriate advice for MOOC design and supporting tools for deployment may be provided.

In this paper, we present an exploratory study, where a criteria-based group formation approach was compared to a baseline grouping function provided by the platform that hosts the intervened MOOC. In our proposal, students were grouped in homogeneous groups based on their previous activity in the course. We analyzed the impact of each grouping approach over group performance, group activity, and students satisfaction. This analysis seeks to show differences, benefits and drawbacks of each grouping approach.

The rest of the paper presents, firstly, an analysis of the group formation problem delving into the scalability issues. Then, we continue explaining the study carried out in a MOOC deployed in the Canvas Network platform. We conclude showing the experiment results and exposing our conclusions and future work.

2 Group Formation Scalability

A basic definition of group formation in educational contexts could be "to put students together in groups with an educative purpose" [23], but effective CL usually requires planning in advance the collaboration to foster the relevant interactions that can better promote learning [9]. Group formation is an essential activity in CL and the method used to define the group composition is a critical function in Computer Supported Collaborative Learning (CSCL) environments [13]. The adequacy of the peers included in a team is a major factor for effective collaboration, and the group composition may affect the group performance and the individual student benefits [14]. Poorly formed groups can lead to many possible negative peer group influences: conformity, anti-intellectualism, intimidation, and leveling-down of quality, which lead to detrimental effects for learning [23]. In her thesis, Ounnas [23] exposed three approaches that can be used to create groups in educational contexts:

- **Random selection of groups**, where the formation is initiated by the teacher who assigns students randomly to groups. It is a simple way of forming groups because there are no constraints to enforce.

- **Self-selection groups**, where students decide the group they want to join and they can negotiate the peers to work with. The allocation of members requires the identification of potential peers which meet the requirements to join the group. This approach is commonly used in communities and networks where participants join together based on common interests. It can also be used in teams where students select their teammates based on interests, (*e.g.*, friendship or confidence, technical capabilities, skills to complete the task). This type of groups have a tendency to homogeneity.

- **Teacher selected groups**, also known as **criteria-based grouping**. This is a very popular approach in task-oriented grouping. The teacher's criteria can be applied in different ways, so that formed groups may have: (i) an homogeneous structure, including members with similarities regarding the criteria, (ii) an heterogeneous structure, including members with differences regarding the criteria, or (iii) a structure based on rules, i.e. several constraints are applied that group members have to meet.

Criteria-based group formation has been largely explored at small-scale educational environments [20,24,12,13], employing different types of criteria (e.g., student's profile, student's learning style), targeting both homogeneity and heterogeneity, as well as applying different types of rules. In the CSCL field, several tools and systems have been proposed to support automatic group formation using different techniques and algorithms [16]. However, MOOCs have particular characteristics, such as their massive and variable scale which hamper a direct extrapolation of conclusions derived in small-scale studies.

