Towards the real-time evaluation of collaborative activities: Integration of an automatic rater of collaboration quality in the classroom from the teacher's perspective

2 authors:

Irene-Angelica Chounta
University of Duisburg-Essen
72 PUBLICATIONS   295 CITATIONS

SEE PROFILE

Nikolaos M. Avouris
University of Patras
288 PUBLICATIONS   4,067 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Computational Learning Analytics for monitoring and guiding View project

Project   Mobile games to enrich cultural experience View project

Towards the real time evaluation of collaborative activities: Integration of an automatic rater of collaboration quality in the classroom from the teacher's perspective

Abstract

This paper presents the integration of a real time evaluation method of collaboration quality in a monitoring application that supports teachers in class orchestration. The method is implemented as an automatic rater of collaboration quality and studied in a real time scenario of use. We argue that automatic and semi-automatic methods which provide assessments of collaboration quality or detect disorders and anomalies in collaborative activities should be further used to assist class monitoring. To that end, we study the routine of teachers when they monitor computer-supported collaborative activities with respect to the size of the classroom and the support of an automatic rater. We focus on the way the rater affects their task and how it can be further used to empower the practice of teachers.

1 Introduction

Computer-Supported Collaborative Learning (CSCL) is an interdisciplinary field that studies the practice of learners who work together towards a common goal. Over the past years research has focused on the analysis of collaborative activities. Through analysis, we explore ways to support the practice of people who work together, to scaffold learning and to enhance the overall outcome. The results of the analysis are often used to support teachers in monitoring and guiding students activities (Kahrimanis, Avouris, and Komis, 2011). For that purpose, digital learning systems make use of information on student activity through collections, measurements and reports on data, commonly referred to as "learning analytics". Learning analytics are presented to teachers either in the form of activity statistics or in the form of various visualizations that depict students interaction through monitoring tools (Soller, Martínez, Jermann, and Muehlenbrock, 2005). However, this may entail dangers of misconceptions. On one hand, the measurements used for the statistical analysis might not be appropriate to portray the activity of users. Statistics do not always relate directly to the quality of collaboration and therefore they are not easy to interpret (Spada, Meier, Rummel, and Hauser, 2005). On the other hand, plain demonstration of statistics and visualizations could mislead non familiar users and create confusion or overload to

teachers who monitor a classroom (Dyckhoff, Zielke, Bültmann, Chatti, and Schroeder, 2012). In addition, teachers tend to focus on the cognitive aspects of collaborative activities and neglect the social ones (Van Leeuwen, Janssen, Erkens, and Brekelmans, 2013). However, the learning outcome of collaborative activities heavily relies on the interaction between learners. Thus, the quality of collaboration is an important factor that may affect the activity turnout.

In this article we introduce a method for the real time evaluation of collaborative activities with respect to the quality of collaboration. The method aims to support teachers to monitor and to provide feedback to people working together. For that purpose, the method was integrated as an automatic rater into a class monitoring tool for teachers. The practice of teachers was studied and further analyzed. We argue that the plain demonstration of statistics and graphs is not enough to support the teacher in real time class monitoring. They demand the teacher to keep track of the activity of students and at the same time investigate and interpret information from various resources without providing clear indication of the collaboration quality. We aim to show that the use of a straightforward evaluation indicator of the quality of collaboration would minimize the overhead and support the teacher's practice more effectively. The article concludes with a discussion on the results, the main findings of the study and future work.

2   Related work and research hypothesis

Teachers in CSCL settings have access to a wide range of information regarding the activity of students. The use of computers in learning allows the detailed recording of student activity in log files. Moreover, it gives access to vast resources for further analysis and allows the automation of the process (Dillenbourg, Järvelä, and Fischer, 2009). A wide range of methodologies have been proposed for the analysis of collaborative activity in order to support teachers. At the same time, an open discussion takes place with respect to how this information should be used and presented in order to facilitate the practice of teachers with minimum overhead. In the next paragraphs we provide a summary on related work.

2.1 Learning Analytics as a tool for teacher support

According to the call for papers of the first international Conference on Learning Analytics and Knowledge (LAK 2011), "*learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs*", (as cited by (Ferguson, 2012)). In a collaborative context, learning analytics focuses on capturing "meaningful" interaction, i.e. episodes of student activity that indicate collaborative knowledge building, fruitful communication and successful coordination that leads eventually to learning (Stahl, Koschmann, and Suthers, 2006), (Woo and Reeves, 2007). Digital learning environments use such information to either support teachers in class monitoring (Stahl, 2007) (Weinberger and Fischer, 2006) or to

build students' profiles in order to detect conflicts and bottlenecks (Casamayor, Amandi, and Campo, 2009). This information is presented in the form of statistics or in the form of graphs or other visual representations to the teachers in order to provide assistance and feedback (Suthers, Dwyer, Medina, and Vatrapu, 2010), (Martínez-Monés, Harrer, and Dimitriadis, 2011), (Voyiatzaki, Polyzos, and Avouris, 2008). Dyckhoff et.al (Dyckhoff et al., 2012) propose the use of a Learning Analytics Toolkit (eLAT) in order to efficiently support class monitoring through graphical indicators. The primary goal of this study was to allow teachers to reflect on their own practice and therefore the toolkit was aimed at usability and usefulness rather than meaningfulness in a pedagogical context. Other frameworks focus on monitoring the activity of students and make it available to the orchestrator through dialogue instances, snapshots of the common workspace, graphs and statistics (Voyiatzaki and Avouris, 2014), (De Groot et al., 2007).

Another approach is to provide feedback regarding student activity by comparing it to a pre-existing ideal solution or to other students' practices (Harrer, McLaren, Walker, Bollen, and Sewall, 2006), (Constantino-González and Suthers, 2007). Harrer et al. (Harrer et al., 2006) propose the BND (Bootstrapping Novice Data) approach for the construction of a cognitive tutor. In this work, the data generated from collaborative sessions is used for the creation of a single representation in the form of a graph that updates and evolves as the collaborative activity progresses. In the end of the process, the tutor author marks the faulty actions, add hints and edits the model in a manual way so as to further use it for tutoring. The results showed that multiple levels of abstraction should be further incorporated into the data analysis in order to represent adequately the collaborative practice. On one hand, it is argued that metrics of interaction may provide false assessment of the quality of collaboration (Stahl, 2002). The main disadvantage is that such metrics do not take into consideration the content of user activity or the collaborative context. In addition, activity metrics, charts and statistics can be interpreted in more than one ways leading to misunderstandings and misinterpretations (Spada et al., 2005). On the other hand, more specialized approaches, such as the use of networks and graphs, would require specialized knowledge and cognitive effort on behalf of the teacher. Rodriguez-Triana et. al. (Rodríguez-Triana, Martínez-Monés, Asensio-Pérez, & Dimitriadis, 2014) suggest that the students interaction should be connected and analyzed with respect to the pedagogical decisions of teachers. To achieve that, they propose the combined use of scripting and monitoring, i.e. the activity of students is monitored and compared to a predefined or ideal state suggested by an existing script. Feedback is provided to the teachers regarding inconsistencies and differences between the current and the desired state. Although this approach aims to analyze collaborative activities on both a social and pedagogical context, the overall assessment is trusted upon the teachers, thus leaving room for possible misunderstanding and misinterpretations. Moreover, it requires the existence of a detailed activity planning (scripting) which creates additional overload to the teacher and entails the risk of limiting and suppressing the creative mechanisms of learners triggered by collaborative practices (Dillenbourg, 2002).

