Predicting the Well-functioning of Learning Groups under Privacy Restrictions

Tobias Hecking
University of Duisburg-Essen
Duisburg, Germany
hecking@collide.info

Dorian Doberstein
University of Duisburg-Essen
Duisburg, Germany
doberstein@collide.info

H. Ulrich Hoppe
University of Duisburg-Essen
Duisburg, Germany
hoppe@collide.info

ABSTRACT

Establishing small learning groups in online courses is a possible way to foster collaborative knowledge building in an engaging and effective learning community. To enable group activities it is not enough to design collaborative tasks and to provide collaboration tools for online scenarios. Collaboration in such learning groups is prone to fail or even not to be initiated without explicit guidance. In the target situations, interventions and guiding mechanisms have to scale with a growing number of course participants. To achieve this under privacy constraints, we aim at identifying target indicators for well-functioning group work that do not rely on any kind of information about individual learners.

KEYWORDS

Group support, Predictive Models, Online courses

ACM Reference Format:
Tobias Hecking, Dorian Doberstein, and H. Ulrich Hoppe. 2019. Predicting the Well-functioning of Learning Groups * under Privacy Restrictions. In *The 9th International Learning Analytics Knowledge Conference (LAK19), March 4–8, 2019, Tempe, AZ, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3303772.3303826

1 INTRODUCTION

The rationale behind most online learning courses is to support individual learners in self-directed knowledge acquisition by providing resources and activities such as learning videos and self-test quizzes for asynchronous usage. In these scenarios, collaboration typically plays a minor role and is often limited to the interaction in discussion forums. A typical problem is a lack of individual support and thus a lack of incentives as they might arise from social interaction in a shared environment. With a high number of participants and limited teaching staff, practical solutions rely on peer-to-peer collaboration and group work [13, 15]. Ultimately, the hope is to compensate the lack of individual support through decentralized peer-help and self-organized discussions as means to improve learning quality and to reduce attrition [18].

Negative phenomena in group work such as social loafing and a lack of commitment of the members become even more problematic in online courses due to anonymity and limitations of communication facilities [11]. Beyond the provision of basic collaboration tools, continuous support and guidance of groups is needed. This requires efficient, well-calibrated and scalable intervention mechanisms, which is in-line with research on automating the scaffolding of group work. This includes predictive models that allow for an early detection of problematic patterns in small group collaboration. In this paper, we elaborate on the prediction of different quality measures for the well-functioning of group work (balanced work distribution and productivity) explicitly considering the following three issues:

Problem 1 (Privacy restrictions): Practical applications of predictive analytics in online courses underly several legal and privacy restrictions [8]. In general, the predictive analytics should be separated from individual assessment. If a course has enough participants a simple anonymization strategy based on pseudonyms can be sufficient to avoid the identification of individual users. The concept of *k-anonymity* [17] allows for controlling the degree of "resolution" of detail on the level of k individuals. In small group work, k-anonymity is no longer ensured for a sufficiently high k. Accordingly, the approach described in this paper only works with events that are not attributable to particular individuals and does not rely on features extracted from learner-generated content.

Problem 2 (Domain dependence of models): A well known problem when making predictions based on data gathered from assignment results is that the ensuing models are often very specific for particular tasks, courses, etc. therefore often not re-usable even in different instances of the same course [4]. In contrast, our approach only relies on the assumption that the group tasks are facilitated by general means of asynchronous communication (e.g. forums) between group members and collaborative content production (e.g. wikis or real-time collaborative editing).

Problem 3 (Amount of data / time needed to make reliable predictions): Predictions in group processes always suffer a cold start problem meaning that it takes a certain amount of time until enough evidence data is available to make a reliable prediction. Furthermore, it is very likely that the most activities in group tasks occur close to the deadline, and thus, the most data is available when an intervention is of limited effect. This paper explicitly addresses the question after how many hours after the start of a group task the outcome (at the end of a seven-day working period) becomes predictable.

The goal of this paper is to explore and positively demonstrate the feasibility of making meaningful predictions under the given restrictions. We are well aware that the underlying models could still be refined and improved.

LAK19, March 4–8, 2019, Tempe, AZ, USA

ACM ISBN 978-1-4503-6256-6/19/03…$15.00
https://doi.org/10.1145/3303772.3303826

2 BACKGROUND AND RELATED WORK

Big anonymous online courses such as MOOCs often offer forum discussions as an opportunity of exchange and collaborative activity. However, participation in these activities usually remains low and limited to a small number of users [9, 20]. Rosé & Ferschke [13] have proposed mechanisms similar to those used in recommender systems to stimulate discussions in such online learning communities. More "interventionist" approaches rely on the explicit assignment of roles (such as Asterhan & Schwarz [2]) or the pre-structuring of activities through scripts (cf. [10]).