Due to the interest for including CL in MOOCs, several authors have tackled the group formation problem in these contexts [29,30,32,31,5] addressing the challenge through different perspectives. These perspectives include a variety of criteria (*e.g.*, knowledge, personality, preferences, affinities, location, motivation), grouping approaches (e.g., criteria-based homogeneity or heterogeneity,
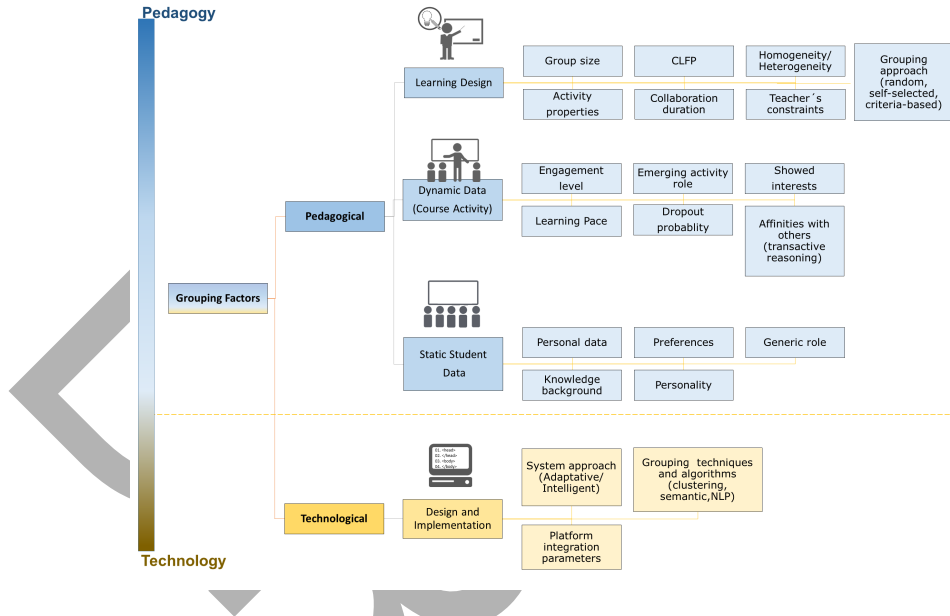
Fig. 1. Categories and factors to be considered for group management in MOOCs.

random grouping) and technological aspects (*e.g.*, social network metrics, natural language processing, classification algorithms) which suggests there are different factors that can be considered for group management in MOOC contexts. Figure 1 shows a hierarchical representation included in our previous framework proposal [27], which depicts four dimensions where grouping factors can be framed: (i) learning design, (ii) student's static data, (iii) course-activity dynamic data and (iv) technological implementation).

The **course-activity dynamic data** can be obtained from the course analytics and may allow us to know *when* and *how* the students work, so these data can reflect some particular features (*e.g.*, irregular level of engagement, variable learning paces) which distinguish MOOCs from other contexts. Therefore, these course-activity dynamic data may be interesting criteria to be considered in the group formation process.

Currently, only a few platforms offer facilities to collaborate in teams (i.e., Canvas, NovoEd, edX), while the students of other platforms (e.g., Coursera, Udacity, FutureLearn), which do not provide these group facilities, have even formed external networks to meet and create study groups. The grouping facilities offered by the aforementioned MOOC platforms include features for students to self-select the teams to join (mostly by the topic). The group may be created by the teacher or by the students. This method leaves out many students who can't manage to join a team [31]. Some platforms also allow teachers to assign manually the members of each group, but this solution is not efficient in a course with a massive number of registered students. Canvas Network includes a func-

tion for splitting students into random teams. All students are distributed in groups with equal number of members. This is a convenient way of ensuring every student will belong to a team. Nevertheless, the criteria-based approach for grouping, which is the preferred method at small-scale context for its pedagogic capabilities, is not covered by MOOC platforms at the moment.

3 Description of the Study

Our research work follows a Design Science Research Methodology (DSRM) [25]. The study reported in this paper is part of the initial iteration of the process. Its main goal is to evaluate the initial ideas of the proposal in order to improve them in the next iterations. We collected quantitative and qualitative data, in order to gain a deeper understanding of the results of the intervention by means of complementarity. This approach is a consequence of our underpinning pragmatic worldview, centered in the problem and oriented to real world practice [6].

3.1 Context

The course was initially designed by teachers of the Faculty of Translation at University of Valladolid and its topic was an introduction to translation from Spanish to English over economic and financial texts. It was originally conceived as an instructor-led MOOC of seven weeks. We formed a co-design team composed of instructors and researchers, and this team redesigned the course to incorporate CL activities in order to identify the emerging challenges [22]. To meet this end, a community glossary and several peer reviewed translation tasks were integrated as optional activities. Moreover, the main collaborative activity included in the MOOC learning design, basis for our experimental study, was a compulsory task presented in the fourth week (see section 3.2 for a full description). All mandatory activities should be completed (one per week) to obtain the certificate, although no grades were included in the assessment of the students.

