Towards Collaborative Convergence: Quantifying Collaboration Quality with Automated Co-located Collaboration Analytics

Sambit Praharaj*
sambit.praharaj@ou.nl
Open University of the Netherlands
Heerlen, Netherlands

Maren Scheffel
maren.scheffel@rub.de
Ruhr-Universität Bochum
Bochum, Germany

Marcel Schmitz
marcel.schmitz@zuyd.nl
Zuyd University of Applied Sciences
Heerlen, Netherlands

Marcus Specht
m.m.specht@tudelft.nl
Delft University of Technology
Delft, Netherlands

Hendrik Drachsler
drachsler@dipf.de
DIPF Leibniz Institute for Research and Information in Education
Frankfurt am Main, Germany
Goethe University
Frankfurt am Main, Germany
Open University of the Netherlands
Heerlen, Limburg, Netherlands

ABSTRACT

Collaboration is one of the four important 21st-century skills. With the pervasive use of sensors, interest on co-located collaboration (CC) has increased lately. Most related literature used the audio modality to detect indicators of collaboration (such as total speaking time and turn taking). CC takes place in physical spaces where group members share their social (i.e., non-verbal audio indicators like speaking time, gestures) and epistemic space (i.e., verbal audio indicators like the content of the conversation). Past literature has mostly focused on the social space to detect the quality of collaboration. In this study, we focus on both social and epistemic space with an emphasis on the epistemic space to understand different evolving collaboration patterns and collaborative convergence and quantify collaboration quality. We conduct field trials by collecting audio recordings in 14 different sessions in a university setting while the university staff and students collaborate over playing a board game to design a learning activity. This collaboration task consists of different phases with each collaborating member having been assigned a pre-fixed role. We analyze the collected group speech data to do role-based profiling and visualize it with the help of a dashboard.

CCS CONCEPTS

• **Human-centered computing → Interaction techniques**; *Computer supported cooperative work*; Social engineering (social sciences).

*Corresponding Author

LAK22, March 21–25, 2022, Online, USA

ACM ISBN 978-1-4503-9573-1/22/03…$15.00
https://doi.org/10.1145/3506860.3506922

KEYWORDS

collaboration, collaboration analytics, multimodal learning analytics, co-located collaboration

ACM Reference Format:
Sambit Praharaj, Maren Scheffel, Marcel Schmitz, Marcus Specht, and Hendrik Drachsler. 2022. Towards Collaborative Convergence: Quantifying Collaboration Quality with Automated Co-located Collaboration Analytics. In *LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22), March 21–25, 2022, Online, USA*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3506860.3506922

1 INTRODUCTION

Collaboration is one of the four important 21st-century skills [15]. Collaboration is said to occur when two or more persons work towards a common goal [9]. The recent interest on co-located (or face-to-face) collaboration (CC) is because of the ubiquity of sensors' usage and the rise of multimodal learning analytics [8, 24]. "CC takes place in *physical* spaces where the group members share each other's *social* and *epistemic* space." [20, p.1, emphasis added]. The social space comprises the non-verbal indicators of collaboration (e.g., non-verbal indicators from speech such as turn taking [13], total speaking time [22] and non-verbal indicators from video such as gestures and postures) and the epistemic space comprises the verbal indicators of collaboration (e.g., the actual content of discussion obtained from the group audio data [25], log data about the content of discussion) [22]. The indicators of collaboration vary depending on the context of the collaboration and these indicators help to determine the quality of collaboration in most of the cases [21, 24]. This can be attributed to the differing goals of collaboration and the group characteristics, team composition (which both can be collectively grouped under the parameters of collaboration) [24]. Majority of studies on CC in the past focused on the audio indicator type [24].

The quality of CC has been detected in the past using various indicators of collaboration derived from audio. For example, *non-verbal audio indicators* like prosody of sound such as pitch, spectral property, tone and intensity [2], total speaking time of group members [1, 3], interruptions [19], overlap or no overlap duration of speech [2]; speaking time of a group member combined with the attention of other group members measured by their eye gaze [31]; linguistic features such as pronouns, sentence length and prepositions [27, 28]; verbal features like the keywords used, topics covered [5], dialogues [11]. Contrary to a majority of studies focusing on non-verbal audio indicators, a handful of studies used *verbal audio indicators* (or the epistemic space) during CC. The verbal audio indicators are more overt as compared to the non-verbal audio indicators in-order to understand the collaboration process [25]. For example, "talk traces" [5], "meeter" [11] used verbal audio indicators of CC for the analysis. In "talk traces", Chandrasegaran et al. [5] did topic modeling during the meeting and then showed topic cluster visualizations as feedback by comparing the automatically generated topics with the topics in the pre-decided meeting agenda. The topic modeling barely scratches the surface of CC analysis based on a collection of representative keywords which is not rich enough to understand the group conversations in-depth. It does not show the proper linkage between these words and the rest of the conversation. This can lead to a loss of holistic meaning of the conversations and a possible under-representation of certain topics when observed contextually. The "meeter" study [11] dealt with dialogue classification of group members based on a lab study to measure information sharing and shared understanding while generating ideas. The performance (or *quality of collaboration*) was measured based on the number of ideas the group members wrote down on the cards after quality check. They did not find significant effects of information sharing and shared understanding on the quality of collaboration. So, these controlled studies on epistemic space of collaboration are too abstract in either choosing representative keyword clusters or few select dialogues' categories which do not affect the collaboration quality. Moreover, epistemic space was manually coded and epistemic network analysis was done to compute the quality of collaboration process during knowledge building [18].

Similar to the partially automated analysis of epistemic space during CC, there has been a battery of works [12, 30] on manually operationalizing the quality of collaboration based on in-depth analysis of the content of the conversations (using *convergence* as a measure) mostly in controlled settings using collaboration scripts and jigsaw scripts. Different types of convergence have been defined in the literature encompassing CC. Knowledge convergence in the context of collaboration has been defined as the increase in common knowledge (i.e., knowledge that all the collaborating group members possess) [12]. The main goal of knowledge convergence is learning together [30]. Similarly, another convergence measure is cognitive convergence which is composed of the different concepts that can be used to describe important processes underlying successful collaboration [30].

To this end, we have the following research questions:

RQ1 What co-located collaboration indicators have been identified from group speech data in the related literature?

RQ2 How can co-located collaboration indicators from group speech data automatically be analyzed?

RQ3 How to visualize quality indicators of collaboration from group speech data?

To answer **RQ1**, we do a brief literature review in Section 2 to get a background overview. Then, to answer **RQ2** and **RQ3**, we design an experimental set up where we collect audio data of 14 different sessions with university staff (pre-assigned with different roles) collaborating while designing a learning activity using a board game. We not only discover emerging role-based collaboration patterns longitudinally across the sessions but also discover collaboration patterns in the session itself. For this analysis, we used automated analytics by visualizing both the epistemic and social space using different methods (as described in Section 3). These methods include the network graph analysis to find rich interconnections of the discussion, how closely each keyword is related to each other and also understand the speaking time and turn taking patterns of group members with different roles. Moreover, we analyze the collaborative convergence patterns in a session automatically motivated by the past manual works on convergence [12, 30]. In the context of our study, we define collaborative convergence from 2 different perspectives: 1) Group level convergence - Convergence between members during collaboration with respect to the expected objectives of the discussion and 2) Individual level convergence - Convergence of group members' role during collaboration with respect to the expected role-based objectives before collaboration. So, instead of analyzing the pre-knowledge and post-knowledge after collaboration of each group member (as has been done in the past), we analyze how the major influential role-role interactions (detected from turn taking and speaking time) contribute to the group's collaboration task temporally across the session by understanding the different conversation patterns. We show these visualizations using a dashboard (in Section 4) and then discuss the future research that can be done on this dashboard (in Section 5). Finally, we discuss our findings (in Section 6) with the limitations and conclude in Section 7.

