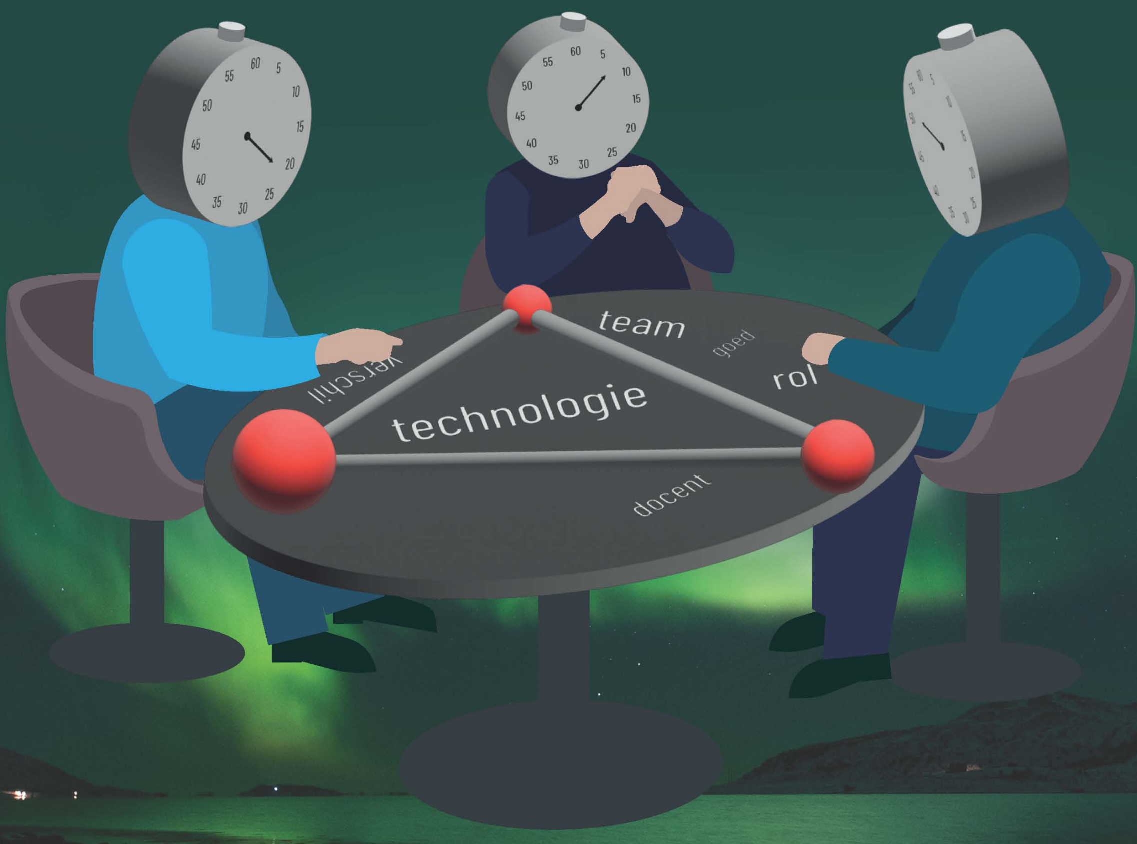
Measuring the Unmeasurable?

Towards Automatic Co-located Collaboration Analytics



Sambit Praharaj

Measuring the Unmeasurable?
Towards Automatic Co-located Collaboration Analytics

The research reported in this thesis was carried out at the Open University of the Netherlands in the Faculty of Educational Sciences, formerly known as Welten Institute – Research Centre for Learning, Teaching and Technology,

Open Universiteit

and under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

SIKS Dissertation Series No. 2022-07

Printed by ProefschriftMaken
Cover idea: Sambit Praharaj
Cover design: Fenna Schaap
Typeset in LaTeX

