

Semantically Meaningful Cohorts Enable Specialized Knowledge Sharing in a Collaborative MOOC

Stian Håklev¹, Kshitij Sharma², Jim Slotta³, Pierre Dillenbourg¹

1. École Polytechnique Fédérale de Lausanne, Switzerland

2. Norwegian University of Science and Technology

3. University of Toronto, Canada

Abstract. This study presents an analysis of a MOOC on inquiry and technology for in-service teachers, which was designed to scaffold multiple disciplinary knowledge communities through common weekly themes, and course-long collaboration scripts happening at different social planes. Using our course design to inform the design of the analysis, we examine how the discourse in each semantically meaningful cohort (Special Interest Groups, SIGs) is indexed to the weekly themes, and develops these themes in areas informed by the discipline, and by the group dynamics. We show that SIG membership influences individual contributions, and that more cohesive disciplinary SIGs are correlated with higher quality student work.

Keywords: inquiry-based learning · MOOCs · Massive Open Online Courses · learning analytics · multi-level analysis · CSCL

1 Introduction

Massive Open Online Courses (MOOCs) attract large numbers of students with very diverse backgrounds and interests. The experiences, ideas, and collective energy of these students could potentially contribute a large amount to the learning experience, however the very number of students also represents an almost insurmountable challenge for teachers wishing to implement a knowledge-community approach in their courses.

Some MOOC platforms offer course cohorts as a solution — assigning students to random groups, and making forums local to each group, as a way to avoid information overload. We posit that grouping students based on their specific interests, and giving them access to rich and diverse knowledge tools, not just forums, can significantly improve the quality and relevance of their discussions.

In this study, we will present an analysis of a MOOC for in-service teachers which ran on the EdX platform. The course, which attracted around 8,500 registrations, and around 2,200 active users, focused on integrating technology and inquiry into the lesson design process, and used a large amount of custom

activities to enable both crowd-sourcing and small group collaboration, with the goal to support transfer from theoretical concepts to students' professional lives.

Too much learning analytics research on MOOCs has treated every course as interchangeable, whereas we argue that taking into account the instructional design and structure of the MOOC is key to understanding the individual and collaborative processes of students [12]. In this paper, we attempt to analyze a MOOC with nested social structures, and complex interactions between multiple pedagogical scripts.

Our main goal is to use learning analytic approaches to explore how this intentional theoretically informed course design actually contributed to structure students' conversations and collaborative work. We also begin to explore factors contributing to higher quality artefacts, although we are not making the argument that this is a valid indicator for individual student learning.

Below, we will present some of the design features relevant to the subsequent analysis (for a more in-depth exploration of the course design, see [6]).

1.1 Course Design

The course design was an attempt at mapping the Knowledge, Community and Inquiry framework [16] to a large-scale setting, which meant ensuring that students' knowledge production was indexed to a knowledge structure representing the learning goals of the course. We conceived of the course design as a matrix, combining specific weekly content themes with course-long collaboration scripts. Students joined a Special Interest Group (SIG), for example "secondary science" with a few hundred others, and within the SIG, had the option of engaging in a lesson design project with up to six others.

As Figure 2 shows, students engaged in collaborative scripts on multiple levels of granularity (whole class, SIG, and design group) beginning in the pre-course lounge, and continuing throughout the course. These scripts were then tied together through the weekly themes, which permeated all scripted activities during a given week, with the scaffolded design of a lesson in small groups of 3-6 students providing the organizing principle throughout the course.

For a given week on the theme of "collaborative learning", a student would begin by watching videos (lectures and mini-documentaries) about collaborative learning, followed by a personal reflection about collaboration (related to their own teaching practice), before responding to several prompts related to collaboration in their SIG. He or she would then look at the in-progress lesson designs in their SIG, and add a review comment addressing how this team could incorporate more collaboration into their design. Finally, a student who was a member of a Lesson Design group would log into their Collaborative Workbench, see the weekly prompt (related to collaboration, see Figure 2), as well as the peer review comments from all their peers, and continue work on improving their lesson design document, informed by all the preceding activities.