2 CO-LOCATED COLLABORATION INDICATORS FROM AUDIO

The majority of past works used audio indicator types to determine the quality of collaboration [24]. The simplest audio indicator of collaboration used was *total speaking time* [1, 3]. The total speaking time of each group member was reflected back to them by a coloured LED light display [1] and concentric circles visualization [3] on the smart table during group meetings. It was found that this helped to regulate the equity of participation during a group conversation. The dominant speakers spoke less and the not so dominant speakers started to speak more. It was later found that the group that had better equity of speaking time also had better quality of collaboration as measured by a post-test. Besides, more frequent speaker changes (i.e., *turn taking*) with *overlap of speech* [13] indicates a good quality of collaboration. Previous research also indicates that overlap in speech is associated with positive group performance [4, 10].

Other *non-verbal audio indicators* used to detect collaboration quality were both group-based (i.e., solo duration, overlap duration of two persons, overlap duration of all three persons) and individual-based (i.e., spectral, temporal, prosodic and tonal) [2].

These indicators along-with manual annotation were fed to a support vector machine classifier to compute the collaboration quality. Similarly, *speaker-based* indicators like the intensity, pitch and jitter were used to detect collaboration quality among working pairs [16]. When two members in a group are speaking at different amplitudes but exhibiting the same pattern of their speech (e.g., the rise and fall of average pitch of both members are similar to each other) then they are showing a high level of synchrony [16]. Lubold & Pon-Barry [16] found a positive correlation between synchrony and rapport (obtained by comparing perceptual rapport from annotators and self-reported rapport) during collaborative interactions. A good rapport between group members can enhance the collaboration [6]. As seen from the examples of past studies, the indicators of collaboration are dependent on the context. This can be attributed to the differing goals and fundamental characteristics or parameters (such as group behaviour, interaction, composition) of the group in each collaboration context. “The parameters of collaboration are primary aspects such as *team composition* (e.g., experts, initiators or roles of being initiators), *behaviour of team members* (e.g., dominance, rapport, conflict), *types of interaction* (e.g., active or passive), *behaviour during collaboration* (e.g., knowledge co-construction, reflection, coherence, misconception, uncertainty)” [25, p. 4].

Additionally, Zhou et al. [19] tracked the speech of students during collaborative maths problem solving. They found that *overlapped speech* is an indicator of constructive problem-solving progress, expertise and collaboration. They used both the *number of overlap* in speech and the *duration of overlap* in speech. Luz & Saturnino [17] used the *non-verbal audio indicators* like speech, silence, pause, transition from group speech to individual speech as indicators to predict performance and expertise on a Maths dataset corpus of groups during collaborative problem solving. Using these non-verbal indicators as features, they trained a model to predict the expertise of the group members and their collaborative performance. They found that these features were able to predict the expertise but not the group performance. Spikol et al. [29] used *audio level* and other *non-verbal indicators* to estimate the success of collaboration activity (i.e., measured by the human observers) while performing open-ended physical tasks around a smart furniture. They found that audio level alone is sufficient to predict the quality of collaboration with high accuracy.

These non-verbal audio indicators studied above are less overt as compared to the verbal audio indicators which analyze the epistemic space of CC. With the rise of automatic speech recognition (ASR) techniques, a handful of studies (for instance, "talk traces" [5], “meeter” [11]) took into account the content of collaboration. In “talk traces”, Chandrasegaran et al. [5] did topic modeling during the meeting and then showed the topic clusters (shown as a representative overview of a group of keywords) as a visualization feedback by comparing with the agenda of the meeting. Although topic modeling shows a representative overview of the different topical word clusters and their evolution during collaboration, it does not show the relationship of the words with each other and their contextualization in the whole conversation. This can lead to a loss of holistic meaning of the conversations and a possible overlooking or under-representation of certain topical themes. In “meeter” [11] the dialogues of the group members were categorized based on a controlled study to measure information sharing and shared understanding while generating ideas. The collaborative task was based on three open ended fixed topics where group members needed to brainstorm and share their ideas in a short session of 10 minutes. Collaboration quality was measured by the number of ideas generated. No significant effects of information sharing and shared understanding was found on the quality of collaboration. So, most of these studies give an abstract overview of the conversations.

Prior to the prevalence of the ASR techniques, there have been some manual studies on in-depth analysis of the content of conversations during CC in controlled settings using collaboration scripts (which describe certain rules for collaboration) and jigsaw scripts (which are individual pieces of knowledge not all identical to each other and shared with all group members so that each member has a unique knowledge piece script). This has been possible by conceptualizing convergence from different perspectives like knowledge convergence [12] and cognitive convergence [30] (i.e., all concepts used to describe important processes underlying successful collaboration). Convergence has been defined as the increase in common knowledge (i.e., knowledge that all the collaborating members possess). This was possible by a pre-test and post-test and comparing the knowledge gain of the group members. So, the main goal of knowledge convergence is to develop better shared mutual understanding and learning together [30]. This effectively improves collaboration.

Table 1 gives an overview of some of the studies on detecting the indicators of collaboration from audio and their operationalization to measure the quality of collaboration.

In the scope of this work, we focus on field trials in a real world setting to semantically understand the content of discussion during collaboration by analyzing and visualizing the epistemic space with some emphasis on the social space too.

3 EXPERIMENTAL SET UP

In this section we describe our experimental context, set up, data collection, processing and the methods used for our experiments, data analysis and visualizations.

3.1 Experimental Context

The collaboration task was to design a learning activity using the Fellowship of Learning Activity and Analytics (FoLA^2)[1] method. We used this task to collect the audio recordings. It is a co-design method which comes with a board game [26] (e.g., of an online version [2] currently under active development) played face-to-face with different themed cards and roles that are used in workshops to create awareness of the connection between learning analytics and learning design. It can also be used as an instrument to collect indicators when planning learning analytics already while designing learning activities. This game was used in 14 face-to-face sessions in between September and October 2020 (with each session varying between 60-90 minutes) among different university staff and students. The collaboration task in each session had different phases which were colour-coded based on the cards supposed to be used in

[1]http://www.fola2.com/
[2]https://game.fola2.com/

Table 1: Indicators of CC and their operationalization of collaboration quality

Parameters	Indicators	Operationalizing collaboration quality	Space tracked	References
Roles (one leader and other non-leaders)	Topics covered detected from keywords, frequently used keywords and phrases	Topical closeness to meeting agenda, proximity of commonly used words and phrases to the roles	Epistemic	[5], [25]
Dominance	Total speaking time	Higher equity of total speaking time means less dominance in the group and higher quality of collaboration	Social	[14], [1], [3], [23]
Active participation	Turn taking frequency	More frequent turn taking changes mean higher active participation and better quality of collaboration	Social	[13]
Expertise	Overlapped speech	Overlap in speech is an indicator of constructive problem solving, expertise and good CC quality	Social	[33], [19]
Rapport	Synchrony in rise and fall of average pitch	Higher synchrony in rise and fall of average pitch indicates higher rapport and better collaboration quality	Social	[16]
Knowledge co-construction	Knowledge convergence (i.e., the amount of shared knowledge in the group), Cognitive convergence	Increase in convergence (i.e., increase in the shared knowledge) implies increase in collaboration quality	Epistemic	[12], [30]

that phase (as *blue, red* and *yellow*) [3]. Each group member performing the task was assigned different roles. The *blue* (card or) phase (varied in length from 28 to 57 minutes across the sessions) defines the interaction steps in the learning activity. The *red* phase (varied in length from 4 to 13 minutes across the sessions) or Learning Enhancing Technology cards are part of the step in the game where we search for enhancements of the interactions using technology such as sensors, virtual reality, smart boards, discussion boards etc. The *yellow* phase (varied in length from 10 to 30 minutes across the sessions) defines what we want to know about the interaction steps or the learning activity, e.g., engagement or how students take initiative. To steer the group conversations during the sessions, there were also prompts on each role card. A demo from one of the game sessions is shown in figure 1.