Collaborative activities cannot be monitored with traditional means since the unit of analysis rapidly changes between the classroom, the group and the individual (Dillenbourg and Jermann, 2010). In some cases the classroom does not even have a physical status but people form virtual groups and classrooms over the internet. In that case, the teacher cannot have visual inspection and the workload for monitoring, detecting problems and interpreting statistics is difficult to manage. Thus, the information overload could possibly lead to teacher's confusion (Duval, 2011) having the opposite than the desired effects. Therefore, there is an increasing need for tools to further support teachers to monitor and orchestrate computer-supported collaborative activities (C. K. K. Chan, 2011).

2.2 Research hypothesis

Earlier studies focus on the behavior of teachers when they monitor CSCL classrooms and how the use of supporting tools affects their practice (Van Leeuwen, Janssen, Erkens, and Brekelmans, 2014). It was shown that for real time classroom inspection (i.e. when the teacher monitors the classroom during the unfolding of the activity) the teacher goes through a number of stages that require time and effort (Voyiatzaki and Avouris, 2014). In some cases, the teachers even fail to gain understanding of the activity and are not able to provide proper feedback (Wichmann, Giemza, Hoppe, and Krauß, 2009). Collaborative learning takes place when interaction between people who work together triggers learning mechanisms (Dillenbourg, 1999). Therefore, teachers should monitor the quality of collaboration along with the cognitive aspects. However, as the classroom size increases, the teacher gradually focuses on students' cognitive activities neglecting the collaborative aspect (Van Leeuwen et al., 2013).

The main research hypothesis of this study is that the task of monitoring becomes more complex as the size of the classroom increases. This is even more difficult in the case of a virtual classroom, where the teacher does not have visual inspection over the students and the monitoring takes place only through the computer. This indicates the need for specialized support with respect to identifying anomalies in collaboration or to assessing collaboration quality. We argue that the use of automatic or semi-automatic tools to assess collaboration quality or to identify disorder in collaborative practice is necessary in order to support the task of classroom monitoring effectively.

3   Method of the study

3.1 Working hypothesis

In order to explore the research hypothesis we conducted a case study. We implemented an automatic rater of collaboration quality and integrated the rater in a monitoring tool for teachers support. We present our working hypothesis in Figure 1. We describe two classrooms, one of very small size (Case A: four students who form two groups) and one of the double size (Case B: eight students who form four groups). For simplicity, we assume that the teacher only monitors the classroom and shifts focus among the

groups in turn. This means that she monitors Group A, then Group B and then she goes back to Group A again. We also assume that the teacher pays the same amount of time to all the groups. Therefore, in Case A the teacher shares half of the overall time to monitor each group:

$$t_{GroupA} = t_{GroupB} = t_{monitor}/2$$

Case A: 4 students, 2 groups
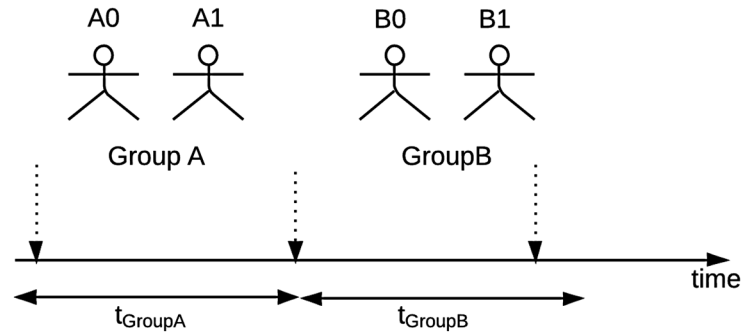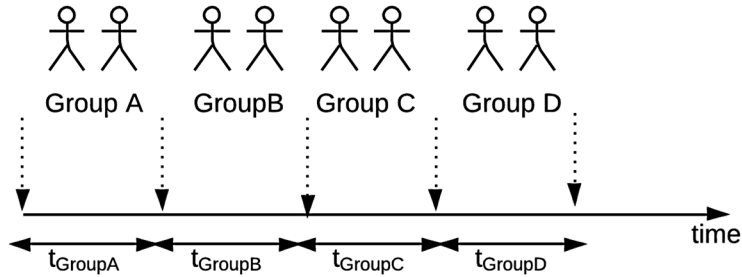

Case B: 8 students, 4 groups

Figure 1. Working hypothesis of the study. In Case A, the teacher monitors a class of four students and in Case B the teacher monitors a class of eight students

By the time the teacher switches views and resumes to follow the same team, she will have to catch up for the same amount of time, $t_{monitor}/2$. In Case B, we have doubled the number of students. The monitoring time is the same ($t_{monitor}$) but since the size of the classroom has doubled, the teacher has less time to dedicate to each group:

$$t_{GroupA} = t_{GroupB} = t_{GroupC} = t_{GroupD} = t_{monitor}/4$$

Keeping in mind that the teacher monitors the groups in the same turn, by the time she resumes to view a group for a second time she will have to catch up for the tripled amount of time than the time she has available ($3*t_{monitor}/4$).

This is an oversimplified scenario that aims at demonstrating the relation between the size of the classroom and the task of monitoring. In a real class, the teacher does

not split her time in equal parts among students or spends her time just by monitoring the activity of groups. On the contrary, the teachers who use class monitoring tools go under phases of activity that require cognitive focus and attention shifting among groups and on different levels or units of analysis, e.g. the classroom, the group, the student. For the purpose of the study, we asked two teachers to monitor two virtual classrooms each through the application "Supervisor". In the end, we analyzed the teachers' practice and their attitude with respect to the automatic rater of collaboration quality.