The collaborative workbench interface where students worked on their lesson designs featured weekly prompts, incoming information from the community (for example review comments), an Etherpad (collaborative scratchpad) for notes

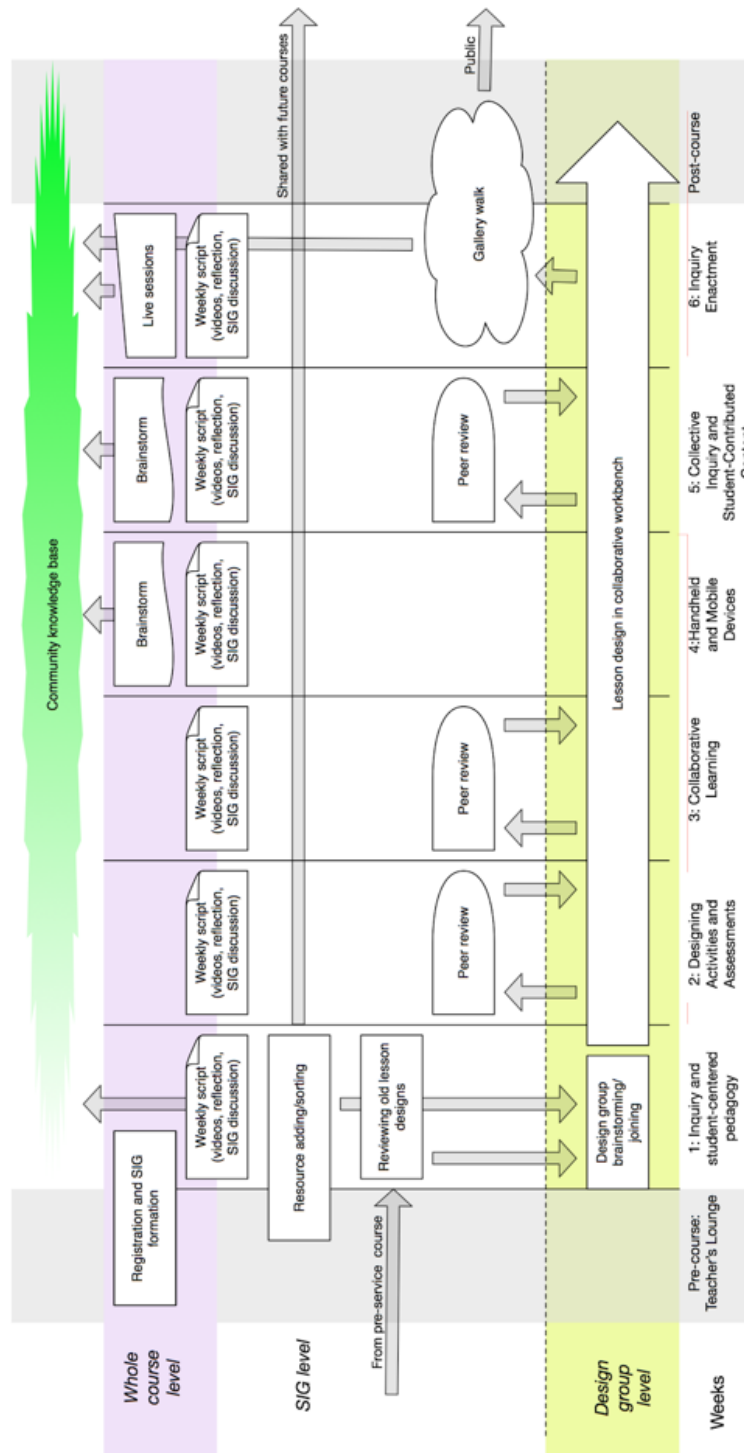


Fig. 1. Graphical depiction of interconnecting MOOC scripts.

Fig. 2. Lesson design outline

1. Describe a typical classroom where this lesson might be enacted.
2. Describe the major theme of the lesson.
3. What are the learning goals of the technology-enhanced lesson?
4. Some aspects of the design (complete any that are relevant)
 - (a) Student-Centered Design
 - (b) Peer Collaboration
 - (c) Use of Handheld or Mobile Computers
 - (d) Supporting Equity and Diversity
5. What is the activity structure of the lesson?
6. Assessment notes.
7. Enactment notes.
 - (a) Ethics or enactment concerns

and ideas private to the group, and a wiki where the group authored the actual lesson design. The wiki page was gradually seeded with template headers which increased in sophistication each week (from learning goals, and activity structure, to technology integration, and assessment), and was exposed to the rest of the class for weekly reviews (see Figure 2).