We recorded the conversations during these sessions (after gathering signed informed consent from the participants) using clip-on microphones attached to each group member along-with their respective audio recorder which recorded and stored their conversation locally in that recorder. The conversations were in Dutch. Each group member was pre-assigned roles during the conversation: *Game master*, *all advisors* (consists of the “technology enhanced learning and learning analytics” advisor and “educational” advisor), *study coach*, *teacher* and *learner* (or student). These roles resemble a real life student, teacher, study coach or advisor while the game master is the game moderator who also helped to steer the group conversations. The roles were also played by actual advisors (age varied from 36–64 years, experience varied from 1–20 years and gender distribution was 10 males and 8 females), teachers (age varied from 28–64 years, experience varied from 1–40 years and gender distribution was 13 males and 1 female), learners (age varied from 19–27 years, experience varied from 0.16–9 years and gender distribution was 13 males and 1 female) and study coaches (age varied from 28–56 years, experience varied from 0–14 years and gender distribution was 12 males and 2 females).

3.2 Methods

Our architecture for data collection, processing and analysis was based on our previous work [25]. We followed a similar data collection, pre-processing and processing approach (as in [25]) only with a minor exception of using Amberscript[4] for speech to text transcription instead of directly using Google Speech to Text. Amberscript uses Google Speech to Text behind the hood but provides

[3] The phases and cards mean the same and are used interchangeably henceforth

[4] https://www.amberscript.com/en/

Figure 1: A game session demo

a much cleaner user interface to play with the transcribed data and make minor modifications when needed.

After data processing (where we cleaned the dataset and made it machine understandable), we found that the 4th and the 13th session needed to be removed because of poor quality of recordings and thus incomplete transcriptions. Then we moved to the data analysis and visualizations with the remaining 12 sessions. First, we visualized them in an exploratory manner to see the frequently used keywords (in the processed text model obtained from speech) by different group members playing different roles by using frequency analysis of the common keywords used in different phases and sessions. To make sense of the visualizations and understand the context of the conversations, we consulted with the game master to generate summarized annotations for each phase (one from each session theme) in English (for example, a sample annotation can be seen in figure 2).

To go in-depth into the influential role-role exchanges, we explored the social space by visualizing with a network graph the speaking time of the group member shown in terms of node size and the turn takings shown in terms of the edges between the nodes. The thickness of the edges is directly proportional to the number of turn takings between the different roles. A sample network graph can be seen in figure 3 in the Results section. Then, we analyzed the words used during the conversations by these roles with the help of bigrams (consecutive two-word phrases) and ranking them by tf-idf to check how often the bigrams have been used. Tf-idf ranking of the bigrams gives an overview of the frequently (with lower tf-idf ranking) and rarely (with higher tf-idf ranking) used bigrams. We used bigrams over trigrams (consecutive three word phrases) because they were more informative in our context.

To analyze the epistemic space (consisting of the content of the conversations), we built a co-occurrence matrix which shows the strength of the different word combinations (i.e., how often certain word combinations occur together). Then, we used this co-occurrence matrix to build an interactive network graph to visualize (as shown in figure 12) the frequency of the different words (denoted by node size) along-with how many times these words co-occur together (denoted by the edge thickness). To make the network graph visualization easier to play with and intuitive for the end user, we built an interactive feature which helps to highlight a specific node and its neighbours in the graph by selecting that specific node or searching for that specific node. Finally, we combined both these social and epistemic components into a single dashboard to get a better understanding of the collaboration sessions. Further, we also explored different centrality measures in the graph network to understand the importance of different keywords used contextually. Moreover, for the analysis of the roles and their interactions, we did not take into account the contribution of the game master as he was only a moderator and steering the discussion. In the next section we describe our findings.

4 RESULTS

As we described before, we consider mainly 4 roles (i.e., *all advisors*, *study coach*, *teacher* and *learner*) for the analysis of the conversations. In total we have 3 phases (i.e., blue, red and yellow) in each session with a total of 12 sessions. First, we observe the social space (as in Figure 3) which shows the speaking time and turn taking of the group members and then we explore the epistemic space (as in Figures 7 and 12). Here, we describe our findings in three different sub-sections.

4.1 Convergence within a phase in a session

We have defined *collaborative convergence* in the Introduction as per the context of our study. It is defined from both group level and individual level. Group level convergence is between members during collaboration with respect to the expected objectives of the discussion and individual level convergence is convergence

Blue		Extra introduction and explanation of blue cards. Explaining Belbin team s which are the topic to be taught.
Blue	Vraag test (Teacher->Learner)	Start of the learning activity. Making groups? Do they already know this? Previous knowledge of previous learning activities are disucssed. We start with the question: do you know abo
Blue		Discussion on what to do if question asked results in that there was no action by students. Option split up groups. Option go back to team last quarter. Making groups based on resul
Blue		Possibility to reflect on Belbin last quarter. Game master intervenes in hearing good ideas, make cards out of it! **Learner -> Learner: Group blok 1 card played.** Discussion on grou
Blue	Belbin roles (Material -> Learner) Discussing each	Learner -> Learner where students discuss what the other one's Belbin role is is discussed and made. **Learning Material -> Learner Belbin material** is placed after that the **L->L c**
Blue	Team assembly ? -> Learner -2bechecked-	Teamfocus could be discussed with the entire class and not persee in the groups now. What is first: team or focus is discussed. Awareness where am I know is important. Is at the er
Blue	List of students and their role card (Leaner -> Mate	**Making a list of students and roles card is made**. Post its or something, like a Padlet. Discussion that the ideal team and focus should be put in the middle.**Class discussion abo**
Blue	Ideal team class discussoin (Teacher -> Learner) S	**What is the ideal team card is made. Class discussion. Teacher -> Student** after that card focus can be discussed. In the general assignment a task is mentioned. A card is ma
Red	Blanco (Moodle assignement), Interaction booster	Introduction cards and parts of cards. **Moodle for the assignement** at the end is made on a blanco red card. Concept mapping tool is looked at as could be used. Smart Screen, Wi
Red		Blog writting is discussed, seems to early for this group. Game master wonders if overview roles per students can be done, mobile phone is used there. Game Master asked if Concep
Red	Mobile phone is added, Concept mapping tool take	Discussion ends
Yellow	Social Interaction, Initiative	Do we want as a teacher know things about the learning proces or the design? Some of the examples of yellow cards are directly available others are not. There are empty cards. Exa
Yellow	Presence, Having Fun	Presence and Activity are discussed. Having fun is discussed. **Presence** and **Having fun** are placed. Perhaps students can search for symbol, avatar, icon belonging to Belbin Roles
Yellow	**Teamname (Student -> Material)** Activity is cha	**Blue card Student -> Material with new teamname of logo is added.** Discussion is put out there by Game Master if all the yellow cards can be used next quarter. Presence or A
Yellow	Social Interaction is taken from the board	How is the social interaction discussed, decibel meter is an option. Observation of the teacher which is not recorded. Do we want to know it. Group takes Social Interaction card away
Yellow		Other yellow cards are discussed. **Having fun** with a laughing meter. Smiley-based check. Game Master suggests **Shakepseak/Interaction** Meter. Doing it with shakespeak the que
Yellow	Initiative is taken of the board. **Moodle Test card**	The smiley-airport-KLM-solution can always be done. The other initiative goes of the board. Discussion is raised that every TEL card has a green card. Do we put some of the available

Figure 2: A sample annotation

of group members' role during collaboration with respect to the expected role-based objectives before collaboration.

