Resource Usage in Online Courses: Analyzing Learner's Active and Passive Participation Patterns

Sabrina Ziebarth, University of Duisburg-Essen, ziebarth@collide.info
German Neubaum, University of Duisburg-Essen, german.neubaum@uni-due.de
Elias Kyewski, University of Duisburg-Essen, elias.kyewski@uni-due.de
Nicole Krämer, University of Duisburg-Essen, nicole.kraemer@uni-due.de
H. Ulrich Hoppe, University of Duisburg-Essen, hoppe@collide.info
Tobias Hecking, University of Duisburg-Essen, hecking@collide.info
Sabrina Eimler, University of Duisburg-Essen, sabrina.eimler@uni-due.de

Abstract: The paper analyzes the experience with an open university course for a very heterogeneous target group in which MOOC-like materials and activities were used. The course was conducted in a specifically prepared and extended Moodle environment. The analysis involves questionnaires as well as performance data that reflect the resource access on the learning platform. A special focus is put on the participants' acceptance and usage of student-generated versus teacher-provided learning content. Network analysis techniques have been used to identify "interest clusters" of students around certain resources.

Keywords: large online courses, learning analytics, student-generated content

Introduction

Massive Open Online Courses (MOOCs) have become popular especially in the US with powerful providers and platforms hosting a large variety of courses (e.g., EdX, Coursera, Udacity). A famous example of European origin is M. Odersky's SCALA Programming MOOC originally offered by EPFL Lausanne and now also available on Coursera. Since many MOOCs are offered by prestigious institutions, yet open for participation without specific prerequisites and fees they accumulate huge world-wide participation. MOOCs suffer from high drop-out rates, and it is still difficult to turn successful participation in a MOCC into "convertible currency" in terms of accepted credits. Still, MOOCs are a promising innovation of in university education since they explore the potential of massification as a resource for learning and are striving for making online learning more interactive. Because of their independence of location and time (Kay, Reimann, Diebold, & Kummerfeld, 2013), MOOCs are also attractive for heterogeneous groups from an inclusion perspective, e.g. for participants with special needs due to non-standard perceptual or language proficiencies and they might offer a better compatibility of family and studies and/or work with benefits for both students and lecturers.

While MOOCs provide an immense potential allowing people a more independent way of learning, instructors and designers of these environments are confronted with a number of challenges regarding the diverse background of students. Learning materials have to be thoroughly conceptualized in order to meet individual needs of a large group of students with different backgrounds. Moreover, since learning can be understood as an inherently social process (e.g. Stahl, 2000), another challenge of these comparably anonymous formats might be to encourage participants to login week after week, and find tasks allowing and motivating collaboration with other participants.

Goals and research questions

In this paper, we concentrate on examining patterns and preferences regarding the usage of learning materials typical for MOOCs in an online university course with a heterogeneous target group. Current research shows dependencies between the usage of learning management systems and archived learning outcomes (e.g. Mödritscher, Andergassen, & Neumann, 2013). In Germany, many study programmes do not only comprise subject-specific courses, but also demand several courses from other subject areas ("optional studies") with the idea to broaden the students' horizons and to foster cross-disciplinary communication. The selected online course (see next section) was offered as "optional study" course targeting students from various study programmes of two large German universities. Accordingly, our target group was particularly heterogeneous regarding prior knowledge and interests.

Using log data from the platform in which the online course was conducted as well as information gained from several questionnaires that were administered during the course we use different methods to analyze the following research questions that we consider important for online courses in general:

- Are there typical usage patterns or preferences regarding learner-generated vs. instructor-generated content?
- Can success predictors be inferred from resource usage?
- Are there characteristic observable differences between students with a prevalent extrinsic motivation (compliance to get credits) versus intrinsically motivated students?

Context and background

Between October 2013 and January 2014 the online course was accessible for eleven weeks and especially advertised for students at two large German universities who could gain credits for participating actively in the course and passing the final exam. Altogether 162 students enrolled in the course.

The course dealt with psychological foundations of computer-mediated communication (cmc) with a special focus on learning and teaching, covering classical theories of cmc to understand the changes that (might) occur by this mediation as compared to face-to-face settings. In doing so, students were provided with hands-on experiences, e.g. in terms of virtual collaborations in small groups of students working on one specific assignment.