In this way, students began each week by receiving abstract and general ideas from the MOOC videos, and continued to engage with these ideas individually and in large and small groups, progressively making them more concrete, and more applied to a specific discipline and a specific lesson design, in the process increasing both the relevance and transferability of abstract concepts.

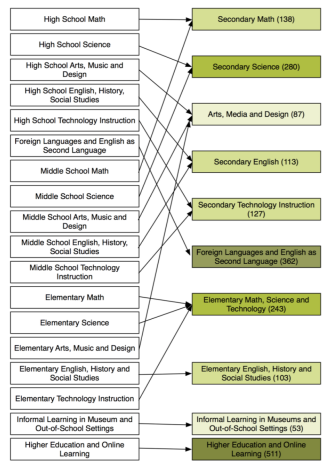


Fig. 3. Special Interest Groups: Initial and final configurations.

1.2 Special Interest Groups

Several weeks before the course had officially begun, we opened the “teachers’ lounge” — a virtual site for teachers to congregate, fill out a survey about their professional interests, and begin contributing resources to the course community. Based on this information, we produced a list of 18 suggested Special Interest Groups, designed to balance homogeneity (teacher discipline/age level taught) and number of participants. MOOC participants were invited to choose one of these SIGs, and based on their actual choices, we then combined a number of SIGs to rebalance the number of group participants (see Figure 3).

The main focus of the course was science and technology in K-12 classrooms, which led to some very specialized SIGs, such as “Secondary Maths”, and some more general ones, like “Elementary Math, Science and Technology”, but also non-science SIGs, like “Arts, Media and Design”, and two non-K-12 SIGs: “Informal learning and museums”, and “Higher education and online learning”.

Our SIGs were different from most MOOC cohorts in two ways. First, they were semantically meaningful, i.e. designed based on actual data on student interests and professional contexts, and then actively chosen by the students. Second, in traditional MOOCs, cohorts are just applied to forum participation, but in this MOOC the integration between the EdX forum cohorts and our external activities, such as group peer review, and lesson design collaborative workbench, provided students with a rich variety of forms of engagement, knowledge exchange, and collaboration.

1.3 Previous Research

In an attempt to understand how this nested social structure, and the semantically meaningful SIGs, contributed to structuring student interaction and discourse, and impacted the quality of the final artefacts (lesson design documents), we have previously analyzed individual student activity traces (video watching and forum access behavior), collaborative actions within design groups, and the social network structures within SIGs [5]. We have presented a coding scheme for the quality of the lesson design documents across five dimensions (Table 1), and correlated these quality metrics with individual and SIG characteristics.

We have found evidence of different SIGs “making the MOOC their own”, with significant differences in video-watching behavior between K-12 SIGs, for whom the MOOC was originally planned, watching more of the K-12 focused videos, and higher education SIGs focusing more on theoretical and conceptual videos. We found strong correlations between SIG reviews and design document quality, but only for the early formative weeks. High network centrality in the SIG discussion forum social networks was also correlated with higher design document quality.

1.4 Research Questions

To continue our analysis of how the design of our course impacted student learning and behavior, and explore how we can conduct an analysis of learning data

that corresponds with the learning design, this study will examine the semantic flow of ideas and concepts between the different social levels, with SIGs as our primary unit of analysis. All students began each week with the same set of new ideas delivered through the videos. We will examine how these common ideas became applied to each disciplinary area in the different SIGs, and how this influenced the knowledge work in the lesson design groups, through individual student uptake from SIG-specific forum discussions and the reviews they received from other SIG group members.

A key question will be whether the sub-community in a SIG adds something beyond what could be expected based on a simple correlation between individual student disciplinary interests, and that student's contributions. Our goal is to understand how the nested social structure, and the sub-communities students formed in SIGs, influenced students' discourse and work in Lesson Design groups. We will also look at the difference between SIGs in terms of disciplinary focus and cohesion, and whether this contributed to the quality of the design documents.