The on-going IKARion project [16] aims at supporting small group work in large online courses through explicit task assignment combined with intelligent feedback and scaffolding. The project is based on the assumption that successful and productive cooperation will not arise from joint self-regulated activities alone but needs explicit support (cf. [3, 14]). In IKARion, the learning activities take place on a Moodle platform enriched with video materials, quizzes, forums and wikis. The group tasks make use of the forum especially for coordination and of the wiki facility for collaborative text production. Feedback and scaffolding interventions are triggered by a backend analytics component that monitors individual and group activities based on temporal aspects.

The work reported here aims at refining and testing the potential of these analysis techniques. It relies on the previous research of Doberstein et al. [7] in which sequences of collaborative interactions were made up of coordination and monitoring messages in the discussion forum as well as major and minor contributions to the text as the final product.

As proposed by Abbott and Tsay [1], sequence alignment techniques, originally known from bioinformatics for finding similarities in amino acid sequences, can be adapted to other fields. In our case, sequence alignment was utilized to characterize groups with similar collaboration patterns. By examining clusters of similar collaboration sequences, it could be shown that certain collaboration patterns go along with various metrics that assess the well-functioning of group work. One of the main findings was that early coordination counter-balances the expected effects of inactivity in the earlier phases of group work. In decision tree analysis, the presence or absence of coordination turned to be more important for the distinction between well-functioning and other groups.

This paper continues this line of research. The finding that sequential collaboration patterns can indicate the well-functioning of learning groups in ex-post analyses suggests that it might also be possible to use the same approach and coding scheme as input for predictive models. Such models would enable automatized interventions during the group work thus enabling to remedy otherwise not well functioning groups.

3 DATA ANALYSIS

3.1 Data Corpus Construction

This work builds upon the dataset assembled by Doberstein et al. [7] based on two instances of an online course on "computer mediated communication" (Course1: N = 270; Course2: N = 111). The courses were open to students of different study programs from two universities and students were supposed to take part in asynchronous small group work. Each topic that the course covered in successive course sections was introduced by a short lecture video, covering the most important aspects. In addition, students were provided with literature and self-test quizzes. Apart from these individual activities there were regular group assignments where groups of four persons had to produce a short text targeting a given problem. Since there were no face-to-face activities, learning groups were provided with a private forum for discussion and planning. Activities were collected from the discussion forums and the collaborative editors of the second course for five group tasks. Overall the activities of 65 different groups were recorded.

3.1.1 Independent variables. As a first step, the collaboration events collected for each group were classified into four categories. In order to not confound the actual prediction objective of this work by possibly misclassifications, no automatic labelling was applied but human evaluators annotated the corpus of group activities according to the following simple rules based on the coding scheme of Curtis and Lawson[5]:

(1) A message in the discussion forum of prospective character, e.g. planning and distribution of work was encoded as a *Coordination* event.
(2) A message in the discussion forum of retrospective character, e.g. reports of what has been done was encoded as *Monitoring* event.
(3) A revision of the collaboratively edited text that significantly changes or extends the content was encoded as *Major revision*. Based on manual judgement of the group tasks in previous work [6] revisions changing the text about more than 600 characters indicates a major revision as a rule of thumb.
(4) All other revisions of the text were considered as *Minor revisions*, which are typically fixing typos or rephrasing of particular sentences.

For this study, the events were hand-classified according to the rules but since the rules are not complex, classification could also be fully automatized in the future. Based on this rule-based classifications each group was encoded as a sequence of those collaboration events. Furthermore, inactivity was taken into account as well. Whenever the temporal distance between two successive events based on their timestamps exceeded 24 hours, a *Gap* "event" was inserted between them, which means that a full day without observable activities occurred. While this does not mean that the group members are not working on the task during this time, such gaps can cause problems in computer mediated group work when students become uncertain about the working progress of the other team members. Note that since sequences are built in a pre-specified period in time, such gaps can also occur at the end of a sequence if the last action took place before the endpoint of the sampling period. An example of a sequence representing the activities of a single group is given in Figure 1.

