



# Using Sequence Analysis to Determine the Well-Functioning of Small Groups in Large Online Courses

H. Ulrich Hoppe, et al. *[full author details at the end of the article]*

Accepted: 8 November 2020 / Published online: 17 December 2020

© The Author(s) 2020

## Abstract

Collaborative learning in small groups can enrich and enhance the learning experience in large online courses by facilitating interaction and collaborative knowledge building between peers. The work reported here addresses scenarios based on asynchronous communication and exchange. As compared to synchronous and face-to-face settings, these scenarios require higher explicit coordination efforts and specific support mechanisms to reach out to inactive group members. This can be achieved by human tutors as well as through automatic interventions by the system in the organization and facilitation of group work. Our approach is based on characterizing the quality or well-functioning of collaboration using sequence alignment techniques that were originally developed in bioinformatics for the comparison of DNA sequences. Sequence alignment provides a similarity measure between pairs of action logs originating from the group work. The ensuing similarity matrix allows for clustering the working groups characterized by the similarity of their complete action logs. Notably, these clusters differ clearly in terms of quality measures related to the activity distribution and group output. This allows for determining quality indicators that can trigger feedback and adaptive scaffolding. Our data stem from inter-university online courses in which regular (weekly or bi-weekly) assignments had to be completed in small groups supported by an online forum for coordination and a separate tool for the actual collaborative writing. The basic actions are interpreted/coded as different contribution types, namely coordination, monitoring, minor/major contributions. Longer periods of inactivity are coded as another type of action (“gap”). These are the building blocks for the sequence alignment. One consequence of our analysis is a revision of the intuitive assumption that inactivity is a main indicator of inefficient group work: We found that inactivity can be compensated by early coordination to the extent that coordination is even the most important factor for successful group work. These insights are now being used to design alert mechanisms to signal inefficient group work and to generate timely interventions.

**Keywords** Online courses · Small learning groups · Sequence analysis · Sequence-based clustering · Collaboration quality

## Introduction

In MOOCs and other types of current online learning courses the main activities are video watching and regular assignments such as self-test quizzes. On the positive side, this allows for supporting individual learners in self-directed knowledge acquisition independent of location and time. However, in the basic version of such online courses participants are left alone without support for individual learning problems. It has been argued that online courses should be better adapted to individual learner needs and learning styles by offering specific assistance and add-on learning activities (Grünewald et al. 2013). For large online courses, this can of course not be achieved through human-orchestrated tutoring services. Although the participants of an online course may be considered as a virtual learning community, these scenarios tend to make little use of peer-to-peer interaction and collaboration. Often, there is not more than a discussion forum to which questions and answers may be posted without additional support mechanisms. This neglects the potential of exploiting and cultivating peer-to-peer interaction and exchange as a powerful learning resource.

Already more than twenty years ago, the idea of using peer interaction and support as a powerful and widely available learning resource in asynchronous networked learning environments was a core principle and motivation for designing the I-Help system (Greer et al. 1998). I-Help was practically introduced and evaluated in technology-rich higher education settings. The Peer Help System module (PHelpS) in I-Help made use of intelligent techniques, including domain knowledge and student models, to recommend peers that would be particularly suited to support other learners with specific problems. It also contained an asynchronous help forum that was similar to forums that are now available for many online courses. Based on experience with the I-Help environment, Bull et al. (2001) have analyzed the interplay and ensuing learning effects of help-seeking and help-giving with an asynchronous help forum. Indeed, it could be corroborated that peer-to-peer learning support was beneficial for both helpers and helpees. Along similar lines, Hecking et al. (2017) studied the dynamics of learner profiles in MOOC (“Massive Open Online Courses”) forums identifying specific combinations of social role patterns and thematic interactions (help seeking/giving related to certain topics). In the discussion of their findings, already Bull et al. (2001) indicated that forming small learning groups with specific task assignments in larger online courses should be beneficial for stimulating engagement and collaborative knowledge building. A specific assumption is that a group setting with a smaller set of social ties as compared to larger groups would help learners to move into more central and active roles over time (Wegerif 1998; Wortham 1999).

Targetting MOOCs and other large online courses, several suggestions have been made to facilitate collaborative learning through system-side mechanisms. These include the automatic formation of small groups or teams with regular task assignments (Erdmann et al. 2017; Staubitz and Meinel 2017) as well as the stimulation of discussions between MOOC participants (Ferschke et al. 2015). In the latter approach, conversational agents have been introduced to steer and coordinate group interactions (Tomar et al. 2016).

The work reported here is in the line of small learning groups with task assignments. A first aspect of potential intelligent system support in such scenarios is “group formation”, i.e. the determination of most promising group constellations based on

prior knowledge of students in relation to expected performance on tasks. This is a well-known issue of Artificial Intelligence in Education (AIED) research. It also relates to the educationally inspired assumption that knowledge exchange would be better facilitated in knowledge-heterogeneous learning groups. There are indeed indications that groups with heterogeneous knowledge levels can lead to better overall learning gains in the group (Webb et al. 2002), with the caveat that the differences in knowledge levels should not be too high (Gijlers and De Jong 2005). In the AIED tradition, the “multiple student modeling” approach (Hoppe, 1995) was an early suggestion for formulating and operationalizing decision rules that could determine the formation of both complementary as well as competitive learning groups. These ideas were taken up and extended in I-Help and PHELPS (Greer et al. 1998) in large-scale practical applications. Also in the context of AIED research, Inaba et al. (2000) have proposed an ontology-based approach to group formation that explicitly represents assumptions from underlying learning theories that are formalized and encoded in the ontology.

More recently, the group formation challenge has been addressed from a general algorithmic perspective as a multi-dimensional optimization problem. In this line, GroupAL (Konert et al. 2014) is a prominent approach that allows for mixing different target criteria (such as gender-homogeneous and knowledge-heterogeneous groups) to achieve an overall optimal grouping in a given set of students. In our recent work we have used GroupAL as a plug-in to the Moodle learning platform to generate heterogeneous (as compared to random) learning groups based on previous course activity (Wichmann et al. 2016). Indeed, there was a positive effect of this treatment on group productivity in a collaborative writing task. However, we also found issues in the coordination and well-functioning of groups that called for specific support after the groups were formed. On this basis, we came to the conclusion that implementing strategies for detecting deficits in the actual group work and providing feedback to overcome these deficits might be even more important than intelligent group formation. This is in-line with findings by Wang et al. (2017), indicating that group formation strategies can be synergetically complemented with dialogue prompts to achieve a better outcome of the group work. Such strategies are particularly important for large online courses in which group work takes place asynchronously, since this requires additional specific efforts in coordination and in collaborative production under the given limitations of the technical environment.