To understand *group level convergence*, we take a simple example of one phase (i.e., blue phase in this case) in a session (i.e., 1st session in this case). During the blue phase participants talked about the steps in the learning activity they were planning, i.e., a group discussion, team focus and the different team roles that can be involved. Each learning activity consists of a sequence of interactions such as learner to teacher, learner with learning environment, material to learner and so on. If we compare Figure 7 and 8, then we observe "Belbin" as a keyword. Belbin team roles are actually nine different team role behaviours that make a high performing team. We can see that in the 1st 10 minutes of the blue phase (i.e., in Figure 7) only advisor, learner and study coach uttered the term "Belbin". But, in the first 20 minutes of blue phase (i.e., in Figure 8), the teacher also started to speak about "Belbin". So, effectively the group level convergence (i.e., shared utterance) increased with reference to a highly relevant term (i.e., "Belbin" in this case). This implies an increase in the quality of collaboration with respect to the group level task objectives because of an increase in task related shared epistemic space. We can also observe a similar increase of convergence with the inclusion of the teacher for the term "docent" ("teacher" in English) across the first 10 and 20 minutes of the blue phase. Towards the end of the conversation, many new terms like "reflecteer" ("reflect" in English) and "klassikaal" ("classical" in English) came up to the top frequently occurring group (when we compare Figures 7 and 8 with Figure 9 or 10). That conveys the change in focus of the conversations initially from "Belbin" roles to later reflect on these roles and also conversations about the classical discussion for their students vs splitting them in groups. Similar to epistemic convergence, social convergence can be observed based on the participation of different group members with pre-assigned roles (i.e., *all advisors*, *study coach*, *teacher* and *learner*) across every 10 more minutes slices of conversation in Figures 3, 4, 5 and 6 of the blue phase. The learner in the 1st 10 minutes did not interact with the study coach and spoke the least amount in terms of speaking time (shown as the smallest node). Towards the end, however a link with the study coach developed because of a turn-taking exchange between them. So, both social and epistemic space give us a holistic understanding of the evolving conversation patterns within the phase of a session.

Regarding *individual convergence*, a rise of the individual role-based words with respect to the group can be considered as an increase in convergence. For example, use of the "team" keyword increased a lot from first 10 minutes to the end of the blue phase (as can be seen in Figures 7, 8, 9 and 10) with respect to its usage by the learner and also in the whole group. These were mainly discussions on how to form an ideal team, team focus and making a new team while discussing the different steps in the learning activity. Next, we move out of one phase to get a high level overview of the conversations in one phase across sessions using bigrams or phrasal analysis.

4.2 Overview of epistemic and social space of red phase across all sessions

We obtained the relevant phase-related bigrams (obtained from high and low tf-idf rankings automatically) for individual roles in the red phase across all the 12 sessions considered for the analysis along-with the dominant role-role exchanges in each session. The dominant role-role exchanges can be useful to get an insight into the groups' main conversations. The red phase was centered around conversation on technologies that can enhance learning. So, we were interested to get an overall idea of the red phase across all the sessions.

As expected many discussions were on different technologies to improve learning. "Concept map", "smart screen", "smart board", "shakespeak" (i.e., an interaction polling system used for interactive lectures in the classroom, which can act as an interaction booster), "powerpoint", "moodle" for assignment, "padlet" and "online collaboration" were different technological terms among many. The advisor consists of the "technology enhanced learning and learning analytics" advisor who was expectedly dominant in this technology phase across most of the sessions. The learner and study coach were having conversations about "moodle" a lot for assignments. Some sessions focused on "surveys" to collect user requirements, moved in the direction of statistics evident from usage of "statistics", "histogram", "graphics", "manner" in which data is collected, saved and how "feedback" is given to "reflect" during the conversations. The

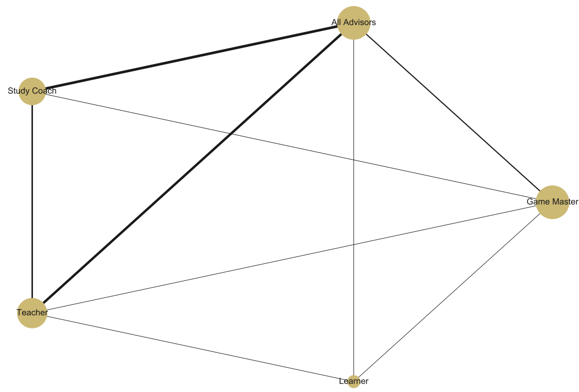

Figure 3: 1st 10 minutes social space (1st session blue phase)

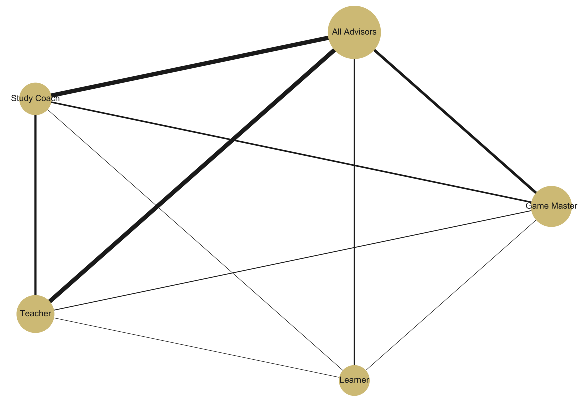

Figure 4: 1st 20 minutes social space (1st session blue phase)

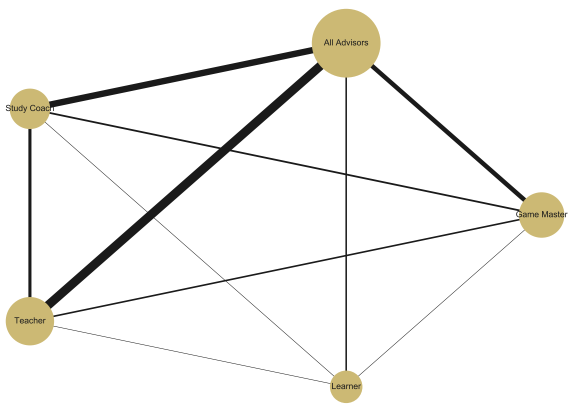

Figure 5: 1st 30 minutes social space (1st session blue phase)