ISBN/EAN: 978-94-93211-43-8

Measuring the Unmeasurable?
Towards Automatic Co-located Collaboration Analytics

Proefschrift

ter verkrijging van de graad van doctor
aan de Open Universiteit
op gezag van de rector magnificus
prof. dr. Th.J. Bastiaens
ten overstaan van een door het
College voor promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 11 maart 2022 te Heerlen
om 13:30 uur precies

door

Sambit Praharaj
geboren op 09 juli 1993 te Rourkela, India

Promotores

Prof. dr. H.J. Drachsler
Open Universiteit / Deutsches Institut für Internationale Pädagogische Forschung - Leibniz Institute for Research and Information in Education / Goethe-Universität Frankfurt

Prof. dr. M.M. Specht
Technische Universiteit Delft

Copromotor

Prof. dr. M. Scheffel
Ruhr-Universität Bochum

Leden beoordelingscommissie

Prof. dr. J.T. Jeuring
Universiteit Utrecht

Prof. dr. A. Weinberger
Saarland University

Prof. dr. R. Klemke
Open Universiteit / Cologne Game Lab

Prof. dr. C.J. Kreijns
Open Universiteit

PD dr. R. Klamma
RWTH Aachen University

The greatest danger for most of us is not that our aim is too high and we miss it, but that it is too low and we reach it.

Michelangelo

Contents

General Introduction

> *The requirement of successful collaboration is complex, multimodal, subtle, and learned over a lifetime.*
>
> Stahl *et al.*, CSCL, (2013)

According to the New Learning Paradigm, students need to be taught the 4Cs super skills that are most in demand in the 21st century (Kivunja, 2015). Collaboration is one of these 4Cs super skills along with critical thinking, communication and creativity. How do we define collaboration? Collaboration occurs when two or more people work towards a common goal (Dillenbourg, 1999). Collaboration can take place either in an online setting or in a co-located (or face-to-face) setting. The measurement of online collaboration processes is possible due to the measurement, collection and analysis of the learner data using learning analytics (Siemens, 2011; Greller and Drachsler, 2012). Lately, with the ubiquitous use of sensors, a new branch of learning analytics that is coined as multimodal learning analytics (MMLA) has risen to popularity (Di Mitri et al., 2018a; Martinez-Maldonado et al., 2017a). Furthermore, sensor technology has become more scalable (Reilly et al., 2018), more affordable and more reliable in the past decade (Starr et al., 2018). With the help of MMLA, focus has shifted to the analysis of co-located collaboration (CC).

"CC takes place in *physical* spaces where the group members share each other's *social* and *epistemic* space" (Praharaj, 2019, p. 1, emphasis added). The social space consist of the non-verbal indicators of collaboration (such as non-verbal indicators from audio like total speaking time, turn taking and non-verbal indicators from video like gesture, posture and eye gaze) and the epistemic space consist of the verbal indicators of collaboration (such as the content of the conversations obtained from the audio) (Praharaj et al., 2018b). "The requirement of successful collaboration is *complex, multimodal, subtle,* and learned over a lifetime. It involves *discourse, gesture, gaze, cognition, social skills, tacit practices,* etc." (Stahl et al., 2013, pp.1–2, emphasis added). The indicators of collaboration help to detect the quality of collaboration.

But, how can quality of collaboration be described and detected by learning analytics indicators? The indicators are the fundamental units for collaboration quality detection obtained after processing and aggregating the sensors' data. One or more indicators can be combined to form the *indexes* which act as measurable markers for collaboration quality. For example, during collaborative brainstorming (Tausch et al., 2014), equality (i.e., the index) of the number of ideas (i.e., the indicator) generated by the group members can denote the quality of collaboration. The group with higher equality can be seen as having a better quality of collaboration because there are less dominating members in the group. Furthermore, depending on the *scenario* and context of CC, the indicators of collaboration vary. For instance, during collaborative programming, certain gestures like grabbing the mouse from the partner and synchrony in body posture are relevant indicators of the collaboration quality; whereas in collaborative brainstorming, the number of ideas generated by each group member is an indicator of the quality of collaboration. This difference in the set of indicators can be attributed to the CC task-based goals and the *parameters*. The parameters of CC are primary characteristics such as team composition (e.g., initiators, role of being initiators or experts), behaviour during collaboration (e.g., reflection, misconception, coherence), behaviour of team members (e.g., dominance, rapport, conflict) and types of interaction (e.g., passive or active).