There is evidence that typical problems of group work such as social loafing, i.e. group members reduce their effort to contribute to the group success and rely on the work done by others, and a lack of commitment of the members are more frequent in online courses as compared to presence settings due to anonymity and limited communication facilities (Piezon and Donaldson 2005). Low productivity or even inactivity of single members can negatively affect the learning experience of the other members. Especially, longer periods of inactivity can cause uncertainty about the willingness of group members to actually participate and contribute. In general, it is desirable to support and increase the feeling of social presence in online courses since there are indications that this has a positive effect on task performance (Weinel et al. 2011). For our target scenario of small group work integrated with larger online courses, this implies that we should try to foster the mutual awareness between group members and provide mechanisms for explicit coordination to mitigate the uncertainty about the participation of the others. The

application of corresponding feedback and scaffolds should be based on an identification of critical conditions.

The main point and guiding rationale of the work described is to start from the analysis and comparison of action sequences, i.e. from the dynamics of the group interactions, to further characterize and compare the quality of collaboration. We claim that the method of sequence analysis, which will be explained in the following chapter, can serve as a kind of microscope to capture differences in the underlying collaboration processes. In this sense, our primary goal is to contribute on the level of the analytic methodology, we do not start from a specific hypothesis about conditions and effects related to online learning. Once we can compare and cluster collaborative learning based on their process dynamics, we characterize the clusters found in terms of quality indicators. Here, we have particularly been inspired by the work of Meier et al. (2007). Among the quality factors or dimensions introduced there, the dimensions “task division”, “time management”, and “technical coordination” were found to be most highly correlated with individual awareness of good collaboration in a post test. We have been able to distinguish our clusters along very similar lines. Using a decision tree analysis, we have identified the relevance of specific factors for the belonging to the one or the other of the clusters. This has led us to revising a standard assumption about predicting the quality of group work based on participation. The prospect and promise of this approach is to induce and fine-tune decision rules as triggers for adaptive interventions in terms of scaffolding or feedback to improve collaboration quality in asynchronous online settings. After preliminary work in this line of research, the approach has been taken up and further pursued in the research project IKARion (cf. Krämer et al. 2017; Constapel et al. 2019).

This article summarizes developments and findings from several studies, two of which are reported in chapters 4 and 5 in some detail. The background and results of the first study (chapter 4) had already been published as a conference paper (Doberstein et al. 2018). Prior work had been focused on the effects of automatic group formation using similar analysis techniques (Doberstein et al. 2017). Since the empirical basis was small and the findings were inconclusive, we did no longer limit our scope of analysis to the effects of automatic group formation (already in study 1). The follow-up study reported on in section [Replication and Refinement \(Study 2\)](#) is based on new data and has not been published yet. The final “[Discussion and Outlook](#)” section will take up the question how the methodology can be practically applied to improve learner support in asynchronous online settings.

## Using Sequence Analysis to Characterize Group Work

Reimann (2009) has pointed out the relevance and the potential of analyzing time series of user actions (events) as compared to simple aggregating “coding and counting” approaches to better understand the process characteristics of Computer-Supported Collaborative Learning (CSCL) interactions. He sees a particular relevance of such analyses for longer term collaborative interactions as they occur in asynchronous networked learning communities. In a similar orientation, Abbott and Tsay (2000) and Cornwell (2015) have adopted sequence analysis techniques to the study of social interactions. Sequence analysis as a method has its origin in bioinformatics where it is

used to compute the similarity of amino acid sequences in proteins or DNA strings (Needleman and Wunsch 1970). This similarity measure is based on “edit distance”, i.e. the computational cost (in terms of minimal number of “edits”) needed to transform one string into another using a certain repertoire of transformations (such “delete”, “insert”, “switch”, etc.). This process is also called “alignment”. It has to be clearly distinguished from sequential pattern mining techniques (Fournier-Viger et al. 2014), which are used to detect specific finite sequences of tokens that occur with a certain frequency in a given sample. The result of sequential pattern mining can be predefined or induced (learned) arrangements of events or items that have a temporal order but not necessarily a direct “followed-by” relation. Sequence analysis does not deal with such reoccurring ordered sequence patterns as parts of given strings but with the similarity of whole strings in terms of their alignment measured by edit distance.

Our approach is based on the sequence alignment between collaboration sequences that arise from different learning groups working on tasks assignments in online course settings. In a first (pre-)processing step, human coding is used to categorize actions in the form of different contribution types: coordination, monitoring, minor contribution, and major contribution. In addition, another descriptor (“gap”) is introduced for one-day inactivity in the group work. Sequences of such action descriptors are then collected for portions of group work corresponding to a specific task assignment. The pair-wise similarities (based on edit distances) between the sequences are captured in a similarity matrix, which forms the basis for a cluster analysis. This processing chain is illustrated in Fig. 1. The ensuing clusters are examined and compared in two aspects, the distribution of inactivity and coordination actions. They can also be compared in using quality measures for group work such as productivity or balanced of participation.

Recently, Boroujeni and Dillenbourg (2019) have presented a related analytic approach that links activity sequences of MOOC participants to their overall studying behavior. The basic activities in this approach are related to video watching, submission

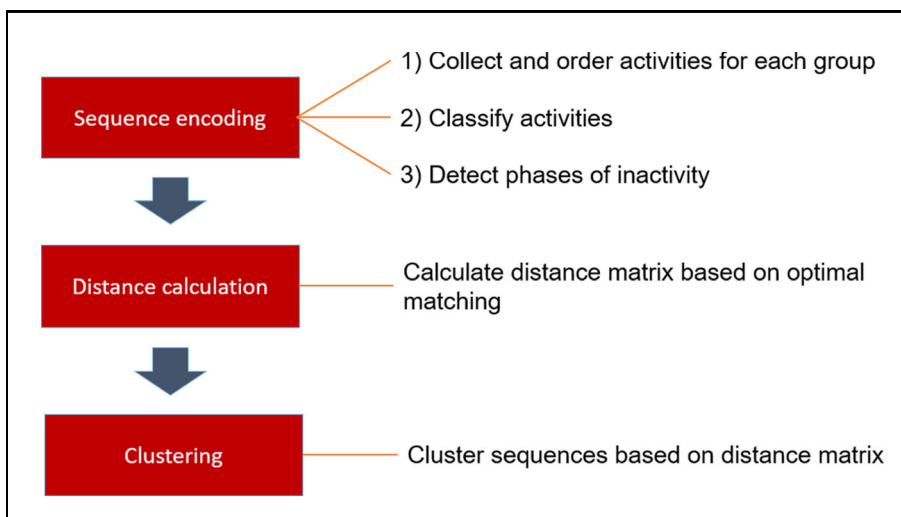


Fig. 1 Phases of sequence analysis and clustering

of solutions to assignments and submissions to a discussion forum. Their analysis uses two different strategies characterized as (1) hypothesis-driven and (2) data-driven. The hypothesis-driven approach performs pattern recognition with predefined sequential of interactions with lecture materials, videos and assignments on the given action logs. One such pattern is the initial interaction with video resources and a subsequent submission of assignments. It was found that 44% of the learners in a MOOC first watched the lecture videos and then completed the given assignments. It was observed that learners temporarily changed their study approach during periods of the course. E.g., learners skipped the videos watching activities for certain topics and submitted assignments without further engagement with lecture material but continued with initial patterns later in the course. 2% of the learners did not interact with video resources for the whole duration of the course. Boroujeni and Dillenbourg (2019) interpret these differences in strategies as indicators of difficulties in the learning process. A shortcoming of the hypothesis-driven approaches is that typical interaction patterns have to be predefined, which implies that unexpected but possibly relevant behavior will not be detected. In contrast, the data-driven approach does not require prior specified patterns. In this approach, unsupervised learning methods are utilized to find patterns in the interactions between learners and course material. Among these, the most frequent of the detected patterns describes video watching activities followed by an inactive period. The second most frequent pattern depicts inactive users that do not access course materials or submit assignments. This pattern describes the typical drop out behavior which is common for online courses.