The constructed corpus of classified group activities over time serves as the only input to the classification models described later on. Every group work process is represented as a sequence of activities while the activities are described only by type and a timestamp without any user or content information. This modeling approach, consequently ensures privacy protection of learners on a very high level because from the input of the classification models it is only

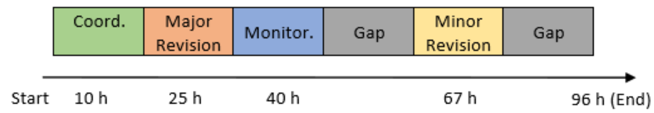

Figure 1: Example of a group timeline

Table 1: Target indicators for the well-functioning of group work

Aspect	Indicator	Mean (sd)
Productivity / Activity	Wordcount	704.23 (256.19)
Work (im)balance	Gini index (wordcnt.)	0.42 (0.18)

possible to observe when an event, for example a forum contribution, has taken place but not who posted to the forum. For this reason the data can even be shared with external analysts to build new models or as a subject for research on group collaboration patterns. This can be seen as analogous to vector space representations encoding texts as numbers, but in contrast capturing processes instead of content. This very generalized view on group work can be applied to many different contexts where group work is characterized by asynchronous communication events and collaborative editing. The drawback is that the lack of details does not allow to extract rich feature sets as input for predictive models.

Of course a predictive model can be integrated into a system so that no humans are involved in generating timely interventions. In those cases privacy is not an issue. However, when data is shared with third-parties, for example, analysts who want to build new models, there is certainly a data privacy issue. Whenever a vendor encodes group activities into sequences as outlined in the paper, however, there are no objections with sharing the data even publicly since it is almost impossible to identify an individual from such encodings.

3.1.2 Target indicators. In order to operationalize the notion of "well-functioning" groups, indicators for the two dimensions *Productivity* and *Work balance* (summarized in Table 1) are defined in the following. These indicators were calculated from the final product - in this case a collaboratively edited text.

Productivity / Activity: The most straightforward method to measure how active or productive a group is, is the *wordcount* of the delivered text. According to the instructions, groups were supposed to write at least 600 words. However, there are many groups that stay below this threshold as well as groups that produced much longer texts. This can also be seen from the standard deviation of the wordcount reported in Table 1. Hence the wordcount is still a distinguishing variable as also reported in previous analyses [7, 19].

Work (im)balance: It is desirable that every group member contributes to the final product in more or less the same magnitude. As a measure of work balance the normalized Gini coefficient of the number of words that each of the four group members contributed to the text was calculated. Lower values are better since the Gini index is *0* if all group members contributed equally and *1* if one did all the work alone.

3.2 Evaluation methodology

In order to answer the question whether the target indicators defined above are predictable early enough to provide a supporting intervention for not functioning groups, the 7 days period of a group task was split into 13 nested segments. The first segment covers the first 24 hours, next segments expand the previous segment by 12 hours until the last segment covers the entire 7 days.

The group sequences described before in Section 3.1.1 were created for each segment separately only based on events that occurred within the corresponding time range. In every segment a feature vector was constructed for each group where the elements denote the number of Coordination, Monitoring, Major- / Minor contribution events, and gaps contained in the group sequence. Consequently, the feature vectors can be built from a growing amount of data.

Next, for each segment two different regression models are evaluated by 10-fold cross-validation that take the group feature vectors of a segment as input and predict the target indicators described in Section 3.1.2. The goal of this study is not to find the best predictive model but to make general statements about how well the well-functioning of learning groups can be predicted to intervene where necessary as early as possible. For this reason, common regression models are used for making the predictions that have the advantage that the models can be predicted by humans such that the most important features for predictions in different points in time can be observed. In particular, standard linear regression and the M5 regression tree [12] are applied. While linear regression assumes linear relationships between the input features and the target indicators, M5 builds a decision tree from the training data with different regression models as leaves, and therefore, can also deal with non-linear relationships in the data.

As baseline, a null model is implemented that always predicts the average of the target indicators over all groups in the training set at any point in time. This baseline is certainly better than random guessing and can be seen as a consensus model over all data points, while it fails to closely adapt to individual data points if the variance is high enough. In the following a predictive model is considered to be reliable if it gives a lower error than the null model.

For further comparison extrapolation models have been created as well. These take the target indicators (e.g. wordcount) calculated up to a day $x \leq 7$ as input to a linear model predicting the actual target indicator after day 7.