2 Literature Review

2.1 Knowledge Community and Inquiry

Knowledge Community and Inquiry (KCI) is a pragmatic framework for curriculum development to foster knowledge communities, which advocates scripting and coordinated grouping to assure comprehensive distribution across a targeted domain, but adds a layer of collective knowledge building, where students engage with Web 2.0 technologies to develop a shared knowledge base that serves as a resource for their subsequent inquiry [16].

KCI projects are designed explicitly to include inquiry activities that lead to the production of artifacts that allow for assessment of learning on a set of pre-specified goals or expectations. Typically, artifacts are evaluated for coherence (presence of mutually conflicting ideas), and completeness [14]. Many KCI designs feature a group project in which students collaborate throughout the term, with new elements or dimensions added as the students gain access to a larger individual and community knowledge base, and become more conceptually sophisticated [13]. Recent examples include students creating a wiki about human disease and body systems, researching Canada's biodiversity [15], or drafting proposals on how to remedy climate change issues [19].

2.2 Grouping and Cohorts in MOOCs

Researchers have looked at forming small groups in MOOCs based on criteria like study habits, time zones, language, learning goals, and collaboration method [20], often aiming to match these characteristics, but in other cases aiming to create culturally heterogeneous groups [10]. Apart from intrinsic student attributes, researchers have also used data about previous student interactions in a course to form more effective groups [18].

Some unique aspects of our study are the nested social structure, with Lesson Design groups that operate within the social context of a Special Interest Group (co-hort), and also that the students could be stratified very naturally based on teaching interest and age group targeted. Because of the number of collaborative elements that we custom-designed and integrated into the course, the social stratification was also much more wide-reaching than in typical studies, where they have often focused on forum discussions or short video meetings.

2.3 Analysis of Text in xMOOC/cMOOC

Forums have been a key focus both in xMOOCs and cMOOCs. In the context of xMOOCs, most of the researchers have used social network analysis (SNA) based variables [4, 8], forum usage statistics [11, 1], and timing patterns [9] to predict grades of the MOOC learners. These methods often use clustering/classification algorithms to cluster/predict the learners' grades. One drawback of such methods is that these methods are used as "black boxes".

On the other hand, in the context of cMOOCs, the main focus is on how learners define their own roles [3], sentiments in the forum posts [2], topic analysis [7], and interaction patterns in the forums [17] to predict/explain the engagement within the MOOC. The primary drawback of these efforts is lack of a universal definition of engagement/dropout, which makes the findings difficult to generalize. In this paper, we present a simple text analysis from a collaborative MOOC to show the relation between the information flow at different social granularities to assess the quality of the artifact produced by each team.

3 Methods and Variables

Coding scheme: Each Lesson Design group was required to produce a design document with the details of a (possibly multi-hour) lesson that would be taught in their classes. Two authors coded these documents, with an inter-rater reliability of 0.82, according to the coding scheme in Table 1.

Table 1. Coding scheme for design document quality

Code	Description
Learning Objectives (LO)	Level of detail put in the learning objectives mentioned
Activity Design (AD)	Richness in the design of the activities according to the learning objectives
Coherence (CO)	Level of coherence in the various parts of the design document
Innovative use of technology (DT)	Depth of thought put into the innovative use of technology in the design document.
Incorporating inquiry-based learning (IB)	The use of inquiry-based learning principles in the design document

Tf-idf: For each SIG, we computed the three Term Frequency-Inverse Document Frequencies (Tf-idf), one each for the forum, reviews, and Etherpads. We computed these three tf-idf matrices for every week. The tf-idf value for each term in the matrix denotes two things simultaneously: 1) how important a term is for one document, and 2) how important the term is across the complete set of documents.

Similarity: In order to compare the different tf-idf matrices, we computed the cosine similarity between two matrices. The cosine similarity will inform us about the conceptual similarities between the two SIGs, or for the same SIG across forum, reviews, or Etherpads. The similarity value is bounded within the interval (0,1), both values included. A similarity value of zero would depict orthogonal concept spaces, that is, there would be no common themes across those two sets of concepts. On the other hand, a similarity value of one would indicate complete similarity, that is, the two sets of concepts would be the same.