This course was open to a large audience of students in two big universities (>30.000 students each) and completely delivered online, but it was not advertised to the outside. Also, the examination was handled inside the two universities in such a way as to obtain "real" credits. Accordingly, this course was not a typical MOOC in terms of targeting a huge, heterogeneous group of learners with different academic and non-academic backgrounds and age levels. Still, it addressed students from very diverse fields such as business administration, education, or media studies. Apart from the target group, the course integrated several elements and activities typical of MOOCs such as instructional videos, discussion forums, weekly assignments and quizzes.

Upon registration to the Moodle-based course environment, participants were informed about organizational conditions and requirements for successful completion. The choice of Moodle (www.moodle.org) as a learning management system was based on the fact that it was already known to most of the potential participants. Also, we could benefit from a number of dedicated extensions (e.g., a video plug-in) that had been developed in a student software development project on beforehand. Course participants were told to access the course environment regularly and not to be absent for more than two subsequent weeks (otherwise they would be automatically excluded from the course). Moreover, they were informed that they would regularly get the chance to complete questionnaires to assess their course experience in order to improve the course. The weekly learning material provided consisted of a video (6 to 9 minutes in length), in which one of the course organizers contextualized the content and introduced the new material and tasks that could be accessed and downloaded. Besides the videos, one or two texts a week, preferably in German language, were provided as core material, as well as some additional material (e.g. video links and text documents) that could optionally be used as "outside the box" material. Also, the completion of quizzes and individual assignments were part of the required activities. A special format we wanted to test, that differs from most conventional MOOCs (e.g. in the sense of time independence), are collaboration tasks. People had to complete two out of three of these tasks in which they worked in a group of up to four students for one week. In the last week of the course, a wiki for exam preparation was provided, in which the participants could bring together their information on the topics addressed in the course. For the optional studies students, the written examination directly followed after the end of the lecture period.

Overall, our course setting is comparable to MOOC conditions with respect to the voluntary (optional) participation, the heterogeneity of the community (no common prerequisites), and the different materials and learning activities offered. However, the course was not really open to the outside and not particularly "massive".

Learning resource-centered learning analytics

In addition to statistical techniques, our analysis and evaluation of the course activities makes use of computational analysis techniques. Concerning the nature of the input data, such learning analytics methods can be categorized as content analysis, sequence analysis and structural analysis. This categorization naturally applies to the structuring of existing methods for learning resource related analysis in online courses. Content analysis is mostly concerned with the identification and visualization of patterns in possibly collaboratively learner generated artefacts (Southavilay, Yacef, Reimann, & Calvo, 2013) as well as discourse analysis of online discussions (De Wever, Schellens, Valcke, & Van Keer, 2006). While the temporal order in which resources are used by students is of particular importance for sequential methods in order to identify sequential patterns of resource access (Perera, Kay, Koprinska, Yacef, & Zaiane, 2009) or frequent learning paths (Bannert, Reimann,

& Sonnenberg, 2014), structural analysis methods can be used to investigate the relations between students and learning resources in general. This also incorporates descriptive statistics. In the work of Nachmias and Segev (2003) it was shown that nearly all learning resources provided in online learning environments are used by at least one student while the concrete collections of resources used by individuals can differ significantly. In this sense it has been argued that different learner types vary in their preferences for types of learning material provided in online courses (Grünewald, Meinel, Totschnig, & Willems, 2013).

In many cases only the relations between students and learning resources are directly observable from log protocols of the learning resource usage of students. However, relations between resources or students can be inferred indirectly. Association rule mining techniques can be used to discover relations between learning resources that are frequently used in combination (Merceron & Yacef, 2008; Romero, Ventura, & García, 2008). This can be of added value for tutors in order to improve course design.

In the work described in Hecking, Ziebarth, and Hoppe (2014), relations between students and learning resources were represented as bipartite networks in which students are connected to learning resources they used in a certain time period. By applying network analysis methods densely connected clusters of students and learning resources can be identified in those networks. Such clusters comprise a group of students who have a common interest in the corresponding group of learning resources manifested by many connections between the two. By investigating the evolution of the student - resource clusters over time it is possible to discover characteristic resource access patterns in resource intensive online courses. This approach is utilized in this work and will be described in more detail in the Methodology section. The approach of modeling interests of individuals in semantic objects as networks was also used by Harrer, Malzahn, Zeini, and Hoppe (2007) in order to facilitate community building, reflection, as well as trend detection.