The course was deployed in the Canvas Network platform and began on February the 6th, 2017. The total number of students enrolled was 1031, but only 875 remained registered when the course ended. Two surveys were employed: an optional welcome survey during the first week, that was completed by 668 students, and a mandatory final satisfaction survey completed by 152 (17,37% of the remaining registered students). 130 students applied for the certificate (12.61% of the initially enrolled students or 14.86% of the students registered at the end of the course).

3.2 Collaborative Activity

We used different data gathering techniques (i.e., questionnaires, interviews and meetings with the MOOC's teachers, and observation) to codesign the compulsory collaborative activity, which was the basis of the grouping experiment. The activity consisted in terminology extraction from some given texts in teams of

six. Each team should create a group artifact including 20 economic or financial English terms and their corresponding Spanish translation referencing the source. The teams should use some of the group-oriented Canvas platform tools (*i.e.*, discussion forums and announcements) for organizing their work, sharing opinions, discussing and reaching agreements in order to select the required terms and choose a spokesman who would be in charge of the task submission. Finally, the activity would be considered as completed, when all members of a team perform an individual revision of the artifact produced by another team. This way, the non-active members of a team would not pass the activity, even if the task was submitted by a member of their group, since the non-active members did not carry out the individual review. The task was assessed as passed/not passed for all the students that completed it and there were no individual or group grades.

3.3   Intervention

This subsection describes the main decisions taken for the design of the experiment. One of the most important decisions was the selection of the criteria to be used for creating the experimental groups. We used dynamic factors (*i.e.*, data from the activity of the students in the platform) to respond to our research question regarding the relevance of these data to reflect some peculiarities of the context (*i.e.*, the variable engagement level). Therefore, we chose three variables to cover three aspects regarding the student engagement level [10]:

- **page views**, to measure their activity,
- **submitted tasks**, to estimate their commitment, and
- **posted messages**, to reveal their active participation.

Another major decision was the application of homogeneity over the criteria instead of heterogeneity. The underlying reason was that, taking into account the group size (six members) and MOOC statistics in literature (5-15% of completion rates), heterogeneity over student's activity criteria could be very similar to a random grouping (feature already covered in the Canvas platform) and could result in many teams with only one active student. The fact that the activity was assessed as pass/not pass and there were no grades strengthened this decision, because this type of homogeneity would have affected the grades.

For the composition of the control group, we chose random grouping because that option can be performed automatically in Canvas and guarantees that all students would be included in a group. However, the fact that in our approach the students with an activity profile type of no-shows [1] were clustered together could be a big advantage over the random teams, where the no-shows students would be spread over the teams. Therefore, we decided to improve the baseline to compare with in order to obtain richer conclusions about the impact of using a criteria-based approach for grouping. Hence, in the control group, we seggregated the students with zero page views by grouping them together prior to the creation of the random teams.

The algorithm selected for implementing the homogeneous grouping was k-means clustering because it is a well known, effective technique that works with big datasets [31]. We combined it with a balancing algorithm to obtain clusters with exactly the same number of members (same size k-means variation[1]). To carry out the experiment the following steps were followed:

1. Finding out the statistical distribution of the selected variables (page views, submitted tasks and forum messages). Using the Kolmogorov & Smirnov, and the D'Agostino & Pearson tests, we found out that all three variables followed a non-gaussian distribution.
2. Data preprocessing. Prior to the clustering process the data was standardized in order to assign the same weight to the three selected variables (page views had a dimension much bigger than the other two) as recommended in [19].
3. Creation of two subsets (the experimental group and the control group) checking their uniformity regarding the variables used as grouping criteria. As a consequence of the non-gaussian distribution of the three variables, a Wilcoxon test was selected to verify that the subsets do not differ regarding them. The array of students was shuffled and splitted in two equal size subsets until the Wilcoxon test returned a p value greater than 0.5 in the three variables used as grouping criteria (if $p < 0.05$, the samples would be different with 95% confidence; if $p > 0.05$ we cannot say that the samples differ; we required a $p > 0.5$ to strengthen the non-difference between samples).
4. Creation of the teams in the control group. Firstly, students with zero page views were seggregated, grouping them together in 11 teams and then, the rest of the students in the control group were distributed randomly in 70 six-members teams.
5. Creation of the teams in the experimental group. The selected clustering algorithms were used to obtain 81 clusters of six members based on homogeneity on the three standardized variables.