Our approach is data-driven (type 2) in the sense of the distinction introduced by Boroujeni & Dillenbourg, but it differs from (2) in that it is based on sequence alignments not based on sequential patterns. In the sequel, we will report on two iterations of applying our method to different instances of a higher education online course.

## Learning Scenario and Initial Findings

Our data source and target scenario is an online course on “computer-mediated communication” (CmC) shared between two major universities in the north-western region of Germany. This course has been offered repeatedly since 2015. The course takes up a series of related topics in the perspective of psychology and the learning sciences (including topics such as “grounding”, “information sharing”, “social presence”, or “CSCL scripts”). The course has been open to students of a variety of study programs from two universities as an elective in “optional studies” (i.e. outside the domain-specific curriculum). Although the students were enrolled in presence-based programs, this course was delivered only online on a Moodle platform including tools for collaborative writing and group forums for communication. Additionally, the platform was modified in such a way that collaboration activities could be fully recorded for latter analyses. The students were instructed that all communication and writing activities had to take place using the dedicated Moodle tools. Although, we cannot completely exclude that the students did not occasionally move to other communication platforms, by closely examining the discussion transcripts no clear evidence was found that other communication tools than the provided ones were used.



Each course section, students were provided with a short introductory video, reading materials, and self-test quizzes to expose the relevant content. For each one of the bi-weekly group assignments, students were assembled into new groups of four participants. These assignments were collaborative writing tasks in which certain subtopics had to be elaborated in an online writing tool. These essays were expected to contain at least 600 words. The students were instructed that all communication and writing activities had to take place in Moodle. Since students came from many different study programs, it was unlikely that they would “discover” peers that they already knew in one of the small learning groups. This corresponds to the typical anonymity condition in a MOOC scenario. On the other hand, different from many MOOCs, this course would give the students actual credits in their respective programs, which explains the limited attrition rate in the range that we also see in some presence courses (around 50% based on the original inscriptions). The inclusion of student data in our analyses was based on informed consent, allowing course participants to “opt out” from being part of the study. These students could still follow the course and were assigned to learning groups together with other “opt out” students.

Technically, the course was run on a Moodle platform and most students were already familiar with Moodle. Two existing Moodle plugins were adapted to support the groups in solving the tasks. For text creation students had to use the collaborative editor Etherpad, which was integrated in the learning platform to enable real-time collaboration. Discussions and coordination activities were supported by separate discussion forums for each group. Discussion forums and Etherpads were linked such that the students could constantly switch between the two. In later instances of the course, the Etherpad tool was replaced by the Wiki tool available as part of the Moodle environment.

One of the first practical experiences was that it was challenging to maintain an adequate level of activity of the learning groups. Longer inactive periods (“gaps”) in the group work appeared as a serious problem, especially in the beginning of the working period. Inactivity also caused uncertainty about the sheer presence and commitment of the other group members even though the list of group members was always visible on the platform. This became evident since in such cases tutors frequently received messages from students expressing their doubts about persons being actually allocated to the same group. Furthermore, in the second study (see chapter 5), a relationship between satisfaction with group work and overall satisfaction with the course could be found based on survey data collected from the first instance of the course (Kyewski et al. 2016). Based on these experiences, significant improvements in course satisfaction were achieved in the next instance of the course through a clearer structuring of the group activities and more precise guidelines for solving the group tasks (Erdmann et al. 2017).

Apart from evaluating the influence of different types of group assignments, one major research goal was to determine the influence of group formation strategies on group productivity (Wichmann et al. 2016). Group productivity was a composite measure comprising the volume of the actually delivered text, domain concepts used in this text, and forum contributions as a quality measure. The parameter used in group formation was based on forum contributions (amount of text) during the previous group tasks. Here, three levels of student activity (“high”, “average”, or “low”) were distinguished. As a result, groups solely composed of students with high previous activity (“homogeneous-high”) showed the highest productivity, as expected. Heterogeneous groups of high-, average-,

and low-level students showed slightly worse productivity than these, but better than the other two homogeneous groups (only average-level or only low-level students). It was also interesting to see that high-level individual students were more productive in heterogeneous groups, whereas low-level students were more productive in homogeneous groups. A plausible explanation would be that low-level students would have to take over a greater part of the work themselves in the homogeneous-low groups. When interpreting these findings, it is important to bear in mind that the performance levels were based on prior activity and not on pre-knowledge.

Based on these findings we saw a need for looking at the actual activities in the group as a determining factor for the well-functioning of groups beyond their initial composition. In this context, both process and product characteristics would be relevant, the product characteristics would rather be used as an *ex post* quality criterion whereas process features (represented as sequences of activities) would characterize the work on a behavioral level. This brought us to considering sequence analysis as an adequate analytic approach.

## Collection and Sequence Analysis of Course Data (Study 1)

The sequence analysis approach was first applied to an instance of the CmC course in which we had introduced informed group composition with four types of groups (see above) in a subset of the task assignments. This initial study reported in Doberstein et al. (2017) used this subset comprised of 19 groups that were subjected to the sequence analysis clustering. The clusters derived through sequence analysis showed interesting differences in terms of late vs. early start (distribution of gaps) and number of coordination actions, but there was no clear dependency between group formation types and the corresponding clusters. As a consequence, we extended the sequence analysis to all 65 groups established during this course. The additional 46 groups were based on random assignments. The group members were drawn from 81 participants who participated in the course and gave their consent to being included in the research study. This instance of the CmC course used Etherpad for the text production and the Moodle discussion forum for communication and coordination. In this sense, all groups had identical working conditions once they were formed.

### Pre-Processing / Coding

Action logs on the group level were collected from the Moodle discussion forum and the Etherpad writing tool in a uniform way and aggregated in one database with synchronized time stamps. The chat function of Etherpad was disabled such that communication happened solely in the forum. The granularity of forum contributions was always a single post (as one action). Continuous Etherpad input originating from the same user without interruptions of more than 60 min or changing to discussion mode was coded as one textual contribution. However, such contributions would be assigned to the group and not to an individual user in the action coding.