3.3 Results

3.3.1 Predictability of group productivity. The accuracy of predicting the target indicators introduced in Section 3.1.2 at different points in time after the start of a group task is measured by the rooted mean squared error (RMSE) between the actual and the predicted indicator value based on 10-fold cross-validation. The results are summarized in Figure 2. Regarding the question about the predictability of well-functioning group work using generalized representations of group collaborations one can see that the work balance (Gini-index) and productivity (wordcount) can be predicted after some days after the start of the group task by linear models and regression trees indicated by a lower error than the baseline. Note that at the end of the group tasks extrapolation yields exactly

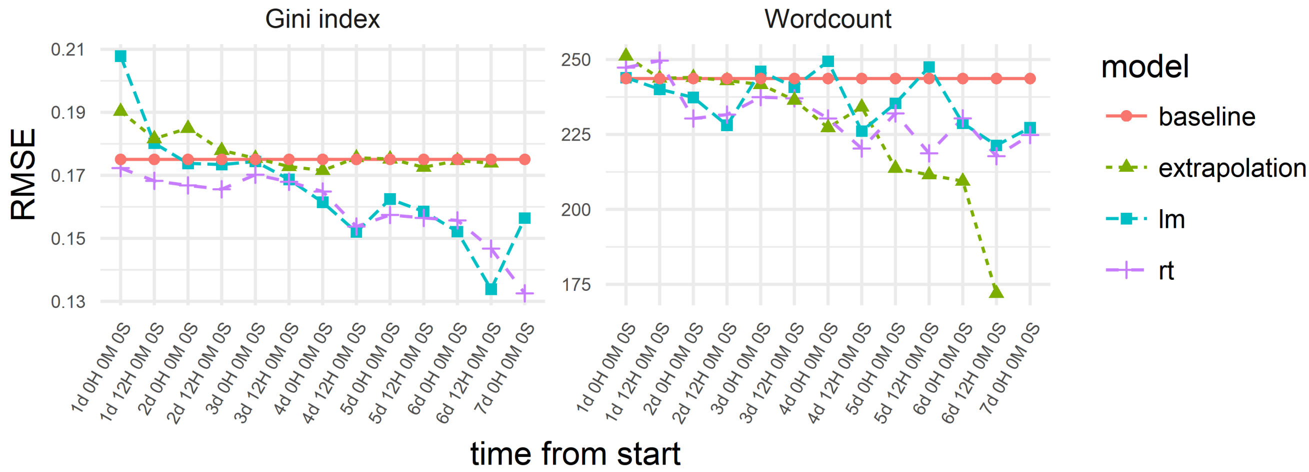

Figure 2: Rooted mean squared error at different times of predictions. (lm: Linear model, rt: Regression tree)

the target values, and thus, the corresponding data points are omitted. The final wordcount produced by the learning groups can be predicted quite early after 3 days. With respect to the balancedness of work distribution (Gini coefficient) reliable predictions can be made after 4 of the 7 days of the group task using sequential encoding (Section 3.1.1). Compared with direct extrapolation of the target indicators (green line) the encoded collaboration sequences perform equally for wordcount in early phases of the group work while extrapolating work imbalance does not work at all. This suggests that the presence of certain collaboration acts in early phases is important for a balanced work distribution at the end. This will be further investigated in Section 3.3.2. One can also see that in contrast to extrapolation, sequential encoding does not lead to continuous improvement of the predictions when the segments of data become larger. This is due to the fact that the sequential prediction models were not built incrementally but learned from scratch in every segment. At certain earlier time points, more reliable predictions are possible as compared to later points in times. An explanation can be that groups tend to increase their productivity close to the deadline and in these periods groups become less distinguishable.

3.3.2 Factors of well-functioning group work. In the following the predictive models are examined in more detail to identify the most important factors for estimate the target indicators for well-functioning group work. The coefficients of the linear models for productivity and work balance (WB) are summarized in Table 2, for both, a prediction model built from the data available at the 4th day (4d) of the group work and a model using the complete data after 7 days (7d). It can be seen that in both dimensions the number of coordination messages in the group forum is by far the strongest predictor after 4 complete days of activity. Surprisingly, the number of major revisions of the text in the first phase is not a significant predictor of final productivity. This suggests that in the early phase coordination and planning are the most important activities as compared to the actual text production. Note that coefficients for work balance

Table 2: Coefficients of the linear regression models derived after 4 days (4d) after start and the end of the group tasks (7d). Significance levels: $p \leq 0.001^{***}, 0.01^{**}, 0.05^{*}$

Coefficient	Prod. (4d)	Prod. (7d)	WB (4d)	WB (7d)
Coord.	44.1^{**}	25.04^{*}	-0.025^{*}	-0.01
Gaps	-8.1	6.44	0.017	0.011
Major rev.	-4.78	81.56^{***}	-0.033	-0.05^{***}
Minor rev.	-11.62	-22.3	-0.011	-0.02
Monitor.	19.74	23.11	0.016	0.01

are mostly negative since a lower Gini index of the group members' contributions indicates a higher work balance. Also periods of inactivity (gaps) are a weaker predictor than coordination messages. This is remarkable since intuitively one would rather focus on the number of major revisions of the text and a low number of activity gaps in order to judge whether a group needs help in solving the task.