3.4 Analysis Methods

To measure the intervention effects we collected data from several sources (i.e., the platform API, the final satisfaction survey and the communications between students and teachers during the collaborative activity) in order to triangulate and complement the results. We monitored team performance during the activity retrieving data about: (i) messages exchanged in each group space, (ii) active participants in each team, and (iii) teams that complete the task submission. On the other hand, the messages that students sent to teachers regarding this activity were collected. Finally, after the end of the activity, we gathered quantitative and qualitative data about student satisfaction by means of open and close ended questions in a survey.

We analyzed the aforementioned data to find out the differences between the experimental (criteria-based) and the control (random) groups regarding: (i) active teams, (ii) active participants per team, (iii) interactions within a team, (iv)

---

[1] https://elki-project.github.io/tutorial/same-size_k_means

task completion rate, (v) student complaints, and (vi) student satisfaction level. This analysis may provide initial evidence about the impact of using criteria-based group formation in order to achieve effective CL in MOOC contexts.

4 Results

4.1 Analysis of the activity of the teams gathered from the platform

After the end of the activity, we collected available data through the Canvas LMS API about the activity within each team. A summary of the gathered information is shown in Table 1. We captured data about the total number of messages (posts and replies in the group discussion forums and announcements) exchanged within each team, as well as the students that produced these messages, in order to detect the team members that were indeed participating in the activity.

Data gathered from the API	Control	Experimental
Teams with registered activity	47/81=58.02%	25/81=30.86%
Teams that submitted the task	46/81=56.79%	26/81=32.1%
Teams with activity which do not submit the task	4	1
Teams without activity which submit the task	3	2
Total number of messages	300	372
Total number of active users	76	78
Average number of messages per active user	3.95	4.77
Standard deviation of messages per active user	2.69	3.67
Average number of messages per active team	6.38	14.88
Standard deviation of messages per active team	5.87	14.92
Median number of messages per active team	3	10

Table 1. Data about teams' activity gathered from the Canvas LMS API.

The method used for the creation of the two subsets ensured a similar number of active users in both subsets (76/78). Due to the homogeneous activity criteria in the experimental group, students with a low activity level were joined together, giving as a result 56 teams with no registered activity (vs. 34 in the control group, out of which 11 were formed in the prior seggregation process for no-shows students). This is an expected result, since there were a big quantity of inactive or low-activity students in the MOOC, and therefore homogeneous groups composed by students with a previous low level of activity will tend to even show less activity, due to negative interdependence. On the contrary, since active users were scattered in the random process, the randomly assigned groups may include some dispersed high-activity members, who will show some activity, even in the presence of inactive team teammates. Nevertheless, there were also 25 experimental teams with a significantly more intense exchange of messages (average of 14.88 messages per team vs. 6.38 in the control group). Moreover, the active users in the experimental group sent a higher number of messages each (mean of 4.77 vs. mean of 3.95). In this case, the homogeneous teams with active
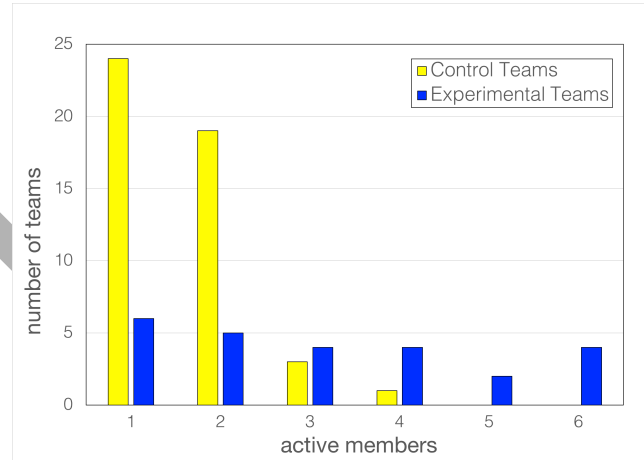
Fig. 2. Number of teams (y axis) with a registered number of active members (x axis).