Class	Example
<b>Major Contribution</b> - Addition to the text > 600 characters	
<b>Minor Contribution</b> - Addition to the text < 600 characters	
<b>Coordination</b> - Prospective Message - Organization / work distribution	<i>"I would like to write the introduction, but I will not have time until tomorrow."</i>
<b>Monitoring</b> - Retrospective Message (often following own contribution)	<i>"I finished the discussion part. Could you have a look if it fits?"</i>
<b>Gap</b> - 24 hours of inactivity	

Fig. 2 Coding scheme for the classification of actions

As depicted in Fig. 2, the next analysis step applied to the group action logs was a human coding of action types. In order to be able to identify inactivity at the beginning of a group task, an artificial start event was added with a time stamp pointing at the release of the task. This action classification was inspired by the coding scheme introduced by Curtis and Lawson (2001), also in a CSCL context. Textual contributions into Etherpad were classified as “major” or “minor”. The idea was to distinguish contributions that added a considerable amount of text and extend the semantic content of the text (major) from small improvements in spelling or smaller text modifications or spelling corrections (minor). Based on a qualitative screening of examples, the distinction was further operationalized using a numerical threshold of 600 characters. Forum posts were classified as “coordination” or “monitoring”. Coordination was used for prospective, forward-looking messages, often related to the planning and distribution of tasks. In contrast, monitoring was used to capture messages of retrospective type typically referring previous posts, text contributions or misunderstandings related to previous messages. Forum posts that could not be subsumed under one of these categories were skipped and deleted from the sequence since they do not provide any valuable information. Such deleted posts were very rare and typically concerned posts like “Bye”, “Thank you for the good work”, etc. The resulting sequence of action types would no longer contain time stamps. However, to capture longer inactivity we introduced another type label: Whenever the action sequence in the original database for a given group showed a difference of more than 24 h between two consecutive actions a “gap” item was included as an additional code.

The “gap” and “contribution” classifications could have been generated automatically but particularly “coordination” and “monitoring” (as well as skipping) were dependent on human judgement. Originally we had one human coder. To check the reliability of this coding, we had a second person re-coding a random selection of 20% of the source material. The ensuing inter-rater reliability (by Cohen’s Kappa) was 0.79, which we consider as sufficient. Figure 3 shows an example sequence for one of the learning groups. We see some inactivity at the beginning (even two idle days), followed by a phase with several coordination actions and major contributions, and the final half

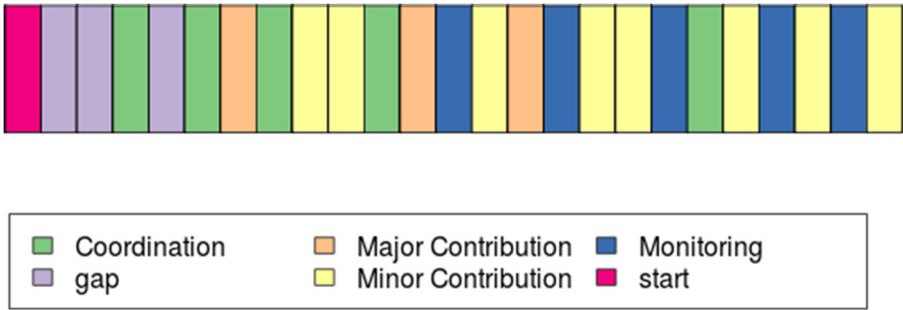


Fig. 3 Example sequence of action types (codes) for one group

dominated by monitoring and minor contributions. Overall, this example shows the profile of a relatively well-functioning group.

### Sequence Alignment and Clustering

The following sequence alignment step computes the similarity between each pair of token sequences in terms of edit distances (cf. Abbott and Tsay 2000). The similarity values correspond to the minimal number of transformation steps (insertion, deletion and substitution) needed to convert one token string into the other. The underlying algorithm is a variant of the well-known Levenshtein distance calculation (Damerau 1964). The method allows for associating each type of matching operation with a specific cost (i.e., a weighting factor). In our alignment of collaboration sequences, the cost for insertion and deletion was set to 1. This was also used for substitution unless a token of type “gap” was involved, either as a target (to be replaced) or as the replacing token. In cases with gaps the cost was 2. This weighting factor reflects the importance of inactivities as indicators of problems in group work.

Edit distances can be directly interpreted as similarities, a smaller number of edits indicating a higher similarity between two different pairs of action sequences and the corresponding groups. The pair-wise similarity values together make up the similarity matrix (or distance matrix). The resulting matrix is the basis for grouping the sequences into clusters using the approach of “partitioning around medoids” (PAM - cf. Kaufman and Rousseeuw 2009). PAM builds clusters around  $k$  different medoids in such a way as to maximize the similarity (i.e. minimize the distance) between all objects and their closest medoid. The build phase in which the medoids are selected is followed by a swap phase in which the clustering is optimized by the switching of objects between clusters. This process is based on a predefined number of medoids and thus clusters (value  $k$ ). To find a proper clustering, the algorithm first searches for a suitable set of representatives (build phase). We have examined values of  $k$  ranging from 2 to 4. To select the best  $k$  for the given dataset, we have applied a silhouette analysis (Rousseeuw 1987) using the measures of average distance inside the clusters (preference for smaller) and average distance between the clusters (preference for higher). The best result was achieved for  $k = 3$ . Figure 4 shows the resulting optimal clustering for the 65 learning groups and corresponding action sequences. It comprises three clusters with 19, 17, and 29 sequences, respectively.

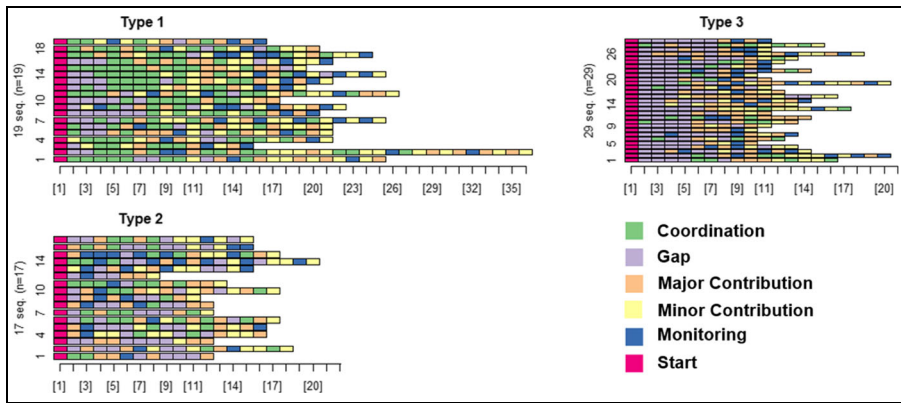


Fig. 4 Optimal clustering for 65 groups / sequences (according to PAMK) resulting in 3 different clusters