3.2 Study setup

In the current study, students were grouped in dyads. They were asked to construct collaboratively an algorithmic flow chart to reproduce the Fibonacci sequence. The duration of the activity was approximately 70 minutes. The students had previously attended a lecture on algorithmic flowcharts and they were given a detailed presentation of the algorithm and its general structure. The collaborative activities took place during a course of computer algorithms at the University of Patras. The teachers who participated in the study monitored virtual classrooms simulated from the aforementioned activities with the use of a playback tool. The playback tool reproduced the recorded activities and presented the information to the teacher as it would have happened in real time in a real classroom. This particular setting was used in similar studies and it can provide a realistic setting with minimum resources (Voyiatzaki and Avouris, 2014). The teachers monitored the classrooms through the Supervisor. The scenario was divided into two phases in order to study how the size of the classroom affected the practice of teachers. In the first phase (phase A), the classroom consisted of two dyads and in the second phase (phase B) the classroom consisted of four dyads. The student groups were chosen so as to represent characteristic cases of collaborative practice. All activities were previously evaluated with respect to the solution and collaboration quality. The teachers were not aware of the evaluation results.

Both teachers were familiar with the construction of algorithmic flowcharts and they had teaching experience on the particular subject. They were provided with the exercise and an ideal solution. One teacher (Teacher A) was familiar with the Supervisor application. She was also familiar with the rating scheme that was used to evaluate collaboration quality. However, she was introduced to the automatic rater component for the first time. The other teacher, (Teacher B) had no prior experience with the Supervisor. Therefore, she went through a training phase in order to become familiar with the application and monitoring process. She was also introduced to the rating scheme, the automatic rater and its results. The practice of the teachers was recorded by an eye tracking device and a video camera. They also used a think aloud protocol in order to express themselves and communicate their practice. After the completion of the study, they were interviewed with respect to their experience. In the following paragraphs we present a qualitative analysis of the activity of teachers. We focus on particular episodes of their activity that provide insight to their needs and how to further support their practice.

4 Real time classification of collaborative activities using time series

In order to provide real time evaluation of collaboration quality (automatic rater), we used a time series classification model of collaborative activities (TSCMoCA). The model was used to classify the activities with respect to the similarity of their time series. The classification schema is based on the hypothesis that collaborative activities of similar quality will be represented by time series of similar structure. The research hypothesis and the effectiveness of the classification schema were explored in a number of previous studies where time series classification was used for the post-evaluation of collaborative activities (Chounta and Avouris, 2012). A number of parameters was also explored, such as the sampling rate and type of activity used for the construction of the time series etc. To evaluate the performance of the model, the results were compared to the ratings of human experts. It was shown that the performance of the model increases for sampling rates of less than 60 seconds and for events that represented basic user activity, such as the sum of chat messages and workspace actions, role switches, etc. This finding confirms earlier studies, where it was shown that for synchronous, collaborative activities the "meaningful" interactions take place within the time frame of about 25 seconds (Schümmer, Strijbos, and Berkel, 2005).

The TSCMoCA model was also used for the real time evaluation of collaborative activities (Chounta and Avouris, 2014). The purpose of this study was twofold: (a) to confirm that the model can be used for the real time evaluation of a collaborative activity and (b) to define the volume of data needed for the successful evaluation of an activity, i.e. how long the time series of user activity should be for the effective function of the model. The results showed that the model was able to perform effectively when used after the first half of the activity. In particular, for an activity of about 90 minutes, the model was able to evaluate adequately the quality of collaboration 40 minutes after the beginning of the activity. The ratings of the model correlated significantly with the ratings of human experts ($\rho > 0.5$) and the prediction error was estimated to be less than one on a 5 point scale.

In Figure 2 we present a common scenario that we aim to support. According to the scenario, the teacher monitors a collaborative activity that takes place among student A and student B in real time. The activity begins on t_o and on t_i the teacher requests an evaluation of the activity. The method (TSCMoCA) is invoked by the teacher and it provides an evaluative value for the quality of collaboration ($CQA[t_i,t_o]$, where CQA: Collaboration Quality Average). The teacher can let the activity unfold or provide additional feedback to the group on a later time (t_F) depending on the evaluation result. The method uses a time series of collaborative activities as input and classifies them using a reference set of pre-evaluated ones. The reference set consists of collaborative activities previously evaluated, with respect to collaboration quality, by human raters with the use of a rating scheme (Kahrimanis et al., 2009), (Kahrimanis, Chounta, and Avouris, 2012). The time series of the collaborative activity ($ts_{CA[t_i,t_o]}$) is compared to the entries of the reference set and further classified with respect to their similarity. The

model TSCMoCA implements a K-Nearest Neighbor (KNN) algorithm. The similarity of time series was computed by the DTW (Dynamic Time Warping) algorithm (Giorgino, 2009).
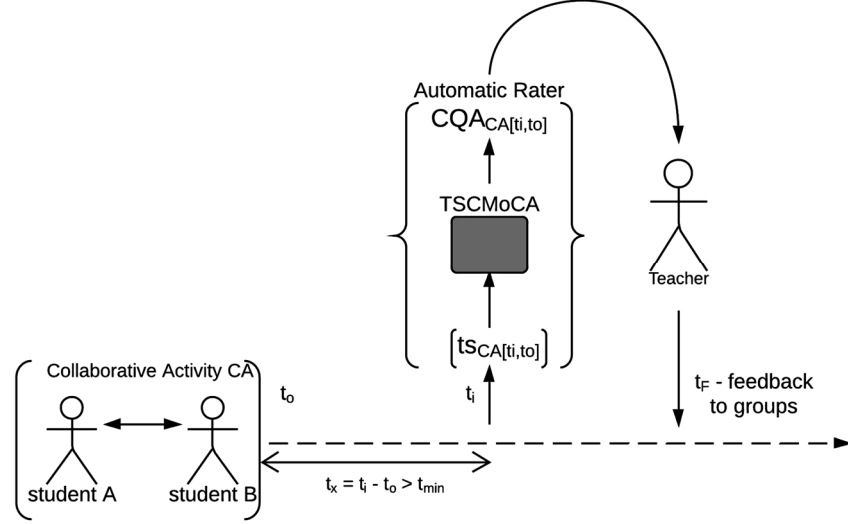

Figure 2. Real time classification scenario of an ongoing collaborative activity CA

The simple structure of the model and the low computational cost make the proposed classification schema an efficient method for the real time evaluation of collaborative activities. The method does not require the completion of the activity but performs effectively even if applied on the first half of the activity. Performance is improved as the activity unfolds, therefore in order to invoke the method there is a minimum time limit from the beginning of the activity ($t_x = t_i - t_o > t_{min}$). This time limit was experimentally set to 25 minutes for an activity of 90 minutes (Chounta and Avouris, 2014).

The method does not require additional resources or coding. It uses the basic user events (such as import/ modify and delete objects, chat messages etc.) that are recorded by the majority of groupware applications. The only pre-requisite is the existence of a reference set of pre-evaluated activities of the same kind. This is a typical characteristic of memory-based learning models and it allows cost-efficient, fast classification since they do not require a training phase.