In this study we are primarily interested in relationships between students and different types of learning resources. Thus, the applied methods mainly concern structural analyses prior to content and sequential analyses in order to gain generic insights into the role of user generated and tutor provided learning material.

Methodology

Data

Table 1 shows the resource types that were used in the course as well as their goals and tasks. All interactions with the course platform were logged.

Table 1: Available resource types and related goals / tasks

Available Resource Type	Goals / Tasks
Video	• Providing an overview of the week's topic • Teaching selected concepts • Giving details to the group tasks
Documents (core and optional)	• Conveying the learning content • Basis for group tasks
Quizzes	• Self-testing
Forums	• Exchange for group tasks • Getting support (questions regarding content and organization) • Collecting possible exam questions • Optional discussion for deepening the knowledge on specific topics
Wiki	• Summarizing contents for exam preparation

In order to analyze the role of student-generated content compared to tutor-provided learning resources from different aspects a triangulation approach comprising different analysis methods was used. In the following, each of the applied methods is introduced.

Resource usage

To get an overview on the general resource usage, the log files accumulated on the Moodle platform were analyzed using descriptive analysis. Furthermore, the Apriori algorithm (Agrawal et al., 1996) has been employed to find rules regarding frequent combinations ("co-occurrences") of learning materials.

Satisfaction with resources and academic motivation

During the course the participants were asked to complete several questionnaires. These contained questions regarding their login behavior as well as their usage of the provided learning material. Additionally, the satisfaction with the learning material was assessed adapting items from Reinhardt (2008) as well as using items from the Training Evaluations Instrument (TEI) by Ritzmann, Hagemann, and Kluge (2014).

Moreover, we were interested in the motivation pattern of participants of this course. We therefore relied upon the concept of the academic self-regulation (Ryan & Connell, 1989). Based on this concept, academic motivation is regarded on a continuum of self-determination (Deci & Ryan, 2002).

Success predictors

For identification of success predictors linear regression was used. Here, several aspects of student behavior (number of course log-ins three weeks before the exam, views of wiki articles, editing of wiki articles) and their performance during the course (measured via performance in the quizzes) were used as predictors for the final grade in the course as assessed in the final exam.

Student profiles based on resource access patterns

For the modeling of student profiles, the approach of Hecking et al. (2014) was used: Based on the log files bipartite student-learning resource networks were created for each week of the course. Applying the bicliques community clustering algorithm (Lehmann, Schwartz, & Hansen, 2008) partially overlapping clusters of students and resources (as depicted in Figure 1) are identified for each time slice. The students in such a cluster form an interest group because they have affiliations to the same set of resources while not having necessarily social relations. If the similarity of student groups of two clusters in two consecutive lecture weeks exceeds a certain threshold, the group is considered to be the same with two occurrences in the corresponding two weeks. These groups do not necessarily form a bipartite cluster with the same group of resources, which indicates a collective shift of interest. Thus, while clusters are identified for each lecture week an interest group of students can persist over several weeks as part of different clusters. Students who are often part of large clusters with a relatively stable group of students during the course are considered to follow a "mainstream" regarding their resource access behavior that is typically influenced by the course design, while there are also students showing more diverse resource access patterns. We define the mainstreaming coefficient (*msc*) for a student i who was part of k_g interest groups in k_c student - resource clusters during the lecture period as: $msc(i) = \exp(\frac{-k_g}{k_c})$. Students who often appear in clusters, but mostly as part of the same student group, receive high values (close to 1) for *msc*. For inactive students who were never part of a cluster *msc* is set to 0.

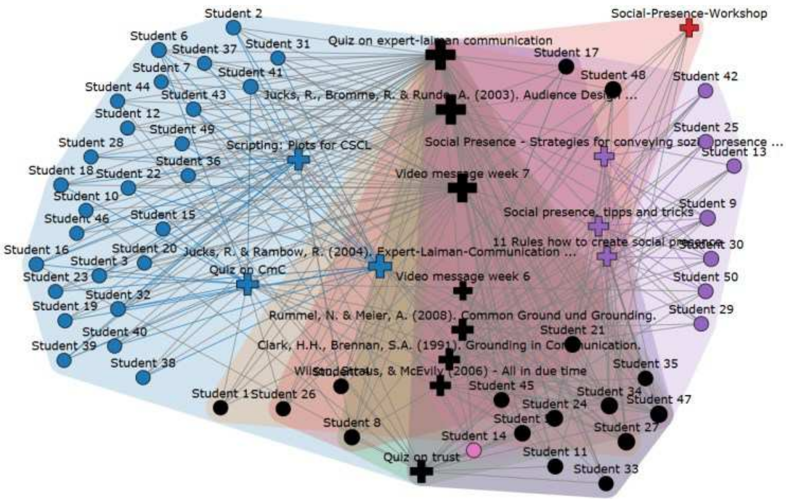

Figure 1 Example of a clustering of a student - resource network. Black nodes belong to more than one cluster.