Differences between the three clusters can be interpreted as differences between corresponding groups. To characterize the clusters (or groups, respectively), we can distinguish measures that are derived from the distribution of actions (behavioral) and measures that characterize the productivity or the overall work pattern of the group. The basic productivity measure is “word count”, i.e. the number of words in the final version of the text produced by the group using the Etherpad tool. As it was indicated in the assignments that each group should write at least 600 words, the word count is of limited significance. However, some groups exceeded this threshold by several hundreds of words so that the word count can still be seen as a measure of engagement and group productivity. Note that word count only reflects the quantitative amount of text produced. Alternatively, one could also calculate a “semantic word count”, which would only reflect the number of important domain concepts. This, however, would require the specification of word lists for each group task prior to the analysis, which can hardly be complete. Since the group tasks cover different topics the number of meaningful words can differ. To ensure comparability of tasks and not to introduce additional biases through the usage of hand-crafted word list, in this work we stick to the simpler word count as it gives a good indication of the overall productivity of a group. A typical measure for the quality of group collaboration is the balancedness of contribution between the participants. We have used the Gini index to calculate this feature. The Gini index is a measure of the deviation of a distribution from an (ideal) uniform distribution with values normalized between 0 and 1 (cf. Cowell 2011). For an equal distribution in which every group member contributed the same amount the Gini index would be 0. It would be 1 in the extreme case that all the work was done by one

Table 1 Behavioral and output parameters per cluster

Cluster	Word count	#Gaps (first half)	#Coordinate (first half)	Work imbalance (Gini index)
1	830	2.7 (1.89)	6.7 (4,56)	0.36
2	695	3.6 (1.84)	2.4 (1,13)	0.4
3	628	4.8 (4.13)	1.2 (0,23)	0.5

participant. In this sense, lower Gini values indicate a higher degree of balancedness. Table 1 shows the average values for word count, number of gaps, number coordination actions, and work balance (Gini) for the groups in each of the three clusters. The action-based measures (no. of gaps and no. of coordination actions) are also counted for the first half of the working period.

Cluster 1 with 19 groups has the highest average productivity (word count) and most balanced average participation, followed by cluster 2 with 17 groups. From Fig. 4 it can be seen that even groups in cluster 1 can have activity gaps at the beginning. However, by closely examining the discussion transcripts it was found that such periods of inactivity were often followed by coordination messages. For example, group 10 in cluster 1 had several days at the beginning because some of the group members were not available in the first days. Later on, however, the group could recover because two members took the lead in coordinating the activities and the members clearly stated when they will be able to make contributions. The biggest cluster 3 (29 groups) has the lowest word count and the highest imbalance of contributions. Pairwise t-tests between the three clusters regarding the Gini values revealed a significant difference between cluster 3 and the other two clusters, but no significant difference between clusters 1 and 2.

It is of particular interest to determine the influence of the action-based parameters on the output-oriented parameters since the actions can be observed and monitored continuously and might be used as indicators of potential suboptimal group performance or failure already during the working period. It is commonplace to assume that inactivity is such an indicator of suboptimal performance: “To find out which groups are not doing well and might need special help just look at the inactivity!”. This is also corroborated by the gap counts with an even sharper distinction for cluster 3 when looking at the occurrences in the first half of the working period. However, the average number of coordination actions shows a clearer advantage for cluster 1, again sharpened if counting is reduced to the first half of the time window.

## Decision Tree Analysis

To further investigate the influence of the different activity profiles (gaps and coordination actions), we have generated a decision tree for determining to which cluster a given group would belong based on the action counts as input values. This was implemented using the CART algorithm (Breiman, Friedman, Olshen, Stone, 1984) available in R (Therneau & Atkinson, 2017). The results is shown in Fig. 5.

Given that the attributes used (number of gaps and coordination counts) were numerical, the regression-based version of CART was applied. By the nature of this construction, at every non-terminal node a binary split is introduced that takes the attribute and criterion that gives the highest information gain for the ensuing classification. For the given data, this means that the degree of coordination discriminates better than any other of the parameters (here: inactivity) in explaining the split of the sequences into clusters. The terminal nodes give the actual classifications after applying one or two splitting criteria. The red (leftmost) leaf node is comprised of 78% of nodes belonging to cluster 1, 13% belonging to cluster 2, and 9% belonging to cluster 3. The share of elements belonging to cluster 1 and ending up in the red leaf node is not directly indicated but corresponds to  $0.35 * 0.78 / 0.29 = 0.94$ , i.e. 94% of the cluster 1

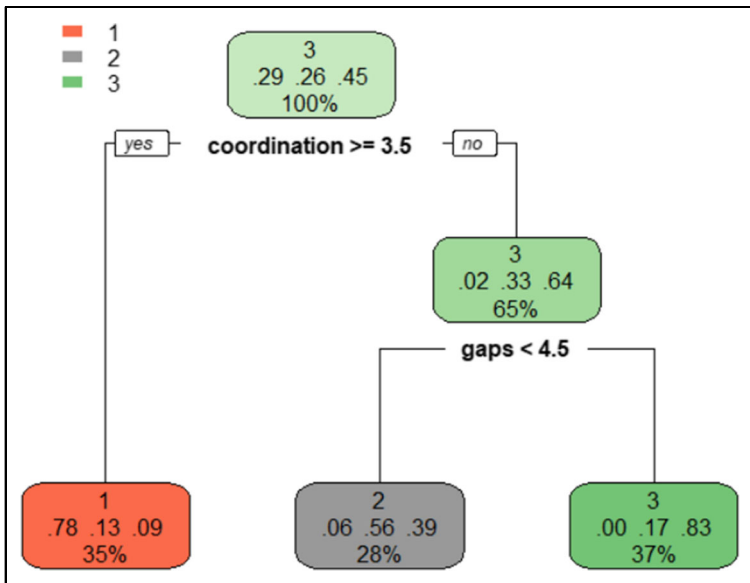


Fig. 5 Decision tree for determining clusters based on action parameters

elements will be subsumed under this node. Similarly, 68% of the cluster 3 elements end up in the green (rightmost) leaf node. The grey (middle) leaf node is dominated by cluster 2 but also contains a considerable share also of cluster 3 elements. Based on the principle of maximizing information gain in every split, we can conclude that the number of coordination actions with the separating value of 3.5 is the best discriminator and already separates out more than 90% of cluster 1 with relatively small mix-ins from the other clusters.

Based on this result, we have to revise the common sense assumption that would see inactivity (gaps) as the most important indicator for the success or failure of group work. Indeed, the number of coordination actions turns out to discriminate better. In other words: coordination actions (which typically occur in the first half of the group working period) can to some extent counter-balance the likely negative effects of inactivity. This should be considered in the design of scaffolding and support mechanisms for small working groups.