Betweenness and withinness: We computed two types of similarities. The first similarity betweenness is computed between the forums and Etherpad from two different SIGs for every week. The second similarity withinness is computed among the forums, reviews, and Etherpad from the same SIG for every week.

Uptake of ideas: In the present MOOC, the flow of ideas among the participants went in three directions: 1) review to Etherpad; 2) forum to Etherpad; 3) reviews to forum. Every design group received peer feedback on their current state of the design document. This feedback was continuously provided during the course and the peer reviewers were given specific weekly theme-related prompts to suggest improvements to the design documents.

To evaluate the uptake of ideas from the reviews by design groups in the different SIGs we computed the similarity between the reviews they received and the Etherpad (internal group discussion) for the subsequent week. Besides reviews, the design groups also received ideas from the discussions in the forums. We also computed the uptake of ideas from forum using the similarity between forum and Etherpads from the same week.

Finally, to measure the effect of the reviews on the forum discussions, we computed the similarity between the reviews and forums from consecutive weeks.

Uptake from videos: The videos represent a common source of ideas for all SIGs. We computed the similarity between the video transcripts of every week with the forums of every SIG, to evaluate the effect of the information provided by the instructors on the discussions in the different SIGs.

Case studies: The SIGs were designed based on participant interests, however participants chose freely which SIG to join. Since participants differed across multiple dimensions (discipline, age group, etc.), some participants with similar disciplinary focus might have joined different SIGs.

To gauge the effect of the SIG discussion (review, forum, Etherpad) on the individual participants, we extracted all the participants from four specific disciplines (Math, Physics, Chemistry, Biology) from their respective SIGs. We then computed the withinness with their own SIGs and the betweenness with the rest of the SIGs. A higher value of betweenness than withinness will denote that

the discussion contributions from individuals are affected by their disciplines; contrary to this, a higher value of withinness will show that the SIG community has a higher effect on the discussions.

4 Results and Discussion

Table 4 shows the average betweenness for respectively forums, reviews and Etherpads across all SIGs during the same week, as well as the pair-wise withinness for forums, reviews, and Etherpads.

Table 2. Average betweenness/withinness for forums, reviews and Etherpads, all SIGs

	Withinness with Reviews	Withinness with Etherpads
Forum (betweenness: $M = 0.50$; $sd = 0.12$)	$M = 0.80$; $sd = 0.03$	$M = 0.60$; $sd = 0.13$
Reviews (betweenness: $M = 0.48$; $sd = 0.12$)	-	$M = 0.67$; $sd = 0.14$
Etherpads (betweenness: $M = 0.22$; $sd = 0.07$)	-	-

Betweenness for two SIGs is highly explained by their respective similarities with the video content (mean adjusted R-sq 0.76). Figure 4 shows the average weekly similarity between the MOOC-wide videos, and a given SIG (forum, reviews, and Etherpads). From Figure 4, we can see that most of the SIGs maintain a consistent similarity profile except SIGs 3 (in weeks 2 and 5) and 4 (in week 5). The reason that these three similarity values are zero, is the absence of any activity from the SIGs 3 and 5 during the respective weeks.

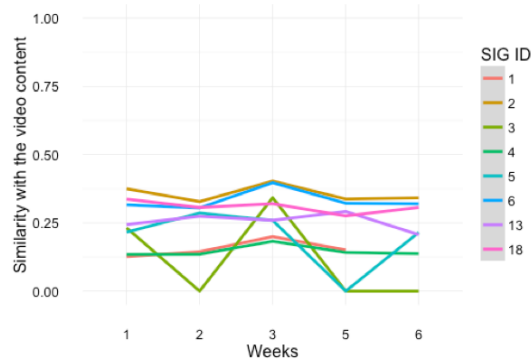


Fig. 4. Similarity between SIG content and videos per week.