Analysis results

Aggregated resource usage

In the following only the log files from the 69 students who took the exam at the end of the course were considered. Figure 2 shows the logins and Figure 3 the usage of different types of resources per week. The activity in week 3 and week 8, in which the first two group tasks took place, was considerably higher than in the other weeks. In week 13, in which the third and last group tasks was conducted, the amount of activity on the platform was much lower than in the first weeks with group tasks. The most activity in the weeks with group tasks was in the group forums that were provided for group exchange.

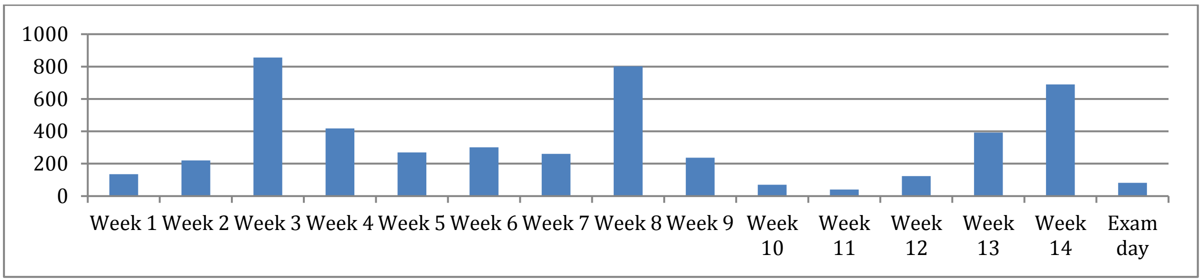

Figure 2. Logins per week

After the Christmas break (week 10 and 11), in which the course was paused, the activity increased until the day of the exam. The wiki for exam preparation that was provided in the last week of the course was highly used. 85.5 % of the students taking the exam actively engaged by editing one, two, four of the 11 articles and 79.7 % by writing comments to one to three articles. Every student read at least one article, the average of articles that were read at least once was 9.7 and the median 11 (all); on average each students accessed 47 times a wiki article. The analysis of association rules with the Apriori algorithm based on Moodle sessions shows that wiki articles were mainly opened in the order in which they were arranged on the index page and which corresponds to the order in which the topics occurred in the course (i.e. the articles to the topics of week 1 and 2, week 2 and 3, etc. were often opened together in on session).