## Replication and Refinement (Study 2)

The dataset reported on in the preceding section was collected in the context of a local cooperation project that took a combined perspective on intelligent group composition followed by monitoring and support. Our findings related to indicators for the well-functioning of groups shifted our attention even more to the monitoring and support side. In a nationally funded follow-up project (IKARion - cf. Krämer et al. 2017) we have further investigated the role of scaffolds integrated with the learning platform. In addition to the Moodle platform, the IKARion system architecture features a dedicated learning analytics backend connected to an expert system for generating the potentially

adaptive scaffolds rendered in turn in Moodle (Constapel et al. 2019). In the context of this project, we have been able to replicate our analysis with an extended set of variables characterizing the group work.

The basis for this extended/replicated study was again the CmC course in an instance delivered during the winter term 2018/19 in the same setting between the same two universities. In this instance, the internal wiki tool of Moodle was used instead of Etherpad. Again, the organization of the work was supported through the Moodle discussion forum. Groups had again four participants, but the working period for each assignment was extended from one to two weeks. This study used different conditions regarding scaffolding and feedback during the different task assignments and ensuing working periods. However, here we are only interested in seeing how the sequence alignment followed by the PAM-based clustering can separate the working groups according to the activity characteristics and how this separation corresponds with variables that characterize the success and well-functioning of the working groups, namely productivity, satisfaction, and work balance.

The sample used for this analysis is based on two task assignments with 16 and 13 groups. Based on an ANOVA for these 29 groups in total, there was no significant influence of the feedback/scaffolding conditions on the target variables (productivity, satisfaction, balance). The PAM clustering applied after sequence alignment produced a good separation (see criteria discussed in the previous section) with only two clusters. The clusters resulting from the sequence analysis are shown in Fig. 6. Cluster 1 contains 19 and cluster 2 contains 10 sequences. Already at a first glance, we can see that cluster 2 has considerably more coordination actions and less gaps. Accordingly, we would expect that the groups in cluster 1 worked better.

Table 2 displays the average (per cluster) numbers of overall group actions (not counting gaps), number of gaps, coordination actions, as well as number of gaps

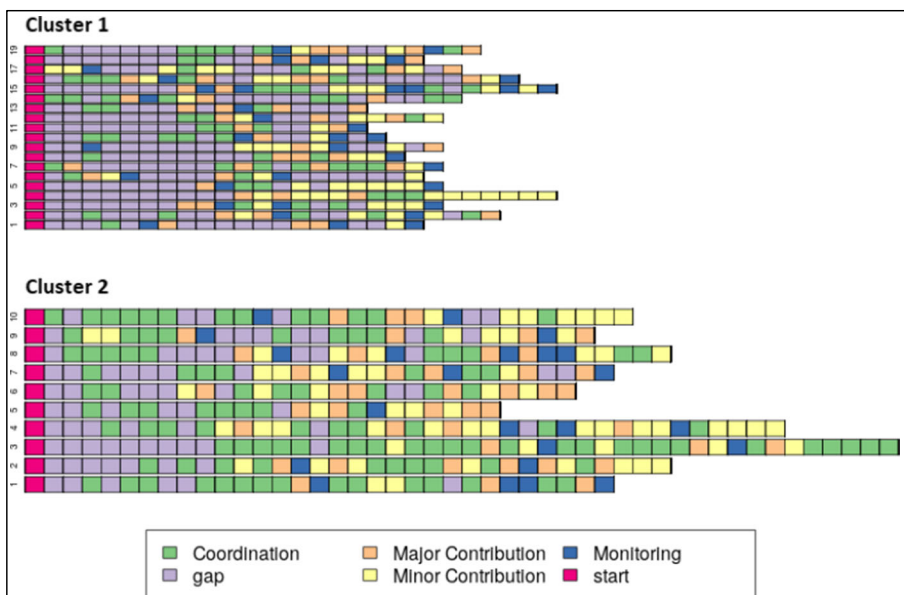


Fig. 6 Clusters after sequence alignment and PAM clustering (k = 2)



**Table 2** Average values per cluster for overall no. of actions (without gaps), gaps, coordinations, gaps in the first half, coordinations in the first half (in parentheses: standard deviation)

Cluster	#Actions overall	#Gaps	#Coordination actions	#Gaps first half	#Coordination actions first half
1	11.7 (3.2)	9.6 (1.2)	3.6 (2.0)	5.8 (1.2)	1.3 (1.6)
2	<b>24.7</b> (5.3)	<b>7.5</b> (1.4)	<b>11.5</b> (5.3)	<b>5.0</b> (1.2)	<b>3.3</b> (2.2)

and coordination actions in the first half (i.e., the first week after the assignment) together with corresponding standard deviations. Due to the doubling of the time window for task completion, the number of gaps is much higher than in the first study. Again, the gaps are more frequent in the first half (about 2/3 fall in the first half). The difference in coordination actions between the two clusters (in favor of cluster 1) is more pronounced than the one in terms of gaps. Regarding the overall number of actions, the value for cluster 2 more than doubles the one for cluster 1. Of course the space occupied by one or more actions in the action sequence diagram is in general not proportional to the duration of the activity. Yet, we know that every gap stands for 24 h of inactivity. There are sequences (esp. in cluster 1) with more than 10 gaps. This implies that in a number of cases the actual online work has been done during a few days.

Table 3 shows the productivity (word count), two variants of satisfaction, together with the work imbalance between group members measured in terms of the Gini index. As can be seen from the standard deviation (in parentheses), values for word count for some of the elaborated texts did not reach the indicated minimum. This was the case for three groups, all from cluster 1. In this study, satisfaction with the group work was measured through a questionnaire administered after completion of the group work. This questionnaire contained two statements related to satisfaction to be rated on a 5-point Likert scale. The first statement just addressed the overall satisfaction with the completed group work, whereas the other addressed the willingness to work again with the same group on future tasks. (This was indeed not an option since groups were always newly composed at random.) Due to the overall small number of groups and the dissimilarity in the length of the sequences, we used a nonparametric method (Mann-Whitney U Test) to test for differences between the two clusters regarding these output parameters. The difference was highly significant for productivity ( $p = 0.0009$ ). The other  $p$ -values were  $p = 0.0057$  for satisfaction 1,  $p = 0.0107$  for satisfaction 2, and  $p = 0.0162$  for the Gini index. This corroborates the assumption

**Table 3** Average values for word count, satisfaction and work imbalance for each cluster (best values in bold face)

Cluster	Word count***	Satisfaction 1**	Satisfaction 2*	Work imbalance (Gini index)*
1	823 (286)	2.97 (0.89)	2.66 (0.94)	0.53 (0.23)
2	<b>1222</b> (239)	<b>3.93</b> (0.55)	<b>3.61</b> (0.87)	<b>0.34</b> (0.15)

Significance levels:  $p \leq 0.001$ \*\*\*,  $0.01$ \*\* ,  $0.05$ \*

that the cluster separation originating from action sequence alignment goes along with an adequate distinction of collaboration quality.