4.1 Does SIG Membership Shape the Discourse of Individual Teachers?

We found four disciplines with a large number of teachers, dispersed across multiple SIGs, and tested whether the similarity between teachers from the same discipline (for example physics teachers) were greater than the similarity between a given teacher and his or her SIG (for example Secondary Science). We found that in all four cases, teachers' contributions were significantly more similar to their SIGs, than to teachers with the same disciplinary interests who had joined other SIGs (see Table 3). This shows that the discourse that developed within SIGs informed individual behavior more than what could be explained by looking at individual interests and demographics. Figure 5 shows the development of these relationships for each week of the course.

Table 3. Comparing betweenness and withinness for four types of teachers

Similarity \sim Case*Type	df1	df2	F (df1, df2)	p-value
Case (physics, chemistry, biology, maths)	1	62	0.50	0.47
Type (between, within): within discipline <within SIG	1	62	6.14	0.01
Case:Type	1	62	0.16	0.68

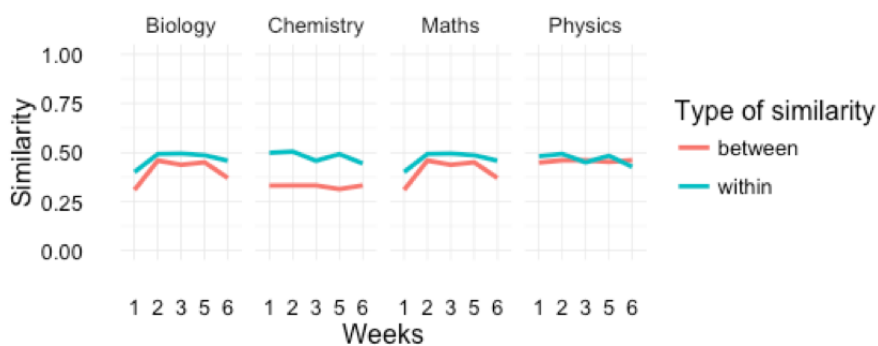


Fig. 5. Comparing betweenness and withinness for four types of teachers.

4.2 Measuring Semantic Diversity of SIGs

We used tf-idf to find the most representative concepts for each SIG (words commonly used in one SIG, and very rarely used in other SIGs). There was a difference between disciplinary-focused SIGs, such as the four listed below, and SIGs focused on a specific age group or audience (museums and informal

learning, higher education). The latter SIGs had very few words that were over-represented, suggesting a larger diversity of internal ideas and directions.

Table 4 shows the most representative disciplinary concepts for four discipline-focused SIGs. The number of common terms across SIGs for each week decreases as the course progresses, suggesting that the SIGs become more unified and perhaps more focused on specific applications, and less on the general concepts that unify the course.

Table 4. Representative terms from four different SIGs across all weeks

Secondary Math	Secondary Science	Arts, Media and Design	Secondary English
percentage	enzyme	atlas	kinesthetic
proportion	hierarchical	morphology	invasion
autograph (math software)	motion	art piece	essence
geometry	PBL (problem-based learning)	pastel	frighten
circle	substrate	watercolor	marginalization
symmetrical	cellular	Pollack	dramatic
lag	PhET lab	Storybird	Hannibal
representation	respiratory	Van Gogh	individual
GCF	ecosystem	melody	Rome
LCM	protein	advocacy	captivate

4.3 Uptake

An important part of the course were the reviews in four weeks of the course, which were disseminated by the participants to their peers scaffolded through the weekly review prompts for each week. In every subsequent week, we found that much of the commonality between the previous weeks' review and SIG discussions and Design groups' Etherpad comments could explain the overall quality of the final design documents. In Table 5 (last two columns), we show the percent of the variance explained of the design document quality ratings by the similarity between reviews a given week and next week's forums and Etherpads respectively. We observe that uptake of reviews in forums is a better predictor of the design document quality than uptake of reviews in Etherpads. One possible explanation could be that the amount of common knowledge in the forum is much higher than that in the Etherpads; as the whole SIG contributes to the forums, while the Etherpads are specific to one design group.

4.4 Correlation Between SIG Characteristics and Quality of the Design Document

We investigated the correlations between different design document quality metrics (as listed in Table 1), and semantic cohesion. In Table 5, we show the pairwise

similarity between Etherpad, reviews, and forums, as well as the individual between similarities between respectively all Etherpads, reviews and forums across SIGs for a given week. We also show the similarity between the forum of a given SIG and the videos of that week (which would indicate idea uptake and focusing on the weekly theme).