4.1 The construction of the reference set

In the present study, the reference set is constructed from 210 collaborative activities that took place during a course of Computer Science in the University of Patras, Greece. The students of the course were grouped in dyads. They worked together in order to create typical algorithmic flowcharts. The duration of the activity ranged from 70 to 90 minutes and the students collaborated synchronously. Each dyad was supported by the groupware application Synergo (Avouris, Margaritis, and Komis, 2004). The students had no face to face contact during the activity.
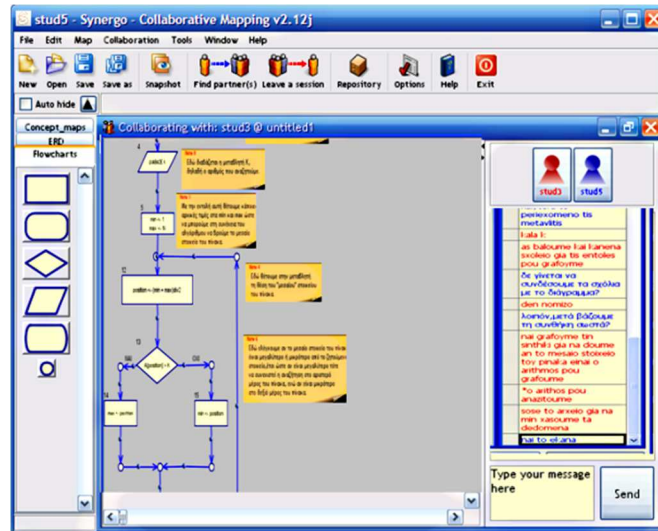
Figure 3. The user interface of the collaborative application Synergo

The groupware application is often used in similar courses for the creation of diagrammatic representations such as entity-relationship diagrams, flowcharts and mind maps. The application consists of two shared spaces: a common workspace for the creation of diagrams and a chat tool that supports communication between users via instant messaging (Figure 3).

4.1.1 Evaluation of the quality of collaboration

In order to evaluate the quality of collaboration, we employed a rating scheme that defines collaboration on four fundamental aspects: Communication, Joint information processing, Coordination and Interpersonal relationship (Kahrimanis et al., 2009). The four fundamental aspects are further analyzed into six collaborative dimensions. These dimensions are:

- Collaboration flow (CF),
- Sustaining mutual understanding (SMU),
- Knowledge exchange (KE),
- Argumentation (Arg),
- Structuring the problem solving process (SPSP),
- Cooperative orientation (CO)

The quality of collaboration (CQA) is computed as the average of the evaluations on the six collaborative dimensions:

$$CQA = average(CF, SMU, KE, Arg, SPSP, CO)$$

In earlier studies, the rating scheme was used to assess the quality of collaboration of joint activities (Kahrimanis, Chounta, & Avouris, 2011), (Kahrimanis et al., 2012).

In particular, two expert evaluators assessed 228 collaborative activities with the use of the rating scheme. Each one of the collaborative dimensions (Collaboration flow, Sustaining mutual understanding, Knowledge exchange, Argumentation, Structuring the problem solving process and Cooperative orientation) was evaluated on a 5-point Likert scale ([-2, +2]). The general dimension of the quality of collaboration (CQA) was computed as the average of the ratings on the six dimensions. The results proved that the rating scheme is a useful means for assessing the overall quality of collaboration. The ratings of the human raters were tested for inter-rater reliability and absolute agreement and the scores for both the ICC and the Cronbach's alpha were satisfying (ICC>0.83, Cronbach's a>0.91). In the current study we used the rating scheme to assess the quality of collaboration of the activities that construct the reference set.

4.1.2   Design and Implementation of the Automatic Rater

The classification method TSCMoCA was designed so as to make the real time evaluation of a group feasible at any time throughout the duration of the activity. In order to study its use in a working paradigm, we integrated the model into a monitoring application (Supervisor) as an automatic rater of collaboration quality. Supervisor (Voyiatzaki et al., 2008) is implemented in Java as part of the groupware suite Synergo (Avouris et al., 2004).

The Synergo suite consists of three basic components:

- **The Synergo Client:** The client supports users who work collaboratively on the creation of diagrammatic representations. It was used in the particular study as the groupware that mediated the collaborative activities.

- **The Relay Server:** the server is responsible for the connection and data exchange between the clients as well as the Supervisor station.

- **The Supervisor:** a tool that supports class monitoring and orchestration. It provides teachers with an overview of the classroom and working groups.

The architecture of the Synergo suite is portrayed in Figure 4. The classification model TSCMoCA was implemented as an R component and integrated in the Supervisor through a Java Interface.
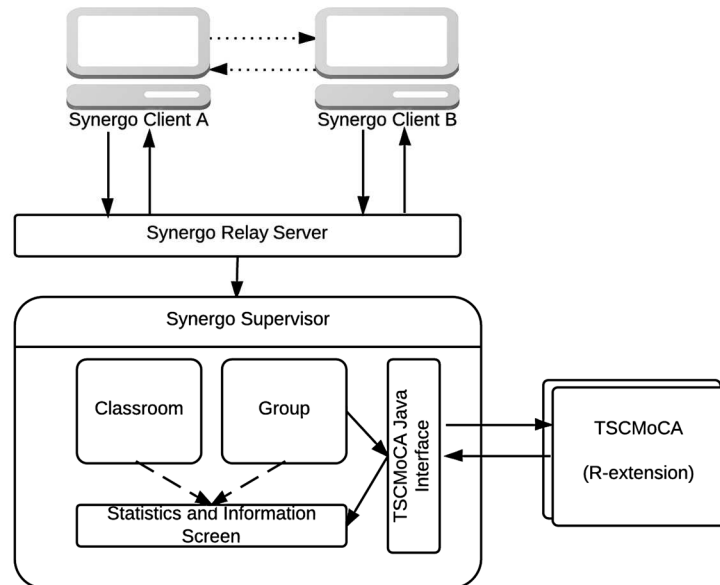
Figure 4. The architecture of the groupware application suite Synergo and the TSCMoCA integration as a Java/R extension plugin to support class orchestration

4.1.3   TSCMoCA User Interface

The TSCMoCA User Interface is visualized as a custom tab that allows the teacher to "execute" a real time automatic evaluation of the collaboration quality of one particular group of students. The automatic rater is called on demand and through the Java Interface. The interface also connects the Supervisor with the TSCMoCA component (Figure 4). The automatic rater uses data already acquired from the Relay Server. Therefore, no additional overhead is caused. The user interface is presented in Figure 5.
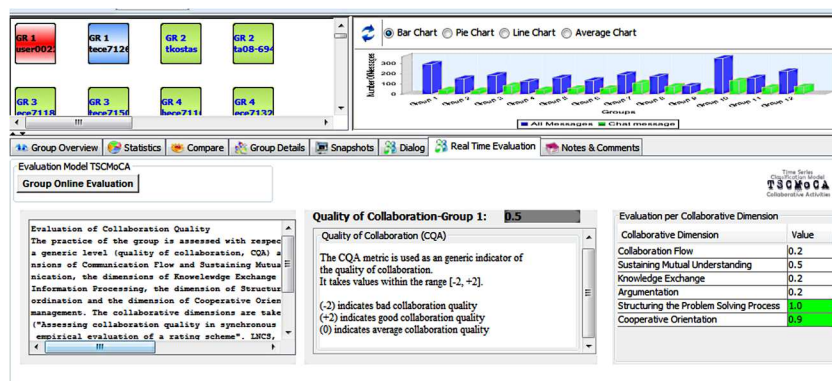