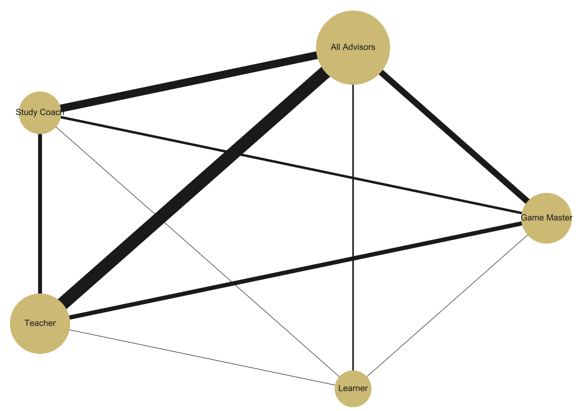

Figure 6: Full social space (1st session blue phase)

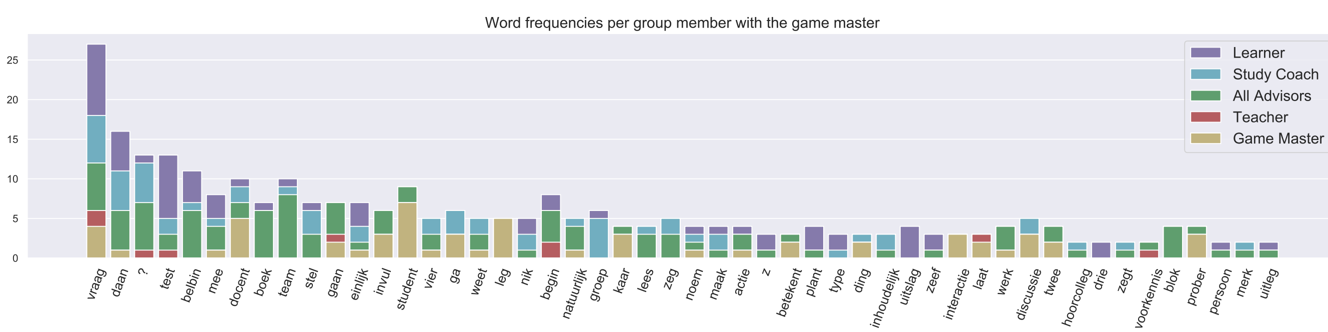

Figure 7: Top 50 word utterance frequency in the 1st session blue phase in 1st 10 minutes with roles

other sessions focused more on “team composition”, “characteristics of students” (i.e., extroverted or not), “collaboration environment” and “online collaboration”. To understand the conversation patterns in-depth, the network graph-based dashboard described in the next sub-section can be helpful.

4.3 Dashboard encompassing the social and epistemic components

Figure 11 shows the dashboard highlighting a node for all advisors in the red phase in session 1. It has four main components.

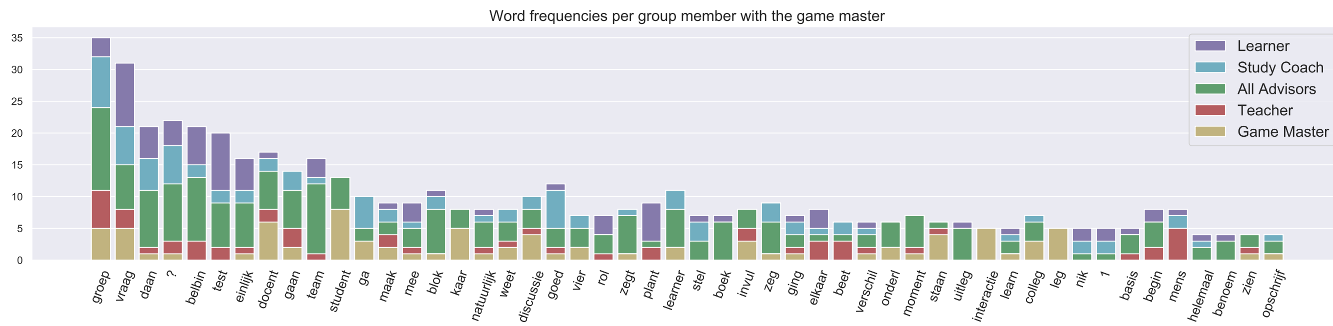

Figure 8: Top 50 word utterance frequency in the 1st session blue phase in 1st 20 minutes with roles

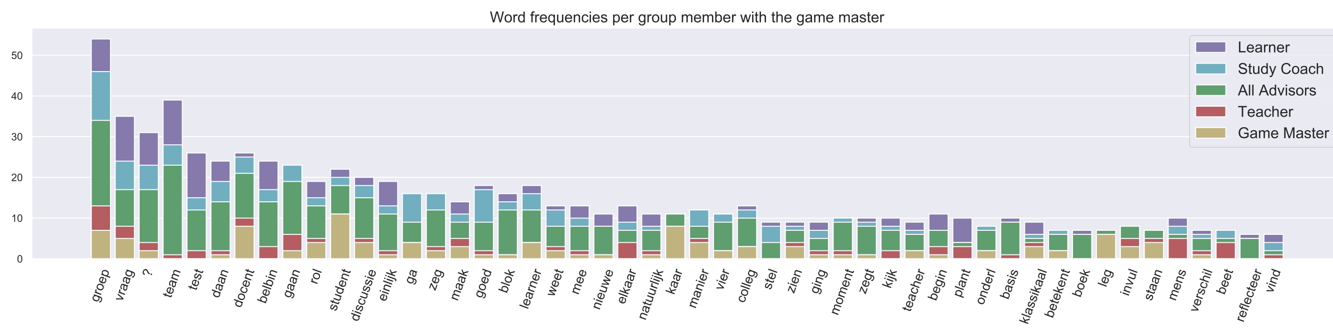

Figure 9: Top 50 word utterance frequency in the 1st session blue phase in 1st 30 minutes with roles

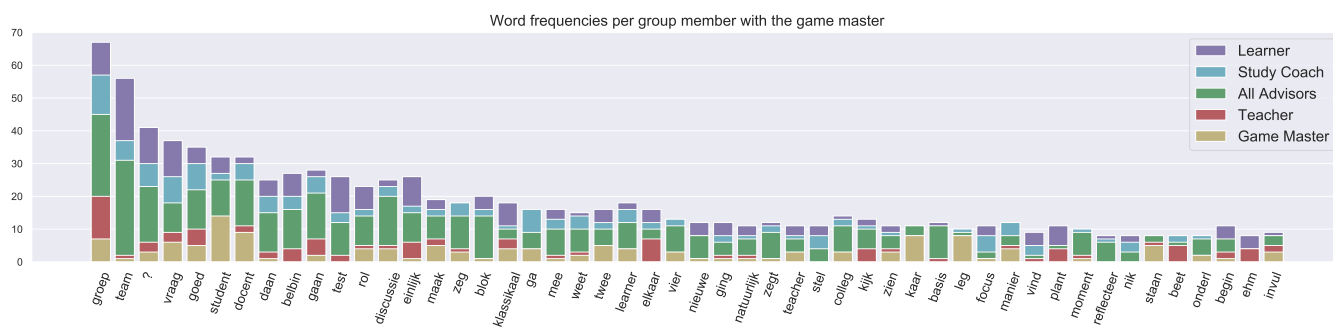

Figure 10: Top 50 word utterance frequency in the full 1st session blue phase with roles

The social space shown by the role network graph, the high level overview of the epistemic space shown by the bar graph, the colourful network graph showing the interaction of a particular role in one phase of a session and the search bar which helps to search and highlight a specific node (which is also possible on clicking on that particular node). Now we have different views for each phase and session with each view showing the conversation of one role in the whole conversation network graph. This will make it easier to compare two roles' conversation patterns when they are seen side by side. This dashboard can be scaled easily and is fully dynamic and interactive.