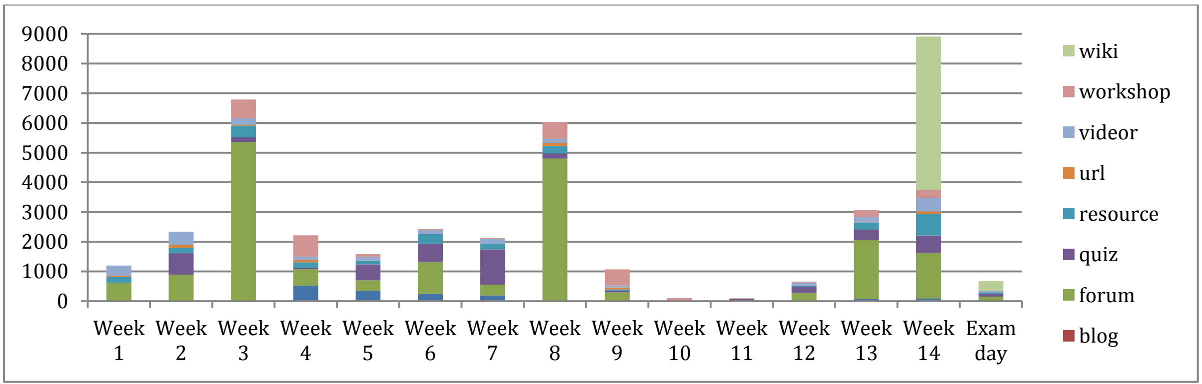

Figure 3. Usage of different types of resources per week

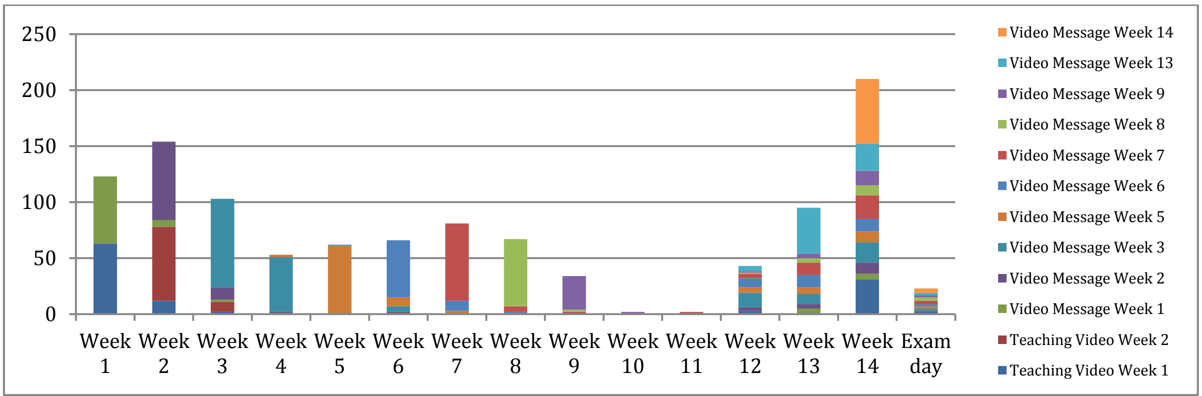

Figure 4. Video usage per week

There are indications that exam preparation takes place in the last three weeks before the exam: Although the last quiz was provided in week 7 the usage of the “old” quizzes increased starting in week 12 (see Figure 3). Furthermore, while in the previous weeks the video usage mainly corresponded with the new video message provided in that week, the video usage became much more diverse (see Figure 4). Similarly, documents provided earlier in the course are accessed.

Success predictors

To determine the influence of the achieved quiz results during the course (predictor) on the final grade of the course (criterion) a linear regression was used. Results revealed a statistically significant relationship between these two variables, R^2= .31, F(1,67)= 29.57, p< .001. A negative relationship was found between the achieved quiz results and the grade of the course (β= -.55) indicating that higher quiz grades during the course predict a better final course grade (as lower grades indicate better performance).

Also, the influence of the course views during exam preparation (predictor) on the course grade (criterion) was of interest. Results of a linear regression showed no significant relationship between these two variables, R^2= .04, F(1,67)= 2.96, p= .09.

In order to test whether the passive usage of learner-generated content and the active contribution to learner-generated content can predict the final grade, we calculated two additional regression analyses. First, we used a linear regression to analyze whether the views of learner generated wiki articles predicted the grade of the course. The results showed that the views of wiki articles are a significant predictor of the course grade, R^2= .16, F(1,67)= 12.43, p< .001 (β = -.40), indicating that the more views of wiki articles are observed the better is the final grade. Moreover, it was of interest if active participation, in the sense of edits on wiki articles, predicted the grade of the course. Results of a linear regression showed that edits on wiki articles are significant predictors of the course grade, R^2= .08, F(1,67)= 6.03, p= .02 (β = -.29). Here, also, the more participants contribute actively, the better is their final grade. However, the explanation of variance is considerably lower compared to the passive usage of wiki articles (8% versus 16%).

Satisfaction with resources

Self-reported usage

One set of questions referred to the usage of the learning material provided, i.e. texts, videos, etc.. The largest group (39.7%) reported to have used the majority of the material that was available, 21.9% indicated to have used half of the material, 9.6% claimed to have used all available material and 6.9% reported to have used only minor parts of the material.

Satisfaction with learning resources

The questionnaire also asked for students overall satisfaction with the learning resources provided by the teachers (availability, accessibility, clarity), e.g. "The digital resources that were needed to complete tasks were always provided on time.", "At all times, it was clear which resources belonged to the specific week", "At all times, the purpose of resources provided was clear to me"). Altogether, six items were adapted from Reinhardt (2008) measured on 5-point Likert scales (1=fully agree, 5=do not agree at all; α =.77). Overall, results show that people were highly satisfied, M=2.05 (SD=0.73). Satisfaction with resources did not correlate with the final test grade.