members have higher chances of developing a higher activity due to the positive influence of their teammates. In the control group there were four teams (vs. one in the experimental group) with registered activity which did not manage to complete the task and therefore, could not obtain the course certificate. All these teams had registered a single active member, which suggests that these students might have felt isolated due to negative interdependence and their motivation regarding the course decreased.

Team size (active members)	1			2			3			4			5			6		
Total/Control/Experiment.	T	C	E	T	C	E	T	C	E	T	C	E	T	C	E	T	C	E
$Number of Teams$	31	24	6	24	19	5	7	3	4	5	1	4	2	0	2	4	0	4
$Msg/ActiveUser - Avg$	1.8	1.9	1.3	4.6	5.0	3.2	5.5	5.0	5.8	4.5	5.3	4.2	5.5	−	5.5	5.8	−	5.8
$Msg/ActiveUser - StdDev$	1.0	1.1	0.5	2.7	2.8	2.1	3.4	2.2	4.1	3.0	2.2	3.2	3.5	−	3.5	4.5	−	4.5

Table 2. Data about teams and users regarding *team size* (num. of active members).

In Figure 2 a significant observation is depicted: the high number of teams with only one or two active participants in the control group (almost fourfold than in the experimental group). We can also observe that full active teams (with five or six active members) can only be found in the experimental group. In this case, due to the homegeneity criterion of the experimental group, it is more likely that all members of some groups may be active. This result confirms that homegeneous group formation may favor some groups, since the most active students are grouped together.

The aforementioned conclusion is further supported by Table 2 which presents data structured according to the number of active participants registered in the team, which we called *team size*. The average number of messages per active user increases with the size of the team with a correlation coefficient of 0.78 stating a strong positive correlation.

4.2 Analysis of the students opinions

A summary of the results of the closed-ended questions of the final survey regarding the collaborative activity on the fourth week is shown in Table 2. The responses *Agree* and *Strongly Agree* in the survey have been agreggated in the category **Agree** in the table, and the responses *Disagree* and *Strongly Disagree* have been agreggated in the categorie **Disagree**. The *Don't Know/No Answer* responses are not included in the table.

Subset	Control		Experimental	
	Agree	Disagree	Agree	Disagree
Satisfaction with the collaboration in my team	35.3%	59.1%	55.0%	36.6%
Inactive students in my team hindered collaboration	78.9%	12.7%	52.1%	32.4%
Inactive students in my team affected negatively to my satisfaction	57.7%	31.0%	40.9%	38.0%
Collaboration in this activity enhaced my motivation	42.3%	42.3%	40.9%	38.0%
Collaboration in this activity enhaced my participation	60.5%	26.7%	67.6%	19.7%

Table 3. Quantitative data collected from the final satisfaction survey.

The students of the homogeneous teams (experimental group) are more satisfied with the collaboration carried out in their teams, while the students in the random teams complain about the presence of inactive students in their group. On the other hand, the collaborative activity was valued as positive regarding participation for both subsets, while collaboration had a neutral effect on motivation for both subsets. These observations confirm previous findings (appeared in the communications between students and teachers during the activity development) regarding the negative effect of inactive students in groups (something that is less prominent in homegenous groups), as well as the positive effect of collaborative activities on participation, even in MOOC contexts.