Figure 5. The User Interface of the TSCMoCA model as integrated in the Supervisor

4.1.4   TSCMoCA Core

The core of the automatic rater (TSCMoCA Core) was implemented in R. R (Team R, 2012) is a software programming language for statistical, efficient computing that offers a wide collection of algorithms and methods for statistical and mathematic analysis, such as support for time series construction (K. S. Chan, 2010) and the DTW algorithm (Giorgino, 2009). Both are used in the automatic rater core component.

The TSCMoCA core returns to the User Interface a set of evaluative values on collaboration quality for the selected group. The model classifies and evaluates the collaborative activity with respect to the reference set of pre-evaluated activities. Thus, the result of the evaluation depends on the type of evaluation that was used for the reference set. In this particular study, the reference set consists of collaborative activities assessed with the use of a rating scheme, as aforementioned. Therefore, the model evaluates the collaborative activity on the dimensions and scale set by the reference set. In particular, the evaluation ratings range in between [-2, +2], where (-2) stands for poor quality while (+2) represents good quality. The choice of the reference set is done by the teacher with respect to the nature of the activity and it does not affect the functionality of the model.

5   Class supervision assisted by an automatic rater of collaboration quality

In the current section we present the analysis of the practice of teachers when monitoring a classroom through the application Supervisor. In similar studies, it was shown that the teachers go through four distinct stages of behavior when they inspect students through monitoring tools (Voyiatzaki and Avouris, 2014). These stages are:

- Steady State Monitoring (SSM): the teacher monitors the activity of students
- Disorder Investigation (DI): the teacher diagnoses a disorder
- Direct Communication (DC): the teacher communicates directly with the student
- Feedback (F): the teacher provides feedback to students

We use these four stages to identify and classify the attention and behavioral shifts of the teacher during monitoring. We argue that the increase of the classroom size will cause bigger information flow than the teacher can process. In this case, the teacher will turn to the automatic rater for additional support.

The study is divided into two cases, according to the working hypothesis. In Case A, the teachers monitor two groups and in Case B, the teachers monitor four groups.

5.1 Case A – Monitoring two dyads

In Case A, the teachers are asked to monitor the activity of four students grouped in two dyads. The two teams were chosen so as to represent a case of good (group 1) and bad (group 2) collaboration quality. We analyze and compare the practice of the

teachers. The timeline of their activity is presented in Figure 6 with screenshots of the teachers' activities provided from the eye tracking device.

5.1.1   Teacher A

Teacher A is a teacher of Computing & Informatics with prior experience in the field. Moreover, she is familiar with the monitoring application. During the first minutes of the activity, the teacher focused on the dialogs between the students and the graphs of students' activity. She was trying to gain some understanding of the relation between the students and the way they communicated (AA.01 – Steady State Monitoring, see Figure 6). She remained focused on the dialogue although she was occasionally looking at the common workspaces of the teams in order to evaluate the progress of the flowchart. She had already formed a clear picture of both teams with respect to their collaboration quality after the first 30 minutes of the activity. In particular, she stated that the first group was good and she was not worried about their progress anymore (AA.02 – Steady State Monitoring). She was mostly concerned about the second group that appeared to have difficulties in communicating. Her attention was shifted to the common workspace of the second team and she focused on the flowchart diagram they were constructing. Almost in the middle of the activity, the teacher focused on the solution quality rather than the collaboration of users (AA.03 - Disorder Investigation). At that point the teacher was in steady monitor stage (SSM) for Group A that she perceived as good and in the disorder investigation stage (DI) for Group B that she perceived as slightly problematic. However she managed to resume to Steady State Monitoring since she had plenty of time to browse through the dialogues and the workspaces of the groups. The teacher used the automatic rater 47 minutes (00:47) after the beginning of the activity, in order to confirm her own opinion of collaboration quality for each team (AA.04). Afterwards she went back to the dialogue to confirm the results of the evaluation. As the activity progressed, the teacher focused again on the common workspaces of the teams, mostly concerned about the correctness of the solution. She used the dialogue of the students to explain the activity in the common workspace and did not use the automatic rater again but only out of curiosity (AA.05)

5.1.1   Teacher B

The second teacher is a teacher of Computing & Informatics who had worked on collaborative activities but she had no prior experience in the monitoring tool. Before the activity, she was given a detailed presentation of the Supervisor functionality and operation. Moreover, she went through a training phase. In the beginning of the activity the teacher browsed the application and reads the dialogue between team members in order to get familiar with the setting. At 00:14 (BA.01) she explained that it was too early to form some idea about the groups and she would like to spend more time browsing between the dialogue and the common workspace in order to understand what is really happening. The teacher mainly switched between the two teams to compare their progress and to monitor the solution quality (BA.02 - Steady Monitoring State).

.

Teacher A - Case A
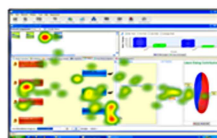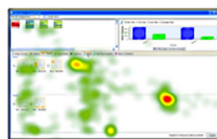
AA.00 (00:00): Activity Start

AA.01 (00:13): Teacher A browsed the dialogues of the groups (SSM)

AA.02 (00:30): Teacher A had a clear picture of all groups. She was able to browse the student activity

AA.03: Teacher A followed the group activity without additional help (SSM)

AA.04 (00:47): The teacher used the automatic rater to validate

AA.05 (01:02): Teacher A returned to the students' dialogues to

Time (HH:MM)
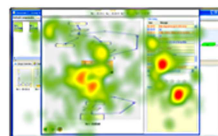
Teacher B - Case

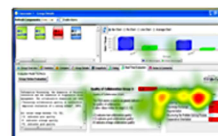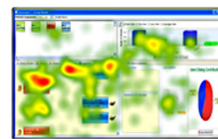**BA.00 (00:00):** Activity Start

BA.01 (00:14): Teacher B browsed the application

BA.02 (00:31): Teacher B browsed the dialogues and compared the progress of

BA.03 (00:41): Teacher B followed the activity and formed an opinion on

BA.04 (00:49): Teacher B used the automatic rater from curiosity

BA.05 (00:57): Teacher B had a clear picture of the classroom and did not need
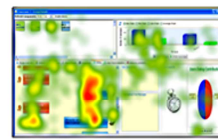
Time (HH:MM)

Figure 6. The timeline of teachers' practice as they monitor the activity of four groups through the Supervisor during Case A

She found it easy to follow the practice of both teams and she formed an opinion with respect to the teams' performance on an early stage (BA.03 - Steady Monitoring State) being confident that she had a clear picture of the class. In the middle of the activity (BA.04) the teacher used the automatic rater "*out of curiosity*", as she noted. Teacher A stated that it confirmed her original opinion but she would not worry if this was not the case. Instead, she would return to have a closer look at the dialogues. Near the ending of the activity (BA.05 - Steady Monitoring State), the teacher stated that she had a clear picture of the classroom, thus she was not discouraged from not having a face-to-face communication with the students. She was able to evaluate the quality of collaboration based on the information she could retrieve from the Supervisor and she did not need an extra indicator of collaboration quality.