Audio is the dominantly used modality in most of the past studies (Bassiou et al., 2016; Lubold and Pon-Barry, 2014; Bachour et al., 2010; Luz, 2013) while detecting the quality of CC. One possible reason could be the ease to capture audio data from the microphones. Most of these studies put emphasis on the social space (i.e., non-verbal audio indicators of collaboration such as total speaking time (Bergstrom and Karahalios, 2007; Bachour et al., 2010), frequency of turn taking (Kim et al., 2015), pitch and rhythm (Lubold and Pon-Barry, 2014; Bassiou et al., 2016)). The non-verbal audio indicators do not convey the true meaning of CC quality because most are using black box machine learning methods and some studies report the indicators (e.g., silence is an indicator for collaboration quality Luz (2013)) without informing about the valence, i.e., how good or bad are these indicators? Very few studies focus on the epistemic space (i.e., the verbal audio indicators or the content of the conversations) while computing the CC quality. Therefore, the emphasis is more on the "how" of the conversations instead of the "what" of the conversations. The "what" of the conversations are more overt as compared to the "how" of the conversations to detect CC quality. For example, higher or lower total speaking time may be a good or bad indicator of collaboration quality while "yes" or "no" will most of the time convey the same semantic meaning in an conversation.

Most of these studies focusing on the epistemic space use a manual approach to detect the CC quality (Jeong and Chi, 2007; Teasley et al., 2008) in a controlled setting. These approaches are laborious and time intensive. Very few use a semi-automatic approach to understand the epistemic space (Huber et al., 2019; Chandrasegaran et al., 2019). These studies are too abstract in either choosing representative keyword clusters (as topics) or classifying dialogues into few selected categories which do not affect the collaboration quality. They do not show the linkage of the conversations between different group members to understand the richness of the conversations.

Therefore, to overcome these limitations, we design a technical set up, conduct experiments with it in real-world authentic settings with the help of field trials. These experiments are conducted in a university setting during a CC task of playing a board game (to design a learning activity) with university staff members across 14 different sessions. The aim of designing this set up is to move towards automatic collaboration analytics and measure the quality of collaboration. Here the group members had pre-assigned roles before starting the board game. We do a holistic analysis of both the social and epistemic space with an emphasis on the inter-linkage of the content of the conversations between different roles to understand the role-role exchanges. This is visualized on a dashboard with the help of the network graph which shows both "what" each role spoke with each other and "how" (i.e., total speaking time and turn taking) they spoke with each other. To quantify the shared understanding (to detect the CC quality) based on the knowledge convergence measure (Jeong and Chi, 2007; Teasley et al., 2008), we also measure the shared epistemic space between the roles with the help of a frequently used keyword analysis and how it varies temporally.

So, the main objectives of the thesis are: (1) to define the constituents helping to detect CC quality (i.e., indicators, indexes and parameters in this case); (2) to design a set up for automatic CC analytics; (3) to move towards quantifying the quality of collaboration by using the CC analytics set up and show the visualizations on a dashboard.

Outline of the thesis

The thesis is structured into three parts that describe the theory to define and understand CC quality and analytics, prototyping a set up using that theory and a field study to detect and visualize collaboration analytics to move towards quantifying the quality of CC. **Part I** (consisting of Chapter 1 and 2) gives a definition of CC quality, analytics and helps to understand it. It describes the indicators of collaboration quality, the high level indexes (composed of one or more of the indicators) which act as the measurable markers of collaboration quality, contextualizing of the indicators based on the scenarios of collaboration and the feedback mechanisms to support collaboration. **Part II** (consisting of Chapter 3) describes the prototyping of an automatic CC analytics set up using the CC quality definition. Here, we develop an architecture for data collection, processing, analysis, visualizations and then test it based on a field study of a specific CC task. We primarily focus on the audio-based indicators of CC. In **Part III** (consisting of Chapter 4) we use the set up built in the second part to visualize CC analytics based on the content of the conversations (or audio-based indicators of collaboration) with an aim to move towards CC quality detection. The content is obtained from the audio recordings during a CC task across 14 different sessions in an university setting.