A finer analysis regarding the *team size* (number of active participants registered in the team) shows that the survey respondents belonging to teams with five or six active members (32 students) were the most satisfied with the collaborative activity (expressing high satisfaction in 75% of the cases), and the survey respondents belonging to teams with one or two active members (68 students) were the less satisfied with the collaborative activity (expressing dissatisfaction in 69,12% of the cases). This result reinforces the need to find the best strategy (based on the most suitable criteria for group formation) in order to include several active members in each group.

The final satisfaction survey also included open-ended questions about the mandatory collaborative activity where the students could explain the aspects they most or less liked of this activity. We used this information together with the messages that students sent to the teachers in the Canvas Network platform to perform an initial content analysis aiming to gain a deeper understanding of what happened in the experiment.

The majority of complaints came from students who were the only active learner of a given group. In many cases the students in teams with one or two active members expressed frustration due to the lack of participation in their group, as well as feelings of having lost the opportunity of an enriching activity. We illustrate the previous observation through a set of comments expressed by students that belonged to groups with only one or two active students:

"I wish my teammates would have been more active, or at least they had contacted me". "My colleagues were noted for their absence. At least they could have introduced themselves and said that they would not participate instead of keep us waiting to see if they appeared". "No teammates showed up, although I sent them messages in the forum asking for their availability. I should say that it was an especially upleasant experience." "In fact, the most interesting aspects of the activity were related to its content and not to the collaborative work, since my teammates did not show any interest for the activity".

The most positive comments belonged to students in teams with five or six active members who expressed their satisfaction of having the opportunity of meeting their mates, helping each other and knowing different points of view. We provide below a characteristic set of comments expressed by this type of students in teams with five or six active members:

"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". "This group has enchanted me because we have all collaborated and we have fit perfectly, something difficult to achieve". "Everything has been very simple. Each one has contributed the terms that he could and when he could, without any pressure". "Although we are partners from all over the world, we managed to finish the activity and maintain a good communication". "What I liked the 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. In my group there were no inactive students". "We were able to distribute the work and see the way that the other colleagues had to work. We learned from each other".

Teams with three or four active members registered more positive comments than negative ones. On the positive side, the students of these teams show their satisfaction in similar terms than the students of full active teams, but in the negative side they express some frustration for the absence of some teammates.

5   Conclusions and Future Work

In this paper we have presented the results of a study in which a small group collaboration activitiy was introduced in a MOOC. More concretely, a criteria-based group formation strategy was compared to the baseline option of random assignment of students to groups provided by the learning platform. On the other hand, we used the dynamic data of previous activity of each student in the course, since such type of data reflects better the large and varying scale context of MOOCs. Such study has provided some insights regarding the introduction

of small group collaboration in MOOCs, and the relative advantages of two group formation strategies. After analyzing the results, we can conclude that the new strategy selected for the creation of the teams (homogeneous groups based on prior activitiy) had a positive impact on student satisfaction and group interactions. We also observe a slight positive impact regarding students dropout.

A key aspect regarding the measures of participation and regarding students' satisfaction is the number of active members in the team (what we called *team size*). Teams with five or six active members registered the most intense activity and the most satisfied students of the experiment. The correlation between the number of messages per active user and the *team size* was relatively high (0.78), which indicates a strong correlation. Therefore, the higher the *team size* the more active the members, probably as an effect of the positive interdependence. As expected, teams with five or six active members promote collaboration, registering the highest number of interactions and a hight student satisfaction. This type of teams were only achieved through the grouping strategy that promoted homegeneous groups based on the dynamic activity data. On the other hand, teams with only one active member did not allow collaboration and generated student frustration, giving as a result several cases of dropout. This fact occurred fourfold less frecuently in the homogeneous teams.

This experience has served to gain insight about grouping solutions which may run smoothly at massive or variable scale. The findings of this study may serve as a seed of the knowledge base to support MOOCs teachers by giving them advice regarding the course design and by developing tools which can help them in the design and deployment of group activities.