5.2 Case B.

In Case B, the classroom consisted of four groups which were monitored by the teachers through the Supervisor. The groups were chosen so as to represent cases of good (group 1 and group 2), average (group 3) and bad (group 4) collaboration quality. The teachers were not aware of the evaluation ratings of each team. The timeline of the activity of teachers throughout the study is presented in Figure 7.

5.2.1   Teacher A

In the second case, teacher A spent the first minutes of the activity monitoring the dialogue of each group in turn. She stated that it was important for her to maintain a steady turn and order to group monitoring (a "*virtual trace*", as she called it) in order to be able to remember effectively the practice of all groups. The teacher was engaged in the dialogue and was not concerned about the collaboration quality since it was "*still too early*" (AB.01 – Steady State Monitoring). As the activity progressed, she became concerned about a team which appeared to have extremely low activity. However she stated she could form an overall judgment for all teams based on the activity metrics and the dialogue. After 24 minutes of the activity (AB.02 – Disorder Investigation), the attention of the teacher shifted to the group she had identified as "problematic" and was studying in detail the common workspace and the dialogues of that particular team. Thirty four minutes after the beginning of the activity (AB.03 – Steady State Monitoring), she stated she was having trouble remembering which team did what because, as she commented, she did not have a view of an actual classroom so as to associate who is who. Moreover she mentioned that the information flow was increasing rapidly and therefore she had no time to go through the dialogues and workspaces of every group, as she would like to do. In order to gain an overall assessment, teacher A used the automatic rater for each group in turn (AB.04). In the case that the results of the evaluation confirmed her opinion, she considered them "*accepted*" and did not explore the findings any further. In the opposite case, she returned to the dialogue of each team to re-assess their practice.

Teacher A - Case B

Teacher B - Case B

AB.00 (00:00):
Activity Start

BB.00 (00:00):
Activity Start

AB.01 (00:12):
Teacher A
monitored each
group in turn
(SSM)

BB.01 (00:11):
Teacher B
monitored the
activity of groups
(SSM)

AB.02 (00:24):
Teacher A was
concerned
about a team
with low
activity (DI)

BB.02 (00:21):
Teacher B
spotted a
problematic team
and followed
them closer

AB.03 (00:34):
Teacher A
experienced
difficulties in
following every
team (SSM, DI)

BB.03 (00:38):
Teacher B
concentrated on
one team and had
difficulties to
follow the rest
(DI)

AB.04 (00:39):
Teacher A used
the automatic
rater to gain an
overall picture of
the class

AB.05 (00:56):
Teacher A
returned to the
students' activity
to confirm the
evaluation result
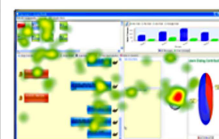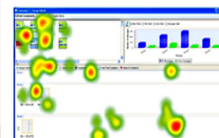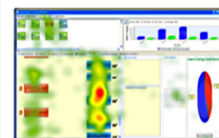

BB.04 (00:41):
Teacher B used
the automatic
rater per team so
as to catch up

Time
(HH:MM)

Time
(HH:MM)

Figure 7. The timeline of teachers' practice as they monitor the activity of four groups through
the Supervisor during Case B

5.2.1   Teacher B

In the second case, the teacher started by browsing the activities of each one of the four groups. She followed a standard path to keep track of the groups. Although the activity of the common workspace was still low, the teacher noted that the dialogue was getting bigger and it was not easy for her to follow (BB.01 –Steady State Monitoring).

Gradually, the teacher quit from following the groups in turn as she spotted some problematic dialogue in certain teams and wanted to follow them closer (BB.02). The teacher was following a steady state monitoring stage for some teams and a disorder investigation stage for others. Her attention shifted on a particular team that she believed it needed support. However, she was not certain whether this was indeed the case or she was missing important information. Therefore, she chose to wait so as not to disturb them. Teacher B was fully concentrated on that team until 00:38 (BB.03) and she was not following the progress of other teams. When she tried to resume, she found it hard to remember the state of each group before she got involved deeper with that one team. The teacher commented that she needed help and used the automatic rater for the teams she was not following (BB.04). After the first half of the activity, the teacher reviewed the automatic evaluation results each time she switched to a different group.

5.3 General Comments on the Activity

After the ending of both cases, the teachers were interviewed with respect to their experience. Teacher A, who had prior experience with orchestrating collaborative activities mediated by computers, stated that she did not trust automatic evaluation. Thus she did not use it regularly when the size of the classroom was small and she could monitor each team without much effort (Case A). During the first case, teacher A used the automatic rater in order to confirm her own assessment. However, she commented that the automatic rater caused her overhead. She had to confirm the results of the rater regardless whether her initial opinion agreed or not. In Case B, the teacher was unable to fully follow the practice of each group. She used the automatic rater as a "guide" or "indicator" and it helped her make quick judgments about the collaboration quality and focus on the correctness of the solution. Therefore, she used the rater before focusing on a particular team so as to have an overall picture. Teacher A stated that although she would still not trust the rater fully, it was useful as an "indicator". It helped her confirm her own perspective and at the same time she was relaxed that in the case of a mistaken, it might have triggered some alarm so that she could re-evaluate. Although she would still not fully trust an automatic rater overall, she would use it as an indication and not as an absolute judgment.

Teacher B was inexperienced in class monitoring through computers and therefore she completed a training phase before using the application. In general, she was cautious with the use of the automatic rater. Teacher B only used it out of curiosity during Case A. She was able to monitor all groups and have a clear picture of their activity. In Case B, the teacher found it hard to keep track of the concurrent activities of all four teams. After the first half of the activity she claimed it was difficult to remember the practice of each team and used the automatic rater as a validation of her

opinion. Although she did not trust automatic raters in general, she commented that she found it useful when the activity was intense and she would use it to a certain extent.