The values in Table 5 are the adjusted R-squared of the linear model between the two variables. The dependent variables are the quality ratings and the independent variables are the various similarities. Due to a low number of teams having all the similarity values, we decided to keep the linear models limited to one dependent and one independent variable, thus getting an early estimate of the feature importance for conducting predictions in future. One might argue that we could have used some feature selection mechanisms for reducing the dimensionality of the feature space. Once again, the number of teams ($n=8$) is not enough to carry out ridge regression. Moreover, it is not less than the number of measures ($p=9$) so that one could carry out dimensionality reduction suitable for $n < p$ situations.

We found that Learning Objectives, Design Thinking and Incorporating inquiry-based learning are all very much explained by the video similarity (theme uptake) and the SIG within similarity (cohesion). Activity Design is loosely related to within similarity, and for Cohesion, there is no relationship. Activity Design could be seen as more of a measure of individual creativity, and Cohesion is a meta-level indicator.

Table 5. Adjusted R-squared for the five design document quality ratings using the different similarity scores.

	Within similarity			Average between similarity			Similarity of forum w/video	Uptake - Review and the next weeks Etherpads and forums	
	ER	EF	RF	simE	simF	simR	simV	RE	RF
LO	34.1	32	26.3	3.6	4.6	2.7	24.9	7.9	30.9
AD	19.7	15.9	16.1	2.4	4.4	3.9	2.2	4.7	36.5
CO	2.4	4.2	4.6	1.8	3.2	0.1	6.8	2.9	32.6
DT	28.1	23.6	29.8	1.7	19.8	11.6	52.3	2.1	55
IB	23.5	17.8	18.7	4.2	15.1	9.8	59.6	4.7	42.3
Mean quality	21.3	18.9	19.7	1.3	9.8	4.9	32.7	1.9	44

5 Conclusions and Future Work

In this paper, we have presented an analysis of the impact of semantically meaningful cohorts in a unique MOOC. We showed how the ideas discussed in different SIG communities (forums, reviews and Etherpads) were seeded by the weekly videos, which indexed the discussions to the course themes, and were informed

by the disciplinary focus of the SIG participants, but were then developed into a coherent discussion that represented something beyond simply a statistical sum of the participants. This can be seen through our analysis of participants with similar disciplinary foci that ended up in different SIGs, and how their expressions of ideas gradually become more similar to the SIG discourse which they are part of, than to the other participants with similar foci in other SIGs.

SIGs and cohorts are an attempt at managing or reducing scale, to avoid overwhelming students, and the large number of students enabled us to form specialized topic-based SIGs in a way that would not have been possible in a small class. However, due to the unequal distribution of interests among students, in what was primarily marketed as a course for STEM K-12 teachers, some of the SIGs were quite specialized around certain disciplines, and others had to group together a number of related disciplines to get a large enough critical mass to support discussions and knowledge work. We have shown that the more specialized SIGs have a higher level of withinness, and are also correlated with a higher quality of the final design documents, perhaps because the forum discussions and reviews were more relevant to the design group efforts.

Student interests have several dimensions, and grouping students in cohorts necessarily prioritizes a subset. We could imagine a physics teacher, working in a high-school, and interested in 3D printing. While she might be grouped with other physics teachers, she would lose out on the comments by the 3D-printing enthusiast in the chemistry SIG. While our analysis has shown the added value of having stable communities whose discourse develops in a coherent manner, future studies could investigate the use of semantic tags (on both participant profiles and content), or text analysis, to promote idea exchange across SIGs.