From the coefficients reported in Table 2 alone it is not possible to fully understand the importance of different events for the predictions because most of the coefficients do not play a significant role since the associated event counts are too low at an early stage of the group work. Therefore, additionally the split points of regression trees derived from the activity data of the first 4 days of the group tasks are depicted in Figure 3 and 4. From top to bottom these models can be read as a ranking of contribution types.

These representations also highlight the importance of coordination messages in the discussion forum as the first feature one has to look at for estimating how well a group will work together. Although the number of major revisions and activity gaps of the edited text is not a significant positive predictor in linear models alone (see Table 2), these features can be seen as predictive variables if they are considered in the context of the presence or absence of sufficient group coordination.

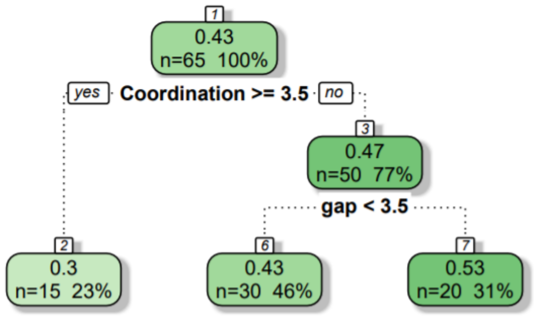

Figure 3: Regression tree of work (im)balance after 4 days from task start

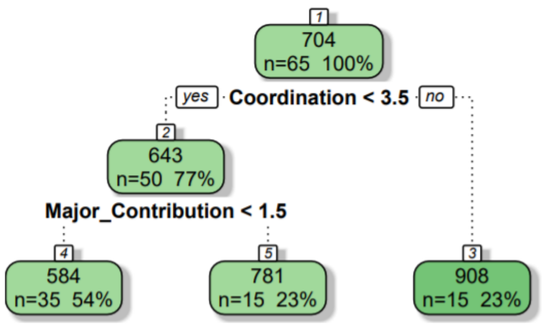

Figure 4: Regression tree of group productivity after 4 days

4 DISCUSSION AND CONCLUSIONS

This paper evaluated an approach to predict the well-functioning of learning groups in online courses only using activity data without any kind of user information. The reported initial results are promising and challenging for the future development of online learning group support at the same time. On the one hand, it could be shown that the expected productivity measured by word count and balanced distribution of work can be estimated before the deadline of a group task which is a prerequisite for targeted interventions. On the other hand, reliable predictions are in general possible after the first half of the group task period. This imposes further requirements to the design of intervention mechanisms since it might be very demanding for groups that already have problems to re-establish effective collaboration within the last 3 or 4 days before the deadline.

Apart from the predictability of productivity and work balance, another result of this paper is the interpretation of the prediction models: It could be revealed that the most important factor for well-functioning group work is enough coordination among the group members. If groups coordinate enough, other factors such as periods of inactivity and the number of contributions to the actual group product are not crucial. Thus, one can conclude that a possible way to promote effective group collaboration in online courses is to foster planning and coordination activities in the early phase, and to consider this aspect also in the design of group tasks. In future work, prediction models from the area of time series forecasting will be evaluated with the objective to improve, both, the accuracy of the predictions and the time when reliable predictions become possible.

REFERENCES

[1] Andrew Abbott and Angela Tsay. 2000. Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect. *Sociological Methods & Research* 29, 1 (2000), 3–33.

[2] Christa S. C. Asterhan and Baruch B. Schwarz. 2009. Argumentation and Explanation in Conceptual Change : Indications From Protocol Analyses of Peer-to-Peer Dialog. *Cognitive Science* 33 (2009), 374–400.

[3] Brigid Barron. 2000. Achieving Coordination in Collaborative Problem-Solving Groups. *Journal of the Learning Sciences* 9, 4 (2000), 403–436.