6   Discussion

In this section we review the effect of automatic rating of collaboration quality on the behavior of teachers who monitor computer-supported collaborative activities. The need for monitoring tools and support systems for teachers is evident especially in the case of large or virtual classrooms. In a real classroom the teacher is usually on the move so as to have constant awareness of the activity of students. However, this is a challenging task when the size of the classroom increases or when the teacher does not have visual inspection of it. It was shown in similar studies that the teachers who use monitoring tools to support class orchestration follow a certain practice consisting of four recursive stages: steady state monitoring, disorder investigation, direct communication and feedback (Voyiatzaki and Avouris, 2014). We used the four stages of teacher activity to group the actions of teachers as recorded in the current study. In both cases (Case A and Case B), the teachers followed a similar course of actions as they monitored the given classroom. Their activity was portrayed as a timeline and the similarity of their practice was evident. Teacher A and teacher B went through phases of Steady State Monitoring and Disorder Investigation. They were able to monitor all groups effectively and in time. Therefore, they did not need extra support. Due to the small size of the classroom, they were able to track the source of a disorder efficiently and to provide the required feedback. They only used the automatic rater as a means of validation close to the end of the activity. Since they were more confident about their own realizations, they perceived it rather as "additional overhead" rather than help. In Case B, we doubled the size of the classroom. The teachers went through the phases of Steady State Monitoring and Disorder Investigation. In this case however, they found it difficult to keep track of all groups since the amount of information doubled as well. When a teacher tracked some disorder, she went over the group's activity in order to track its source. By the time she resumed to monitor the rest of the classroom, the information available and the progress of the rest of the groups had increased so much that the teacher was in need of extra support. In Case B, the automatic rater was used sooner than in Case A. The teachers used the rater in order to get an indication of each group's collaboration quality. In the case it confirmed their initial opinion, they shifted their attention to the solution quality or to what kind of feedback they should provide. In the opposite case, the teachers perceived the automatic rater as an alarm so as to re-evaluate the practice of the team in question.

The timeline of the teachers' activity reveal that they both follow similar practices with respect to the stages they go through, the order and the rhythm of their activity. Despite the fact that one teacher is experienced and has good knowledge of the monitoring application while the other is newly introduced to it, they both face the same difficulties and their attention is shifted in similar ways.

The steady state monitoring and the disorder investigation stages are the ones that are mostly affected by the size of the classroom increases. In the steady state monitoring stage, teachers have to go through the group activities faster and to remember more information. When they become aware of some anomaly or disorder, they focus on the particular group and even deeper, on the individual student. This prevents them from following the rest of the class and creates additional workload. In Case A the teachers were able to track disorders and return to state monitoring without any particular issue. In Case B, the teachers entered the disorder investigation stage sooner than Case A and it was harder for them to resume. This could be an indication that in the second case teachers were unable to process all the available information. Therefore, they investigated, as disorders, phenomena that otherwise they would consider normal. We should also note that when the size of the classroom increases the stages of steady state monitoring and disorder investigation will inevitably overlap for different groups, this leading to additional overhead for the teacher.

The automatic rater aims to support the teacher to overcome this overhead. The study showed that the teachers did not use the automatic rater when they were confident of having a clear picture of every group in the classroom. They were even negatively disposed towards it and perceived it as nuisance as they were mentally compelled to re-evaluated the automatic evaluation results from their perspective. The teachers stated before the study that they did not trust automatic raters in general and they would only feel confident after an extensive personal experience of use. One teacher in particular questioned the use of automatic raters because "*it is impossible to interpret adequately the content of the dialogues or the colloquial terminology and figures of expression that students use when they chat*". However, their practice revealed that they overcame the initial cautiousness as the complexity of monitoring increased. Even in the latter case, the automatic rater was mostly perceived as an indicator rather than an absolute rating of collaboration quality. The teachers used it to validate their original opinion or to gain a rough estimation of the quality of collaboration before they got engaged with a particular team on a deeper level. They accepted the results of the automatic evaluation when they mirrored their own opinion and they were alarmed only when the results were negative with respect to their evaluation.

The overall metric of the quality of collaboration (CQA) was perceived by the teachers as most useful. The collaborative dimensions were characterized as "details" that might be used if the evaluation on a deeper level was required. However, the teachers commented that in a real-classroom they are not interested in such a deep level of analysis. The teachers were satisfied that they could invoke the automatic rater on demand, rather than have it as a constant, global indicator. They would have to check constantly the results since they did not trust it fully, and that would cause them additional overhead. On the contrary, they would like to have the possibility to "save" the evaluation results so as to be able to view them later. That would help them maintain the "history" of a certain group for post evaluation.

7 Conclusions and future work

In this article we present the implementation of a method for real time evaluation of collaborative activities and its integration into Supervisor, a monitoring tool that supports teachers to monitor computer-supported collaborative activities. The method was integrated as an automatic rater and it could provide the teachers with real time on demand evaluations of collaboration quality. The research agenda was divided into two parts. In the first part, we presented the real time classification and evaluation method and its implementation as an on demand automatic rater. The results of the method were compared with ratings of human evaluators. In the second part, two teachers were asked to monitor collaborative activities with the use of Supervisor and assisted by the automatic rater. The aim of the study was to record the practice of teachers with respect to the automatic rater and the size of the classroom.

It was shown that the teachers in general were reluctant to use the automatic rater when the size of the classroom was small and they could keep track of students' activity. They mostly browsed the dialogues of groups because they could extract valuable information out of the content of written messages. The dialogue was also their main tool for forming an opinion about collaboration quality. The automatic rater was perceived as "extra workload" because they felt that they had to confirm these ratings. Therefore, it was rarely used and not fully trusted. When the size of the classroom increased, the teachers found it difficult to follow the classroom with no additional help. They used the automatic rater quite often in order to gain an overview of collaboration quality or to validate their own opinion. They stated that they trusted it as an "indication" of the quality of collaboration and it helped them make faster judgments on the progress of certain teams. They were also confident that the use of the automated rater could prevent them from making wrong decisions.

The use of a real time evaluation method of collaboration quality aims to support teachers to empower their practice and does not intend to substitute them. The emergence of new technologies and their integration into teaching practices is a promising research field. However, the teachers face additional overhead as they have to adapt to new learning settings (virtual classrooms, distance learning) and adopt new tools (mobile devices, tablets, specific-purpose applications etc). The reported study involved activities of a specific collaborative setting. In future work we plan to generalize the findings by investigating collaborative activities of various characteristics, i.e. asynchronous communication, various tasks, different team sizes etc.

References

Avouris, N., Margaritis, M., & Komis, V. (2004). Modelling interaction during small-group synchronous problem-solving activities: The Synergo approach. In *Proceedings of ITS 2004 workshop on designing computational models of collaborative learning interaction* (pp. 13–18).