Given the iterative nature of the DSRM methodology, we plan to carry out new iterations, in which we plan to study several aspects, such as: (a) other alternatives of criteria-based strategies, (b) other types of data (dynamic or static) according to the factors included in our framework, (c) the impact and usability of a user interface for the instructors - instructional designers regarding the criteria to be used for group formation. These studies will be performed in the context of real MOOCs, that have been scheduled in the upcoming months.

Acknowledgements

References

1. Alario-Hoyos, C., Pérez-Sanagustín, M., Delgado-Kloos, C., Parada-G., H.A., Muñoz-Organero, M.: Delving into participants' profiles and use of social tools in MOOCs. IEEE Transactions on Learning Technologies 7(3), 260–266 (2014)

2. Alario-Hoyos, C., Pérez-Sanagustín, M., Delgado-Kloos, C., Parada-G., H.A., Muñoz-Organero, M., Rodríguez-De-Las-Heras, A.: Analysing the impact of built-in and external social tools in a MOOC on educational technologies. In: Proceedings of the 8th European Conference, on Technology Enhanced Learning (EC-TEL-2013), Paphos (Cyprus), 17-21 September 2013. vol. 8095 LNCS, pp. 5–18 (2013)
3. Blom, J., Li, N., Dillenbourg, P.: MOOCs are more social than you believe. eLearning Papers 33, 1–3 (2013)
4. Brinton, C.G., Chiang, M., Jain, S., Lam, H., Liu, Z., Wong, F.M.F.: Learning about social learning in MOOCs: From statistical analysis to generative model. IEEE transactions on Learning Technologies 7(4), 346–359 (2013)
5. Cheng, H.F., Yu, B., Park, Y.H., Zhu, H.: ProjectLens : Supporting Project-based Collaborative Learning on MOOCs (2017)
6. Creswell, J.W.: Research design: qualitative, quantitative, and mixed methods approaches. SAGE Publications (2014)
7. Daniel, J.: Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility. Journal of Interactive Media in Education 2012(3), 18 (2012)
8. Dillenbourg, P., Fox, A., Kirchner, C., Wirsing, M.: Massive Open Online Courses: Current State and Perspectives. Tech. Rep. 1 (2014)
9. Dillenbourg, P., Tchounikine, P.: Flexibility in macro-scripts for computer-supported collaborative learning. Journal of Computer Assisted Learning 23(1), 1–13 (2007)
10. Ferguson, R., Clow, D., Beale, R., Cooper, A.J., Morris, N., Bayne, S., Woodgate, A.: Moving through MOOCS: Pedagogy, learning design and Patterns of Engagement. In: Proceedings of the 10th European Conference on Technology Enhanced Learning (EC-TEL-2015), 15-18 September, 2015, Toledo (Spain). vol. 9307, pp. 70–84. Springer Verlag, Toledo, Spain (2015)
11. Grünewald, F., Meinel, C., Totschnig, M., Willems, C.: Designing MOOCs for the Support of Multiple Learning Styles. In: Proceedings of the 8th European Conference on Technology Enhanced Learning (EC-TEL-2013), Paphos, (Cyprus), 17-21 September 2013. p. 12 (2013)
12. Inaba, A., Supnithi, T., Ikeda, M.: How Can We Form Effective Collaborative Learning Groups? In: Proceedings of the 5th International Conference on Intelligent Tutoring Systems (ITS-2000), Montreal, (Canada), 13-23 June 2000. pp. 282–291 (2000)
13. Isotani, S., Inaba, A., Ikeda, M., Mizoguchi, R.: An ontology engineering approach to the realization of theory-driven group formation. International Journal of Computer-Supported Collaborative Learning 4(4), 445–478 (2009)
14. Konert, J., Burlak, D., Steinmetz, R.: 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 (2014)
15. Mackness, J., Mak, S.F.J., Williams, R.: The Ideals and Reality of Participating in a MOOC. In: Proceedings of the 7th International Conference on Networked Learning, Aalborg, (Denmark), 3-4 May 2019. vol. 10, pp. 266–274 (2010)
16. Magnisalis, I., Demetriadis, S., Karakostas, A.: Adaptive and Intelligent Systems for Collaborative Learning Support: A Review of the Field. IEEE Transactions on Learning Technologies 4(1), 5–20 (2011)
17. Manathunga, K., Hernández-Leo, D.: Has Research on Collaborative Learning Technologies Addressed Massiveness? A Literature Review. Educational Technology & Society 4522, 1–14 (2015)