Figure 12 shows a zoomed in version of the advisor role among other roles with different shape and colour. The colour and shape of the node helps in the distinction of roles. The neighbours of each node (or in other words which words co-occur with each other) are shown on hovering the mouse over the node. Similarly, the strength of the words that co-occur (shown by the thickness of the edge) is also shown when we hover the mouse over the edges. This graph helps us to understand the different contextual keywords, how often they have been used, what are they associated with strongly and weakly (measured based on the edge strength of the nodes).

To analyze the network graph in depth, we looked at different *centrality measures* such as the betweenness centrality (BC) and eigenvector centrality (EC) of these words. Betweenness centrality shows how often a node (or word) acts as a bridge node, that is the

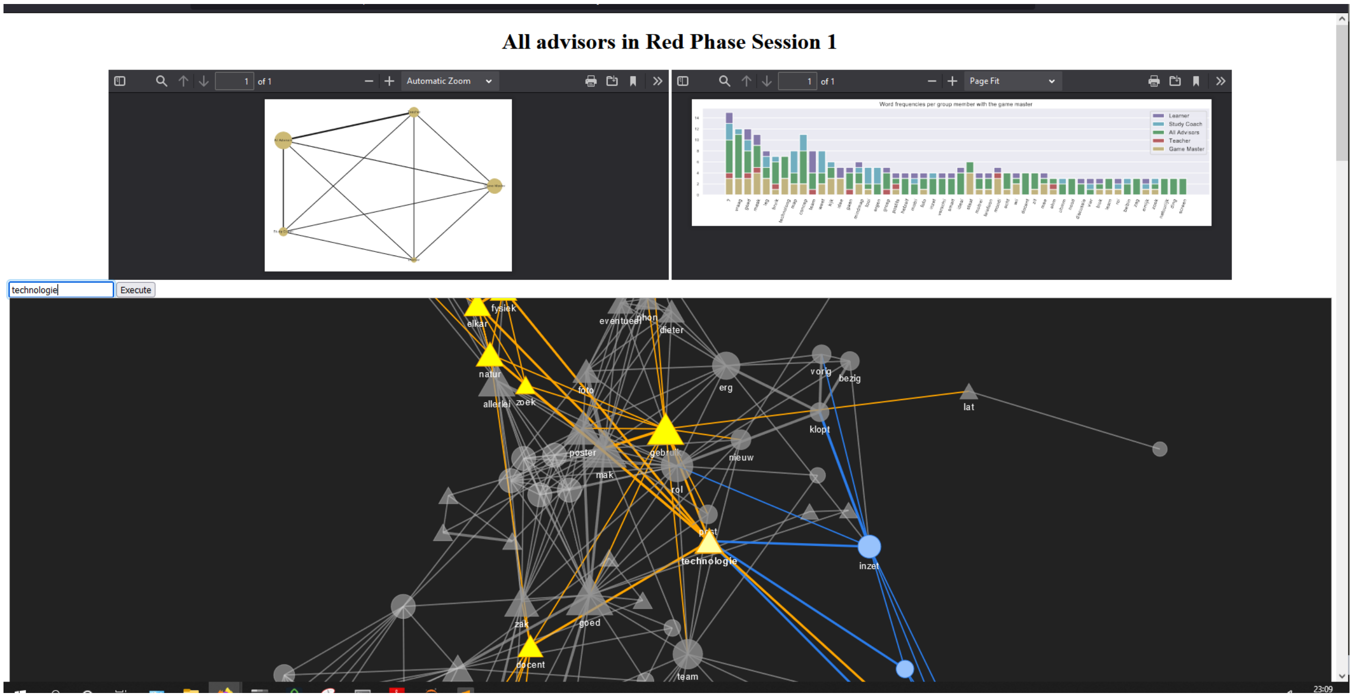

Figure 11: Screenshot of the dashboard with social and epistemic components

number of times a node lies on the shortest path between other nodes. This means that a node (or a word) with high betweenness centrality would have more control over the network. Another centrality measure that can be a good indicator of the influence of a node is eigenvector centrality. Therefore, a node with a high eigenvector centrality score must be connected to many other nodes who themselves have high scores. For example, in the red phase of the 1st session, frequency wise four words in decreasing order were “good”, “make”, “moodle” and “use”. But, BC-wise it was “good”, “team”, “use” and “technology”, and EC-wise it was “make”, “poster”, “good” and “role”. So, this example shows that centrality measures can elevate the ranking of even less frequently used words (i.e., “team”, “technology” and “role” in this example) in that particular context.

5 FUTURE RESEARCH ABOUT THE DASHBOARD

We have built a generic dashboard to quantify collaboration quality based on different collaboration indicators in the social and epistemic space. This dashboard is useful to show how each role interacted during the collaboration task and with whom. Now, the important question is: “Who would use it and why?”. This question will be answered by understanding the needs of the dashboard design.

The design of the dashboard will be driven by the temporal needs (i.e., whether updated in real-time every few minutes or shown as a summary at the end of collaboration) and the stakeholders’ (teacher or task moderator or the group members themselves) needs. To cater to the temporal needs, we need to first differentiate what can be shown as an immediate formative feedback and what can be shown as a summative feedback at the end of collaboration. To this end, we need to do a qualitative study by interviewing different stakeholders to identify the user requirements. This will give us an idea as to what type of feedback is relevant for which stakeholder group and can be shown to them accordingly. For example, this type of dashboard for a teacher (as the stakeholder) could be useful to determine scaffolding strategies during collaboration and also planning the collaboration sessions. For the group members, it can be a useful tool to self-reflect and adapt their collaboration accordingly.

Based on that we can also do design enhancements and modifications in the dashboard using different visualization filters to capture and compare temporal role-based snapshots. The customizability should be extended to the users using the dashboard.

6 DISCUSSION

To answer “RQ1: What co-located collaboration indicators have been identified from group speech data in the related literature?”, we did a short literature review where we identified different indicators of CC quality from group speech data and defined how they have been operationalized contextually to measure the quality of CC. We find that most studies [1, 14] in the past focused on the analysis of the social space of collaboration and few studies [5, 11] that focused on the epistemic space were abstract in nature or labour intensive because of manual coding. We overcome this limitation in RQ2 by conducting field trials.