Also, items from the Training Evaluations Instrument by Ritzmann et al. (2014) were used for the overall course evaluation. From the instrument that assesses subjective evaluation of the course as well as aspects of the course design (1=does not apply at all, 5=fully applies) the sub-dimensions reported enjoyment, perceived difficulty, knowledge and demonstration were used. The demonstration dimension was based on asking if the learning goals had been clarified and been reached and if the (available) media used were considered helpful and suitable. Internal consistency shows to be good for all sub-dimensions (see Table 2). All means range around the scale mean and show significantly positive correlations with the final test grade (see r-values in Table 2).

Table 2: Means, standard deviation, internal consistency and correlation with final grade for the dimensions enjoyment, difficulty, knowledge and demonstration

Dimension	M	SD	α	p	r
Enjoyment	3.12	1.01	.90	.04	.28
Difficulty	3.05	0.76	.74	.03	.30
Knowledge	3.10	0.97	.94	.04	.27
Demonstration	3.10	0.83	.88	.05*	.26

Student profiles based on bipartite clustering

In a further explorative analysis, we have studied contingencies between the mainstreaming coefficient based on the bipartite clustering approach and other variables. This coefficient characterizes a standard, non-deviant behavior regarding the usage of certain resources. Students with a high *msc* have been "flowing the crowd" regarding their interests in terms of resource access. For an overview, Figure 5 depicts the distribution of different interest groups each student belonged to during the course. In order to exclude effects caused by non-active students or students who drop out of the course early, the calculation was restricted to students who were at least part of 5 student - resource clusters during the lecture period. This does not automatically imply that those students were also part of 5 different interest groups. It can be seen that the majority of students belong to less than 3 or less different interest groups but a smaller number has a more diverse resource access behavior, and thus, were allocated to more different clusters.

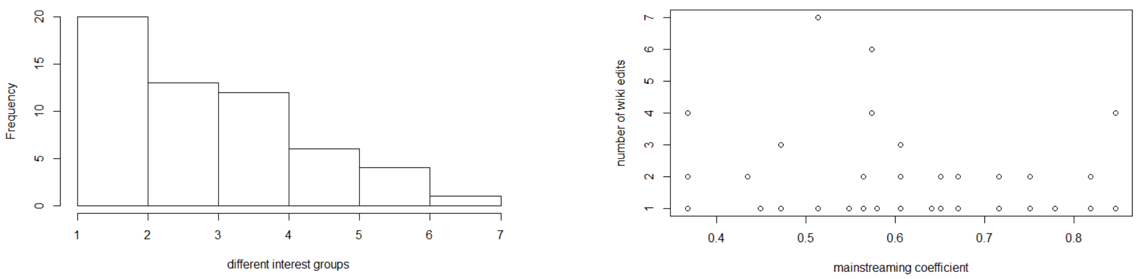

Figure 5. Distribution of the number of different interest groups the students belong to during the course (left) and mainstreaming behavior compared to wiki edits (right).

Mainstreaming often goes along with a high course activity, and accordingly success in the final exam. However, it could not be corroborated that non-mainstreamers performed worse in general. Interestingly, the mainstreaming coefficient is positively correlated with the students' extrinsic motivation ($r=0.30$, $p<.05$) as captured by the questionnaires described in the Methods section. This can be explained by the plausible assumption that extrinsically motivated students avoid experimental or explorative behavior because of unclear benefit and likely additional effort.

As the results in the previous section have shown, the exam preparation wiki was used by nearly every student in the week before the exam. However, in total only a small set comprising 44% of all students actively contributed content to the articles while 89% of the editors were students who participated in the exam. Among these content editing students there is a surprisingly high amount of individuals who cannot be classified as mainstreamers. By investigating the number of edits in the exam preparation wiki compared to the mainstreaming coefficient (right part of Figure 5) it can be seen that 50% of the contributors have a mainstreaming coefficient less than 0.61. The comparison of wiki edits and the mainstreaming coefficient of the students suggests that the resource access behavior of students during the lecture do not necessarily have an effect on wiki editing close to the exam. Moreover, student-generated content seems to be important also for some students who were less active or who used learning resources in a more individual way in the weeks before. One possible explanation could be that these students use the wiki to verify their knowledge by presenting it to the community for discussion.