14

18. Margaryan, A., Bianco, M., Littlejohn, A.: Instructional Quality of Massive Open Online Courses (MOOCs). Computers & Education 80, 77–83 (2015)

19. Mohamad, I.B., Usman, D.: Standardization and its effects on K-means clustering algorithm. Research Journal of Applied Sciences, Engineering and Technology 6(17), 3299–3303 (2013)

20. Muehlenbrock, M.: Learning Group Formation Based on Learner Profile and Context. In: Duval, E., Ternier, S., Assche, F.V. (eds.) Learning Objects in Context, pp. 19–25. AACE (2008)

21. Onah, D.F., Sinclair, J., Bollat, R.: Dropout rates of massive open online courses : behavioural patterns. In: Proceedings of the 6th International Conference on Education and New Learning Technologies, Barcelona, Spain, 7-9 Jul 2014. pp. 14–15 (2014)

22. Ortega-Arranz, A., Sanz-Martínez, L., Alvarez-Alvarez, S., Muñoz-Cristóbal, J.A., Bote-Lorenzo, M.L., Martínez-Monés, A., Dimitriadis, Y.: From Low-Scale to Collaborative, Gamified and Massive-Scale Courses: Redesigning a MOOC. In: Proceedings of the 5th European MOOCs Stakeholders Summit (eMOOCs 2017) (2017)

23. Ounnas, A.: Enhancing the Automation of Forming Groups for Education with Semantics. Phd thesis, University of Southampton (2010)

24. Paredes, P., Ortigosa, A., Rodriguez, P.: A Method for Supporting Heterogeneous-Group Formation through Heuristics and Visualization. Journal of Universal Computer Science 16(19), 2882–2901 (2010)

25. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. Journal of Management Information Systems 24(3), 45–77 (2007)

26. Roschelle, J., Teasley, S.D.: The construction of shared knowledge in collaborative problem solving. In: O'Malley, C. (ed.) Computer-Supported Collaborative Learning, pp. 69–97 (1995)

27. Sanz-Martínez, L., Dimitriadis, Y., Martínez-Monés, A., Alario-Hoyos, C., Bote-Lorenzo, M.L., Rubia-Avi, B., Ortega-Arranz, A.: Influential factors for managing virtual groups in massive and variable scale courses. In: 2016 International Symposium on Computers in Education (SIIE). pp. 1–4 (2016)

28. Sharples, M., Delgado-Kloos, C., Dimitriadis, Y., Garlatti, S., Specht, M.: Mobile and Accessible Learning for MOOCs. Journal of Interactive Media in Education pp. 1–8 (2014)

29. Sinha, T.: 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 (2014)

30. Spoelstra, H., Van Rosmalen, P., Sloep, P.: Toward Project-based Learning and Team Formation in Open Learning Environments. Journal of Universal Computer Science 20(1), 57–76 (2014)

31. Wen, M.: Investigating Virtual Teams in Massive Open Online Courses: Deliberation-based Virtual Team Formation, Discussion Mining and Support. Phd thesis proposal., Carnegie Mellon University (2015)

32. Zheng, Z., Vogelsang, T., Berlin, B., Pinkwart, N.: The impact of small learning group composition on student engagement and success in a MOOC. In: Proceedings of the 8th International Conference of Educational Data Mining. pp. 500–503 (2015)