References

1. Cobo, G., García-Solórzano, D., Morán, J.A., Santamaría, E., Monzo, C., Melenchón, J.: Using agglomerative hierarchical clustering to model learner participation profiles in online discussion forums. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. pp. 248–251. ACM (2012)
2. Dmshinskaiia, N.: Dropout prediction in MOOCs: using sentiment analysis of users' comments to predict engagement. Master's thesis, University of Twente (2016)
3. Dubosson, M., Emad, S.: The forum community, the connectivist element of an xmooc. *Universal Journal of Educational Research* **3**(10), 680–690 (2015)
4. Fancsali, S.: Variable construction and causal modeling of online education messaging data: initial results. In: Educational Data Mining 2011. Citeseer (2010)
5. Håklev, S., Sharma, K., Slotta, J.D., Dillenbourg, P.: Contextualizing the co-creation of artefacts within the nested social structure of a collaborative mooc. In: Lavoué, É., Drachsler, H., Verbert, K., Broisin, J., Pérez-Sanagustín, M. (eds.) *Data Driven Approaches in Digital Education 12th European Conference on Technology Enhanced Learning, EC-TEL 2017, Tallinn, Estonia, September 12–15, 2017* (2017)
6. Håklev, S., Slotta, J.D.: A principled approach to the design of collaborative MOOC curricula. In: Kloos, C.D., Jermann, P., Pérez-Sanagustín, M., Seaton,

- D., White, S. (eds.) *Digital Education: Out to the World and Back to the Campus*. EMOOCs 2017. Lecture Notes in Computer Science, vol. 10254, pp. 58–67. Springer, Cham (2017)
7. Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., Hatala, M.: What do cmoc participants talk about in social media?: a topic analysis of discourse in a cmoc. In: *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*. pp. 156–165. LAK '15, ACM, ACM, New York, NY, USA (2015). <https://doi.org/10.1145/2723576.2723609>, <http://doi.acm.org/10.1145/2723576.2723609>
 8. Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V., De Kereki, I.F.: Translating network position into performance: importance of centrality in different network configurations. In: *Proceedings of the sixth international conference on learning analytics & knowledge*. pp. 314–323. ACM (2016)
 9. Khan, T.M., Clear, F., Sajadi, S.S.: The relationship between educational performance and online access routines: analysis of students' access to an online discussion forum. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. pp. 226–229. ACM (2012)
 10. Kulkarni, C., Cambre, J., Kotturi, Y., Bernstein, M.S., Klemmer, S.: Talkabout: Making distance matter with small groups in massive classes. In: *Design Thinking Research*, pp. 67–92. Springer (2016)
 11. Lopez, M.I., Luna, J., Romero, C., Ventura, S.: Classification via clustering for predicting final marks based on student participation in forums. *International Educational Data Mining Society* (2012)
 12. Mor, Y., Ferguson, R., Wasson, B.: Learning design, teacher inquiry into student learning and learning analytics: A call for action. *British Journal of Educational Technology* **46**(2), 221–229 (2015)
 13. Najafi, H.: *Transforming Learning in Science Classrooms: A Blended Knowledge Community Approach*. Ph.D. thesis, University of Toronto (2012)
 14. Peters, V.L., Slotta, J.D.: Analyzing collaborative knowledge construction in secondary school biology. In: *Proceedings of the 9th International Conference of the Learning Sciences-Volume 1*. pp. 548–555. International Society of the Learning Sciences (2010)
 15. Peters, V.L., Slotta, J.D.: Scaffolding knowledge communities in the classroom: New opportunities in the web 2.0 era. *Designs for Learning Environments of the Future* pp. 205–232 (2010)
 16. Slotta, J.: *Knowledge Community and Inquiry*. Paper presented and published for the Network of Associated Programs in the Learning Sciences (NAPLES). Tech. rep. (2014)
 17. Wang, Z., Anderson, T., Chen, L., Barbera, E.: Interaction pattern analysis in cmocs based on the connectivist interaction and engagement framework. *British Journal of Educational Technology* **48**(2), 683–699 (2017)
 18. Wen, M., Maki, K., Wang, X., Dow, S., Herbsleb, J.D., Rosé, C.P.: Transactivity as a predictor of future collaborative knowledge integration in team-based learning in online courses. In: *EDM*. pp. 533–538 (2016)
 19. Zhao, N., Najafi, H., Slotta, J.D.: An analysis of teacher-students interactions in three science classes: A pilot study. In: *Proceedings of the Ninth International Computer-Supported Collaborative Learning Conference*, Hong Kong. International Society of the Learning Sciences. (2011)
 20. Zheng, Z., Vogelsang, T., 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)