Casamayor, A., Amandi, A., & Campo, M. (2009). Intelligent assistance for teachers in collaborative e-learning environments. *Computers & Education*, *53*(4), 1147–1154.

Chan, C. K. K. (2011). Bridging research and practice: Implementing and sustaining knowledge building in Hong Kong classrooms. *International Journal of Computer-Supported Collaborative Learning*, *6*(2), 147–186.

Chan, K. S. (2010). *TSA: time series analysis. R package version 0.98.* URL: http://CRAN. R-project.org/package= TSA.

Chounta, I.-A., & Avouris, N. (2012). Time Series Analysis of Collaborative Activities. In V. Herskovic, H. U. Hoppe, M. Jansen, & J. Ziegler (Eds.), *Collaboration and Technology* (pp. 145–152). Springer Berlin Heidelberg.

Chounta, I.-A., & Avouris, N. (2014). It's all about time: towards the real-time evaluation of collaborative activities. In *14th International Conference on Advanced Learning Technologies, ICALT2014* (pp. 283–285). Athens, Greece: IEEE.

Constantino-González, M. de los Á., & Suthers, D. D. (2007). An approach for coaching collaboration based on difference recognition and participation tracking. In H. U. Hoppe, H. Ogata, & A. Soller (Eds.), *The Role of Technology in CSCL* (pp. 87–113). Springer US.

De Groot, R., Drachman, R., Hever, R., Schwarz, B. B., Hoppe, U., Harrer, A., Baurens, B. (2007). Computer supported moderation of e-discussions: the ARGUNAUT approach. In *Proceedings of the 8th iternational conference on Computer supported collaborative learning* (pp. 168–170). International Society of the Learning Sciences.

Dillenbourg, P. (1999). What do you mean by collaborative learning? *Collaborative-Learning: Cognitive and Computational Approaches*, 1–19.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL?* (pp. 61–91). Heerlen, Open Universiteit Nederland.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In *Technology-enhanced learning* (pp. 3–19). Springer.

Dillenbourg, P., & Jermann, P. (2010). Technology for Classroom Orchestration. In M. S. Khine & I. M. Saleh (Eds.), *New Science of Learning* (pp. 525–552). Springer New York.

Duval, E. (2011). Attention Please!: Learning Analytics for Visualization and Recommendation. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 9–17). New York, NY, USA: ACM.

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and Implementation of a Learning Analytics Toolkit for Teachers. *Educational Technology & Society*, *15*(3), 58–76.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5), 304–317.

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: the dtw package. *Journal of Statistical Software*, *31*(7), 1–24.

Harrer, A., McLaren, B. M., Walker, E., Bollen, L., & Sewall, J. (2006). Creating cognitive tutors for collaborative learning: steps toward realization. *User Modeling and User-Adapted Interaction*, *16*(3-4), 175–209.

Kahrimanis, G., Avouris, N., & Komis, V. (2011). Interaction Analysis as a Tool for Supporting Collaboration: An Overview. In T. Daradoumis, S. Caballé, A. A. Juan, & F. Xhafa (Eds.), *Technology-Enhanced Systems and Tools for Collaborative Learning Scaffolding* (pp. 93–114). Springer Berlin Heidelberg.

Kahrimanis, G., Chounta, I.-A., & Avouris, N. (2011). Rating dimensions of collaboration quality in synchronous collaborating dyads: findings and interpretations. In H. Spada, G. Stahl, N. Miyake, & N. Law (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings* (Vol. 1, pp. 446–453). Hong Kong, China.

Kahrimanis, G., Chounta, I.-A., & Avouris, N. (2012). Validating empirically a rating approach for quantifying the quality of collaboration. In *Intelligent adaptation and personalization techniques in computer-supported collaborative learning* (pp. 295–310). Springer.

Kahrimanis, G., Meier, A., Chounta, I.-A., Voyiatzaki, E., Spada, H., Rummel, N., & Avouris, N. (2009). Assessing collaboration quality in synchronous CSCL problem-solving activities: Adaptation and empirical evaluation of a rating scheme. In *Learning in the Synergy of Multiple Disciplines* (pp. 267–272). Springer.

Martínez-Monés, A., Harrer, A., & Dimitriadis, Y. (2011). An Interaction-Aware Design Process for the Integration of Interaction Analysis into Mainstream CSCL Practices. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 269–291). Springer US.

Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2014). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology*.

Schümmer, T., Strijbos, J.-W., & Berkel, T. (2005). A new direction for log file analysis in CSCL: Experiences with a spatio-temporal metric. In *Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!* (pp. 567–576). International Society of the Learning Sciences.

Soller, A., Martínez, A., Jermann, P., & Muehlenbrock, M. (2005). From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning. *International Journal of Artificial Intelligence in Education*, *15*(4), 261–290.

Spada, H., Meier, A., Rummel, N., & Hauser, S. (2005). A New Method to Assess the Quality of Collaborative Process in CSCL. In *Proceedings of th 2005 Conference on Computer Support for Collaborative Learning: Learning 2005: The Next 10 Years!* (pp. 622–631). Taipei, Taiwan: International Society of the Learning Sciences.

Stahl, G. (2002). Rediscovering CSCL. In *Cscl* (Vol. 2, pp. 169–181).

Stahl, G. (2007). Meaning Making in CSCL: Conditions and Preconditions for Cognitive Processes by Groups. In *Proceedings of the 8th International Conference on Computer Supported Collaborative Learning* (pp. 652–661). New Brunswick, New Jersey, USA: International Society of the Learning Sciences.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. *Cambridge Handbook of the Learning Sciences*, *2006*.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, *5*(1), 5–42.

Team, R. C., & others. (2012). R: A language and environment for statistical computing.

Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2014). Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. *Computers & Education*.

Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2013). Teacher interventions in a synchronous, co-located CSCL setting: Analyzing focus, means, and temporality. *Computers in Human Behavior*, *29*(4), 1377–1386.

Voyiatzaki, E., & Avouris, N. (2014). Support for the teacher in technology-enhanced collaborative classroom. *Education and Information Technologies*, *19*(1), 129–154.

Voyiatzaki, E., Polyzos, P., & Avouris, N. (2008). Teacher tools in a networked learning classroom: monitor, view and interpret interaction data. In *Networked Learning Conference, Halkidiki, Greece*.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*(1), 71–95.

Wichmann, A., Giemza, A., Hoppe, U., & Krauß, M. (2009). Effects of Awareness Support on Moderating Multiple Parallel E-discussions. In *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning - Volume 1* (pp. 646–650). Rhodes, Greece: International Society of the Learning Sciences.

Woo, Y., & Reeves, T. C. (2007). Meaningful interaction in web-based learning: A social constructivist interpretation. *The Internet and Higher Education*, *10*(1), 15–25.