To check the interdependence between output-oriented quality parameters, Spearman rank correlation coefficients were computed: There was a rather strong and almost identical correlation between both variants of satisfaction and the Gini index. This correlation has a negative sign because a higher balancedness corresponds to a smaller Gini index ( $r = -0.662$ ,  $p < 0.001$ , for overall perceived satisfaction and  $r = -0.661$ ,  $p < 0.001$ , for the willingness to work again with the same group). This indicates that the perceived group climate is indeed closely coupled with the actual balancedness of contributions. A weaker correlation was found between the willingness to work together in the future (satisfaction 2) and the word count ( $r = 0.39$ ,  $p = .035$ ). This can be interpreted in such a way as to assume that joint achievement is a good motivation for continued future collaboration.

## Discussion and Outlook

This article provides a synopsis of our experience with applying sequence alignment and clustering techniques to analyze and characterize the well-functioning of small learning groups in larger online courses. The work was conducted in the context of two consecutive cooperation projects with continuity in the reference course from which the data were collected. This paper is focused on the underlying analytics and not, e.g., on strategies of scaffolding feedback or the enabling system architecture.

As explained in section [Using Sequence Analysis to Characterize Group Work](#), sequence analysis (with alignment and clustering) reveals the overall similarity of action sequences. This should not be confounded with sequential pattern mining, in which the frequently re-occurring fixed subsequences are identified (be these predefined or inductively determined). In this sense, sequence analysis is not exactly about a strict structural mapping of partial subsequences but it is about a similar distribution of actions over time. If we completely ignored the sequential characteristics, we would end up with “coding and counting” (cf. Reimann 2009) in a way that would not even consider the overall ordering of the actions in a given sequence. Sequence analysis still depends on and reflects the ordering of the actions. However, once the clusters were generated, we have indeed used “coding and counting” to identify differences between the clusters on the behavioral level. The distinction between action counts in the first half and the overall counts reflects again a limited sequential characteristic. We have also seen that the clusters are not only different in the behavioral level (of actions in process) but also in the output or achievement-oriented quality of the corresponding groups (productivity and balancedness of contributions). The decision tree analysis revealed that a certain minimal degree of coordination, usually concentrated in the first half, is a better indicator for group well-functioning than inactivity, as is commonly assumed.

From a learning analytics perspective, our analysis covers steps 1–3 of the “learning analytics cycle” (Clow 2012) in that data are collected from learners, analyzed and turned into meaningful measures, but the final step of feeding these

back to the learning environment or other stakeholders is still missing. In the learning analytics perspective, the predictive potential of our approach has recently been explored and exemplified by Hecking et al. (2019). In this study, it could be shown that 3–4 days of recorded collaboration actions and activity gaps are a sufficient basis to estimate the work imbalance and productivity at the end of the seven-day group works. Since encoded collaboration actions (see Section [Collection and Sequence Analysis of Course Data \(Study 1\)](#)) are not attributed to a particular group member but to the group as a whole, our approach furthermore shows the potential to be a building block for the early identification of groups at risk ensuring a high level of privacy.

Feedback and scaffolding strategies that would close the learning analytics cycle have been an important goal of the IKARion project (Krämer et al. 2017) in which different types of group scaffolds have been introduced that were, among other principles, also motivated by the dependencies explained in this article. The technical framework and architecture enabling this transition has been reported by Constapel et al. (2019). This architecture contains an analytics backend that analyzes the action logs received from the learning platform (Moodle) to generate a group model. This model represents the current context for each of the learning groups and instantiates certain variables such as inactivity (gaps), imbalance, and productivity (word count). The model also allows for counting only such terms that have been included in a semantic dictionary extracted from the texts of the corresponding learning materials (“semantic word count”). In the IKARion architecture, the group model feeds into a rule-based component that steers the feedback and adaptation/reconfiguration in the Moodle platform. The decision tree that we “learn” from the cluster characteristics gives us indicators for critical situations that we should react to. The sequence analysis is a preliminary step to determine the clusters. Once we know about the specific influence of certain behavioral parameters such as coordination actions and gaps (and the interplay of both) on quality-oriented output parameters, we do no longer need to actually perform a sequence analysis but can use the decision rules as shortcuts. Two consecutive versions of the IKARion architecture have been used in parallel with two versions of the CmC and other smaller courses. The architecture is fully functional and complies with the real-time requirement of the course settings. However, the experiments with this architecture did not exploit the potential of flexible adaptation. The point here was to examine the effects on different types of scaffolds in the Moodle environment that were triggered by clearly controlled and fixed conditions. Nonetheless, we have an operational target environment with adaptation mechanisms based on our analytic findings. First results of this work were presented by Dorian Doberstein (Doberstein et al. 2017) at the conference on Collaboration and Technology (CRIWG 2017) held in August 2017 in Saskatoon, Canada. Jim Greer was present at this event; it was the last time that we could talk to him. There was so much shared understanding backed by the history of intelligent peer help and tutoring that only minimal explanations were needed to reach a productive common ground. Smoothly and directly we could dive into details of our ongoing research and shared interests in the new field of learning analytics. More of this would have been so much appreciated.

**Acknowledgements** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology review and prospect. *Sociological Methods & Research*, 29, 3–33.
- Boroujeni, M. S., & Dillenbourg, P. (2019). Discovery and temporal analysis of MOOC study patterns. *Journal of Learning Analytics*, 6(1), 16–33.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC Press.
- Bull, S., Greer, J., McCalla, G., & Kettel, L. (2001). Help-seeking in an asynchronous help forum. Proceedings of the AIED 2001 workshop on help provision and help seeking in interactive learning environments. San Antonio (USA), may 19-23, 2001.
- Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. In proceedings of the 2nd international conference on learning analytics and knowledge - LAK '12 (pp. 134-138). Vancouver (Canada), April 2012.
- Constapel, M., Doberstein, D., Hoppe, H. U., & Hellbrück, H. (2019). IKARion: Enhancing a learning platform with intelligent feedback to improve team collaboration and interaction in small groups. Proceedings of the 18th international conference on information technology based higher education and training - ITHET 2019 (pp. 1-10). IEEE press.
- Cornwell, B. (2015). Social sequence analysis: Methods and applications (Vol. 37). Cambridge University press.
- Cowell, F. (2011). Measuring inequality. Oxford University Press.
- Curtis, D. D., & Lawson, M. J. (2001). Exploring collaborative online learning. *Journal of Asynchronous Learning Networks*, 5(1), 21–34.
- Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3), 171–176.
- Doberstein, D., Hecking, T., & Hoppe, H. U. (2017). Sequence patterns in small group work within a large online course. In CYTED-RITOS international workshop on groupware - CRIWG 2017 (pp. 104-117). Springer LNCS 10391.
- Doberstein, D., Hecking, T., & Hoppe, H. U. (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.
- Erdmann, J., Rummel, N., Christmann, N., Elson, M., Hecking, T., Herrmann, T., ... & Wichmann, A. (2017). Challenges in implementing small group collaboration in large online courses. Proceedings of the International Conference of Computer Supported Collaborative Learning - CSCL 2017 (pp. 625–628). Philadelphia, USA: International Society of the Learning Sciences (ISLS).
- Ferschke, O., Howley, I., Tomar, G., Yang, D., Liu, Y., & Rosé, C. P. (2015). Fostering discussion across communication media in massive open online courses. In proceedings of the 11th international conference on computer supported collaborative learning - CSCL 2015 (pp. 459-466). Gothenburgh, Sweden: International Society of the Learning Sciences (ISLS).
- Fournier-Viger, P., Gomariz, A., Gueniche, T., Soltani, A., Wu, C. W., & Tseng, V. S. (2014). SPMF: A Java open-source pattern mining library. *The Journal of Machine Learning Research*, 15(1), 3389–3393.
- Gijlers, H., & De Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 42(3), 264–282.