[4] Christopher Brooks and Craig Thompson. 2017. Predictive Modelling in Teaching and Learning. In *The Handbook of Learning Analytics* (1 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, and Dragan GaÅąevic (Eds.). Society for Learning Analytics Research (SoLAR), Alberta, Canada, 61–68. http://solaresearch.org/hla-17/hla17-chapter1

[5] David. D. Curtis and Michael. J. Lawson. 2001. Exploring Collaborative Online Learning. *Journal of Asynchronous Learning Networks* 5, 1 (2001), 21–34.

[6] Dorian Doberstein, Tobias Hecking, and H. Ulrich Hoppe. 2017. Sequence Patterns in Small Group Work Within a Large Online Course. In *Proceedings of the 23rd International Conference on Collaboration and Technology (CRIWG)*. Springer, Cham, 104–117. https://doi.org/10.1007/978-3-319-63874-4

[7] Dorian Doberstein, Tobias Hecking, and H. Ulrich Hoppe. 2018. Using Sequence Analysis to Characterize the Efficiency of Small Groups in Large Online Courses. In *Proceedings of the 26th International Conference on Computers in Education*. APSCE, Manila, Philippines.

[8] Hendrik Drachsler and Wolfgang Greller. 2016. Privacy and Analytics: It's a DELICATE Issue. A Checklist for Trusted Learning Analytics. In *Proceedings of the 6th International Conference on Learning Analytics & Knowledge*. ACM, New York, NY, USA, 89–98.

[9] Jonathan Huang, Anirban Dasgupta, Arpita Ghosh, Jane Manning, and Marc Sanders. 2014. Superposter Behavior in MOOC Forums. In *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S '14)*. ACM, New York, NY, USA, 117–126.

[10] Ingo Kollar, Frank Fischer, and Friedrich W. Hesse. 2006. Computer-supported collaboration scripts - a conceptual analysis. *Educational Psychology Review* 18, 2 (2006), 159–185.

[11] Sherry L. Piezon and Robin L. Donaldson. 2005. Online groups and social loafing: Understanding student-group interactions. *Online Journal of Distance Learning Administration* 8, 4 (2005), 1.

[12] John R. Quinlan. 1992. Learning with continuous classes. In *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*. World Scientific, 343–348.

[13] Carolyn Penstein Rosé and Oliver Ferschke. 2016. Technology Support for Discussion Based Learning: From Computer Supported Collaborative Learning to the Future of Massive Open Online Courses. *International Journal of Artificial Intelligence in Education* 26, 2 (2016), 660–678.

[14] Nikol Rummel and Hans Spada. 2005. Learning to collaborate: An Instructional Approach to Promoting Collaborative Problem Solving in Computer-Mediated Settings. *Journal of the Learning Sciences* 14, 2 (2005), 201–241.

[15] Thomas Staubitz, Tobias Pfeiffer, Jan Renz, Christian Willems, and Christoph Meinel. 2015. Collaborative learning in a MOOC environment. In *Proceedings of the 8th Annual International Conference of Education, Research and Innovation*. 8237–8246.

[16] Sebastian Strauß, Nikol Rummel, Filipa Stoyanova, and Nicole Krämer. 2018. Developing a library of typical problems for collaborative learning in online courses. In *Proceedings of the 13th International Conference of the Learning Sciences (ICLS)*. London, UK, 1045–1048.

[17] Latanya Sweeney. 2002. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, 05 (2002), 557–570.

[18] Gaurav Singh Tomar, Sreecharan Sankaranarayanan, and Carolyn Penstein Rose. 2016. Intelligent conversational agents as facilitators and coordinators for group work in distributed learning environments (MOOCs). *AAAI Spring Symposium - Technical Report* SS-16-01 - (2016), 298–302.

[19] Astrid Wichmann, Tobias Hecking, Malte Elson, Nina Christmann, Thomas Herrmann, and H. Ulrich Hoppe. 2016. Group Formation for Small-Group Learning: Are Heterogeneous Groups More Productive?. In *Proceedings of the 12th International Symposium on Open Collaboration (OpenSym '16)*. ACM, New York, NY, USA, 14:1–14:4. https://doi.org/10.1145/2957792.2965662

[20] Jian-Syuan Wong, Bart Pursel, Anna Divinsky, and Bernard J. Jansen. 2015. Analyzing MOOC discussion forum messages to identify cognitive learning information exchanges. *Proceedings of the Association for Information Science and Technology* 52, 1 (2015), 1–10.