To answer “RQ2: How can co-located collaboration indicators from group speech data automatically be analyzed?”, we conducted

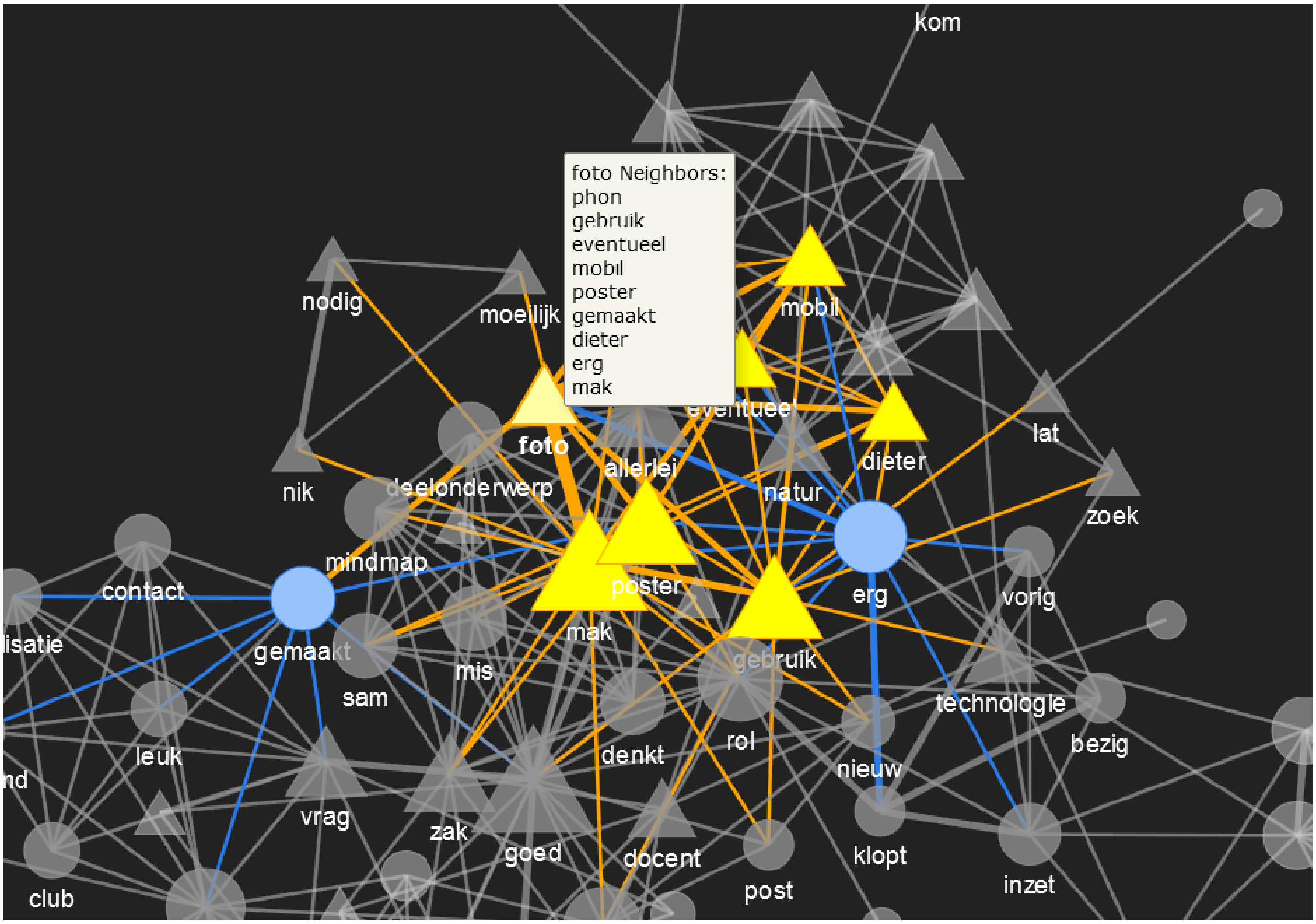

Figure 12: Zoomed in network graph highlighting a node of the advisor in yellow rectangles and rest others in blue circles in red phase

field trials in 14 different sessions (only 12 of which were later used for data analysis) where we collected the audio recordings. The collaboration task was to design a learning activity with each group member having been assigned a pre-fixed role (such as teacher, all advisors, study coach, learner and game master) before collaboration. Each session had 3 different phases (i.e., blue, red and yellow), each with different objectives. Here, we used the already defined indicators of collaboration quality covered in the literature review. We analyzed the collaboration convergence (i.e., increase in shared utterance of specific phase related keywords) automatically as evolving conversations in a phase motivated by manual knowledge convergence (i.e., increase in shared knowledge) and cognitive convergence studies done earlier [12, 30]. We find that analysis of utterance frequency of specific keywords for different roles helps in this regard to understand the change in role-based conversation patterns with time. This is because the more utterances we have of a specific phase-related keyword, the more is its usage in that context and hence, more importance. The convergence patterns help us to understand how specific conversations were discussed by all roles or specific roles. Combined with the social space analysis (shown as a role-role interaction network graph), the holistic overview of how the conversations evolved can be obtained. This helped us to quantify the collaboration quality. So, we do not categorize whether higher or lower convergence is good or bad. We just show an approach to quantify collaboration and categorizing is up to the context of collaboration. For instance, in our study if there is higher convergence for on-topic conversations then it is good for the quality of collaboration but higher convergence for off-topic conversations is bad for collaboration quality. As we do not define fixed objectives before collaboration and do not conduct a lab-based study, so it is quite open to interpretation.

To answer "RQ3: How to visualize quality indicators of collaboration from group speech data?", we built a dashboard encompassing both the social and epistemic components. We used network graphs and bar graphs to show the role-role interaction in both the social and epistemic space respectively. To understand the epistemic space further in-depth, we built an extended interactive network graph to do role-based profiling of the conversations during collaboration. This helped us to understand which words have been frequently used (shown as node size) by different roles and what are the strength of the word co-occurrences (i.e., how often multiple words co-occur together). This network graph is an intuitive, interactive dynamic representation which can reveal more information by mouse hover such as the strength of linkage between keywords and the keywords connected to one. It is a modified large scale network similar to online SNA [32] and better than the abstract representation of topics [5]. We also computed the centrality measures [7] and found that graph centrality measures convey richer information about the importance of less frequent keyword in the context of the conversation. Finally, these visualizations are combined to form a dashboard to analyze and understand evolving collaboration patterns, hence compute the quality of the collaboration. Besides, the dashboard needs to be customized in the future depending on the use-case and the stakeholders who will use it as discussed above. This will also determine whether the dashboard is updated every few minutes for real-time feedback or given as a post-hoc summative feedback.

So, our main contribution is twofold: 1) To provide definition of CC quality, 2) understanding the social and epistemic space of collaboration with an in-depth analysis on the content of the conversations using the network graphs and bar graphs in a dashboard. To this end, we also detected convergence to quantify the quality of collaboration, defined it in the context of our study and left this dashboard as an open option for anyone to build and customize it further.

However, there are certain limitations in terms of the architecture, analysis and the visualizations. The transcription of audio data needs human intervention to do sanity checks especially when any names are concerned. The dashboard is now a generic version which is information rich instead of being bereft of information. Even though it provides a holistic overview of the collaboration patterns, it is not suited for a specific stakeholder group which can be customized further depending on their needs to make it suitable for them. Finally, the network graph can be overwhelming when it is generated for a long time duration where the number of nodes and edges can cause overcrowding, clutter and the popular hairball problem in large network graphs. This should be tackled while considering the dashboard design.

7 CONCLUSIONS

First we did a brief literature review to identify the indicators of collaboration quality from group speech data and define the operationalization of the CC quality. Then, we conducted field trials. Here, we analyzed and visualized the audio recordings collected in 12 different sessions of a collaboration task of designing a learning activity. This is a starting step in the direction of automated collaboration analytics to understand co-located collaboration patterns and give feedback. For this, we analyzed both the epistemic space (i.e., the content of the conversations) and the social space (i.e., the speaking time and turn takings) to get a holistic understanding of the evolving collaboration patterns and CC quality.

Apart from the simple analysis of the frequency of the keywords, we also analyzed the richness of these conversations with an interactive network graph to understand the contribution of each role in terms of what they spoke and how strongly a specific keyword or phrase is related to each other. To understand the role-level contribution, we explored the network graph and also different convergence patterns across a phase in a session. We found that this can be temporally computed by finding the shared utterance of the frequently used keywords among different roles and will be helpful to quantify the CC quality. For visualizing the social space, we used network graph to show the role-role interaction in terms of speaking time and turn-taking. Finally, we built a dashboard which is made up of both the social and epistemic components to analyze the emerging collaboration patterns and get a holistic understanding of CC quality.