Conclusion

In general, most of the students (the "mainstream") access all mandatory material and stick to the sequence of the course (they access the videos, documents and quizzes in the week they are provided). In the exam preparation phase which starts approximately three weeks before the exam also many resources of the previous weeks are accessed. The activity on the platform increases in this time until the exam takes place. Students were in general very satisfied with the teacher provided course material. While their satisfaction does not correlate with the final test grade, the reported enjoyment, perceived difficulty, knowledge and demonstration correlate significantly. But also the resource usage has some impact: The quiz grades during the course are a positive predictor for the final exam grade. Overall, our aggregated data analysis also corroborates the assumption that students make productive and successful use of peer-generated content. The different analysis results highlight the importance of the wiki for exam preparation: Although the participation was optional and no external motivation like bonus points was provided, most (ca. 86 %) of the students participating in the exam engaged actively in collaboratively creating a common knowledge base. Not only "mainstreaming" students who were very active during the course participated, but also students who were less active or who used learning resources in a more individual way in the weeks before. While active participation is a weak predictor for the grades in the exam, "passive" reading of the learner-generated wiki articles is a stronger predictor. This indicates that the collaboratively created and edited wiki articles were of high quality and well trusted by the students.

In spite of these encouraging findings, we have to bear in mind that resource access data do not tell a complete story of user behavior. The actual communication and collaboration activities might happen largely off-line and do not appear in the traces that were accessible to us. Nevertheless, the limited perspective on interaction with resources on the platform already allows for relevant insights. Since our results are based on a limited sample of students, the online course will be held and evaluated again in summer term 2015 to also get more fine-grained information on the students' actions (e.g. by click-streams).

References

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Knowledge Discovery and Data Mining, chapter 12, 307-328. AAI Press / The MIT Press.

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning, 9*(2), 161-185.

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Comput.Educ., 46*(1), 6-28.

Deci, E. L., & Ryan, R. (2002). Overview of self-determination theory: An organismic dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), Handbook of self-determination research (pp. 3-33). Rochester: University of Rochester Press.

Grünewald, F., Meinel, C., Totschnig, M., & Willems, C. (2013). Designing MOOCs for the support of multiple learning styles.*8095*, 371-382.

Harrer, A., Malzahn, N., Zeini, S., & Hoppe, H. U. (2007). Combining social network analysis with semantic relations to support the evolution of a scientific community. *Proceedings of the 8th Iternational Conference on Computer Supported Collaborative Learning,* New Brunswick, New Jersey, USA. 270-279.

Hecking, T., Ziebarth, S., & Hoppe, H. U. (2014). Analysis of dynamic resource access patterns in a blended learning course. *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge,* Indianapolis, Indiana. 173-182.

Kay, J., Reimann, P., Diebold, E., & Kummerfeld, B. (2013). MOOCs: So many learners, so much potential ... *Intelligent Systems, IEEE, 28*(3), 70-77.

Lehmann, S., Schwartz, M., & Hansen, L. K. (2008). Biclique communities. *Phys.Rev.E, 78*(1).

Mödritscher, F., Andergassen, M., & Neumann, G. (2013). Dependencies between E-learning usage patterns and learning results. *Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies,* Graz, Austria. 24:1-24:8.

Merceron, A., & Yacef, K. (2008). Interestingness measures for associations rules in educational data. *EDM, 8*, 57-66.

Nachmias, R., & Segev, L. (2003). Students' use of content in web-supported academic courses. *The Internet and Higher Education, 6*(2), 145.

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *Knowledge and Data Engineering, IEEE Transactions On, 21*(6), 759-772.

Reinhardt, J. (2008). http://www.cedis.fu-berlin.de/e-learning/beratung/werkzeuge/fragebogen/frageboegen/evaluation_blendedlearning_pdf.pdf

Ritzmann, S., Hagemann, V., & Kluge, A. (2014). The training evaluation inventory (TEI) - evaluation of training design and measurement of training outcomes for predicting training success. *Vocations and Learning, 7*(1), 41-73.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education, 51*(1), 368-384.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*(5), 749.

Southavilay, V., Yacef, K., Reimann, P., & Calvo, R. A. (2013). Analysis of collaborative writing processes using revision maps and probabilistic topic models. *LAK,* 38-47.

Stahl, G. (2000). A model of collaborative knowledge-building. *Fourth International Conference of the Learning Sciences,* Ann Arbor, MI*, 10* 70-77.