- Greer, J., McCalla, G., Cooke, J., Collins, J., Kumar, V., Bishop, A. and Vassileva, J. (1998) The Intelligent HelpDesk: Supporting Peer Help in a University Course. Proceedings of "Intelligent Tutoring Systems" - ITS '98 (pp. 494–503). San Antonio (USA), August 16–19, 1998. Springer LNCS 1452.
- Grünewald, F., Meinel, C., Totschnig, M., & Willems, C. (2013). Designing MOOCs for the support of multiple learning styles. In European conference on technology enhanced learning - EC-TEL 2013 (pp. 371–382). Springer LNCS 8095.
- Hecking, T., Chounta, I. A., & Hoppe, H. U. (2017). Role modelling in MOOC discussion forums. *Journal of Learning Analytics*, 4(1), 85–116.
- Hecking, T., Doberstein, D., & Hoppe, H. U. (2019). Predicting the well-functioning of learning groups under privacy restrictions. In proceedings of the 9th international conference on Learning Analytics & Knowledge - LAK '19 (pp. 245–249). Tempe (USA), march 4–8, 2019.
- Hoppe, H. U. (1995). The use of multiple student modeling to parameterize group learning. In J. Greer (ed.), Proceedings of the Seventh World Conference on Artificial Intelligence in Education (pp. 234–241). Charlottesville (USA): AACE.
- Inaba, A., Supnithi, T., Ikeda, M., Mizoguchi, R., & Toyoda, J. I. (2000). How can we form effective collaborative learning groups? In international conference on intelligent tutoring systems - ITS 2000 (pp. 282–291). Springer LNCS 1839.
- Kaufman, L., & Rousseeuw, P. J. (2009). Finding groups in data: An introduction to cluster analysis (Vol. 344). John Wiley & Sons.
- Konert, J., Burlak, D., & Steinmetz, R. (2014). The group formation problem: An algorithmic approach to learning group formation. In European conference on technology enhanced learning - EC-TEL 2013 (pp. 221–234). Springer LNCS 8719.
- Krämer, N., Rummel, N., Hoppe, H. U., & Janneck, M. (2017). Intelligente Unterstützung von Kleingruppenarbeit in der online-gestützten Hochschullehre. *E-Teaching.org*: [https://www.e-teaching.org/etresources/pdf/erfahrungsbericht\\_2017\\_kaemer\\_et\\_al\\_intelligente\\_unterstuetzung\\_von\\_kleingruppenarbeit\\_in\\_der\\_online\\_gestuetzten\\_hochschullehre.pdf](https://www.e-teaching.org/etresources/pdf/erfahrungsbericht_2017_kaemer_et_al_intelligente_unterstuetzung_von_kleingruppenarbeit_in_der_online_gestuetzten_hochschullehre.pdf)
- Kyewski, E., Krämer, N. C., Christmann, N., Elson, M., Erdmann, J., Hecking, T., ... & Wichmann, A. (2016). Is small group collaboration beneficial in large scale online courses? An investigation of factors influencing satisfaction and performance in group MOOCs. Proceedings of the 12th International Conference of the Learning Sciences - ICLS 2016 (pp. 918–922). Singapore: International Society of the Learning Sciences (ISLS).
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 63–86.
- Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3), 443–453.
- Piezon, S. L., & Donaldson, R. L. (2005). Online groups and social loafing: Understanding student-group interactions. *Online Journal of Distance Learning Administration*, 8(4).
- Reimann, P. (2009). Time is precious: Variable-and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 239–257.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Staubitz, T., & Meinel, C. (2017). Collaboration and teamwork on a MOOC platform: A toolset. In Proceedings of the 4th ACM Conference on Learning@Scale (pp. 165–168). ACM.
- Themeau, T.M. & Atkinson, E.J. (2017). An Introduction to Recursive Partitioning using the RPART Routines. Retrieved September 26, 2017, from <https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf>
- Tomar, G. S., Sankaranarayanan, S., & Rosé, C. P. (2016). Intelligent conversational agents as facilitators and coordinators for group work in distributed learning environments (MOOCs). In 2016 AAAI spring symposium series. Stanford University, Palo Alto, USA, march 2016.
- Wang, X., Wen, M., & Rose, C. (2017). *Contrasting explicit and implicit support for transactive exchange in team oriented project based learning*. Philadelphia: International Society of the Learning Sciences.
- Webb, N. M., Nemer, K. M., & Zuniga, S. (2002). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. *American Educational Research Journal*, 39(4), 943–989.
- Weinel, M., Bannert, M., Zumbach, J., Hoppe, H. U., & Malzahn, N. (2011). A closer look on social presence as a causing factor in computer-mediated collaboration. *Computers in Human Behavior*, 27(1), 513–521.
- Wegerif, R. (1998). The social dimension of asynchronous learning networks. *Journal of Asynchronous Learning Networks*, 2(1), 34–49.
- Wichmann, A., Hecking, T., Elson, M., Christmann, N., Herrmann, T., & Hoppe, H. U. (2016). Group formation for small-group learning: Are heterogeneous groups more productive? In proceedings of the 12th international symposium on open collaboration (pp. 1–4), ACM.

Wortham, D.W. (1999). Nodal and matrix analyses of communication patterns in small groups. In proceedings of the international conference on computer support for collaborative learning" - CSCL 1999 (pp. 681-687). Stanford University, Palo Alto (USA), Dec. 12-15, 1999. Lawrence Erlbaum associates.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

**H. Ulrich Hoppe<sup>1</sup> · Dorian Doberstein<sup>1</sup> · Tobias Hecking<sup>1</sup>**

✉ H. Ulrich Hoppe  
hoppe@collide.info

<sup>1</sup> COLLIDE Research Group - Department of Computer Science and Applied Cognitive Science, University of Duisburg-Essen, 47057 Duisburg, Germany