REFERENCES

[1] Khaled Bachour, Frederic Kaplan, and Pierre Dillenbourg. 2010. An interactive table for supporting participation balance in face-to-face collaborative learning. *IEEE Transactions on Learning Technologies* 3, 3 (2010), 203–213.
[2] Nikoletta Bassiou, Andreas Tsiartas, Jennifer Smith, Harry Bratt, Colleen Richey, Elizabeth Shriberg, Cynthia D'Angelo, and Nonye Alozie. 2016. Privacy-Preserving Speech Analytics for Automatic Assessment of Student Collaboration.. In *INTERSPEECH*. 888–892.
[3] Tony Bergstrom and Karrie Karahalios. 2007. Conversation Clock: Visualizing audio patterns in co-located groups. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. IEEE, 78–78.
[4] Özgür Çetin and Elizabeth Shriberg. 2006. Overlap in meetings: ASR effects and analysis by dialog factors, speakers, and collection site. In *International Workshop on Machine Learning for Multimodal Interaction*. Springer, 212–224.
[5] Senthil Chandrasegaran, Chris Bryan, Hidekazu Shidara, Tung-Yen Chuang, and Kwan-Liu Ma. 2019. TalkTraces: Real-time capture and visualization of verbal content in meetings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.
[6] Carole Chapman, Leonie Ramondt, and Glenn Smiley. 2005. Strong community, deep learning: Exploring the link. *Innovations in education and teaching international* 42, 3 (2005), 217–230.
[7] Kousik Das, Sovan Samanta, and Madhumangal Pal. 2018. Study on centrality measures in social networks: a survey. *Social network analysis and mining* 8, 1 (2018), 1–11.
[8] Daniele Di Mitri, Jan Schneider, Marcus Specht, and Hendrik Drachsler. 2018. From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning* 34, 4 (2018), 338–349.
[9] Pierre Dillenbourg. 1999. What do you mean by collaborative learning?
[10] Wen Dong, Taemie Kim, and Alex Pentland. 2009. A quantitative analysis of the collective creativity in playing 20-questions games. In *Proceedings of the seventh ACM conference on Creativity and cognition*. 365–366.
[11] Bernd Huber, Stuart Shieber, and Krzysztof Z Gajos. 2019. Automatically analyzing brainstorming language behavior with Meeter. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–17.
[12] Heisawn Jeong and Michelene TH Chi. 2007. Knowledge convergence and collaborative learning. *Instructional Science* 35, 4 (2007), 287–315.
[13] Jaebok Kim, Khiet P Truong, Vicky Charisi, Cristina Zaga, Manja Lohse, Dirk Heylen, and Vanessa Evers. 2015. Vocal turn-taking patterns in groups of children performing collaborative tasks: an exploratory study. In *Sixteenth Annual Conference of the International Speech Communication Association*.
[14] Taemie Kim, Agnes Chang, Lindsey Holland, and Alex Sandy Pentland. 2008. Meeting mediator: enhancing group collaboration using sociometric feedback. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, 457–466.
[15] Charles Kivunja. 2015. Exploring the pedagogical meaning and implications of the 4Cs 'super skills' for the 21st century through Bruner's 5E lenses of knowledge construction to improve pedagogies of the new learning paradigm. *Creative Education* 6, 02 (2015), 224.
[16] Nichola Lubold and Heather Pon-Barry. 2014. Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*. ACM, 5–12.
[17] Saturnino Luz. 2013. Automatic identification of experts and performance prediction in the multimodal math data corpus through analysis of speech interaction. In *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 575–582.
[18] Jun Oshima, Ritsuko Oshima, and Shunsuke Saruwatari. 2020. Analysis of students' ideas and conceptual artifacts in knowledge-building discourse. *British Journal of Educational Technology* 51, 4 (2020), 1308–1321.

[19] Sharon Oviatt, Kevin Hang, Jianlong Zhou, and Fang Chen. 2015. Spoken Interruptions Signal Productive Problem Solving and Domain Expertise in Mathematics. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 311–318.

[20] Sambit Praharaj. 2019. Co-located Collaboration Analytics. In *2019 International Conference on Multimodal Interaction*. 473–476.

[21] Sambit Praharaj, Maren Scheffel, Hendrik Drachsler, and Marcus Specht. 2018. MULTIFOCUS: MULTImodal Learning Analytics FOr Co-located Collaboration Understanding and Support.. In *EC-TEL (Doctoral Consortium)*.

[22] Sambit Praharaj, Maren Scheffel, Hendrik Drachsler, and Marcus Specht. 2018. Multimodal Analytics for Real-Time Feedback in Co-located Collaboration. In *European Conference on Technology Enhanced Learning*. Springer, 187–201.

[23] Sambit Praharaj, Maren Scheffel, Hendrik Drachsler, and Marcus Specht. 2019. Group Coach for Co-located Collaboration. In *European Conference on Technology Enhanced Learning*. Springer, 732–736.

[24] Sambit Praharaj, Maren Scheffel, Hendrik Drachsler, and Marcus Specht. 2021. Literature Review on Co-Located Collaboration Modeling Using Multimodal Learning Analytics—Can We Go the Whole Nine Yards? *IEEE Transactions on Learning Technologies* 14, 3 (2021), 367–385.

[25] Sambit Praharaj, Maren Scheffel, Marcel Schmitz, Marcus Specht, and Hendrik Drachsler. May 2021. Towards automatic collaboration analytics for group speech data using learning analytics. *Sensors* 21, 9 (May 2021), 3156.

[26] Marcel Schmitz, Maren Scheffel, Roger Bemelmans, and Hendrik Drachsler. 2019. Fellowship Of The Learning Activity - Learning Analytics 4 Learning Design Game. https://doi.org/10.25385/zuyd.9884279

[27] Bertrand Schneider and Roy Pea. 2014. The Effect of Mutual Gaze Perception on Students' Verbal Coordination.. In *EDM*. 138–144.

[28] Bertrand Schneider and Roy Pea. 2015. Does seeing one another's gaze affect group dialogue? A computational approach. *Journal of Learning Analytics* 2, 2 (2015), 107–133.

[29] Daniel Spikol, Emanuele Ruffaldi, Lorenzo Landolfi, and Mutlu Cukurova. 2017. Estimation of success in collaborative learning based on multimodal learning analytics features. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 269–273.

[30] Stephanie Teasley, Frank Fischer, Pierre Dillenbourg, Manu Kapur, Michelene Chi, Armin Weinberger, and Karsten Stegmann. 2008. Cognitive convergence in collaborative learning. (2008).

[31] Jacques Terken and Janienke Sturm. 2010. Multimodal support for social dynamics in co-located meetings. *Personal and Ubiquitous Computing* 14, 8 (2010), 703–714.

[32] Kui Xie, Gennaro Di Tosto, Lin Lu, and Young Suk Cho. 2018. Detecting leadership in peer-moderated online collaborative learning through text mining and social network analysis. *The Internet and Higher Education* 38 (2018), 9–17.

[33] Jianlong Zhou, Kevin Hang, Sharon Oviatt, Kun Yu, and Fang Chen. 2014. Combining empirical and machine learning techniques to predict math expertise using pen signal features. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*. ACM, 29–36.