In **Chapter 1** we do an exploratory scanning of the CC landscape to understand the state-of-the-art studies on the indicators to detect the quality of collaboration. Then we look into the studies on feedback during CC which take the help of these indicators to facilitate collaboration. For example, during collaborative meeting, total speaking time was used as an indicator of collaboration quality to give real-time feedback to facilitate collaboration. The real-time feedback was shown as a reflection of the total speaking time by glowing the required number of LED lights (i.e., proportional to the total speaking time) in front of that group member on a smart table. This helped to reduce the participation (i.e., total speaking time in this case) of the over participators and increased the participation of the under participators; thereby improving the quality of collaboration.

So, based on these indicator, feedback examples from several studies, we designed a quick hybrid set up (with the combination of humans and sensors) to test real-time feedback with the help of a small field study. Here, we used group PhD meetings to track audio-based indicators (such as total speaking time and turn taking) using microphones and human observers. Using these indicators, we designed a real-time feedback on a public shared display to show as a reflection to the group members. The aim of this feedback set up was to get a feel of it instead of testing the efficacy of the feedback on collaboration quality.

After the exploration, we do an in-depth literature review of the indicators of collaboration quality in **Chapter 2**. It is because the indicators of collaboration vary depending on the scenarios and the context of collaboration. Thus, a diverse set of indicators of collaboration can help in the detection of quality of CC depending on different scenarios. The quality of CC can be measured by a event-process framework made up of the indicators and indexes. The indicators are low-level events obtained after processing and aggregation from the sensors. The indexes (i.e., high level processes) which act as the measurable markers help in detecting the quality of collaboration are made up of one or more indicators. For instance, in collaborative meetings (i.e., the scenario of CC), the equality (i.e., the index) of the total speaking time (i.e., the indicator) measures the quality of collaboration. If all group members have similar total speaking time with no one dominating the conversation then there is higher equality of total speaking time for the group and better quality of collaboration.

These indicators also vary across different scenarios because of the differing goals and parameters (i.e., primary aspects such as team composition, behaviour of team members and behaviour during collaboration) of CC. For example, indicators of CC quality for collaborative programming can differ completely from indicators of collaborative brainstorming. Therefore, with the help of this literature review, we define a conceptual model that encompasses the indicators, indexes and parameters to detect the CC quality. In this model, we map the parameters in different scenarios onto the indicators and indexes to support the design of a CC quality detection and prediction system.

Using this model, we zoomed in on audio-based indicators of collaboration in **Chapter 3**. Audio is the dominantly used modality as found in the literature review and it is also easy to capture with microphones. We found that most of the prior works focused on the "how" of the conversations and not on the "what" of the conversations while detecting CC quality. Thus, the focus was on the social space (comprising of non-verbal audio indicators such as total speaking time, change in pitch) and not that much on the epistemic space (comprising of the content of the conversations) of the audio modality. A handful of works focused on the epistemic space semi-automatically in lab setting and some coded the epistemic space manually. These approaches were based on predefined conditions, gave abstract overview of the topics of discussion and laborious to implement. So, we build a prototype to overcome this and analyze the richness of the epistemic space in an authentic real world setting with the help of field trials.

For this, we record the audio conversations during a CC task where university staff play a board game with pre-assigned roles to create awareness of the connection between learning analytics and learning design. We transcribe these audio recordings (i.e., convert from speech to text), process them and then visualize them (using network graphs to understand the interconnected nature of the spoken text) by doing a role-based profiling to get a holistic overview of the conversations in an automatic manner. We test this prototype for one CC session and also discuss the limitations in automation of the prototype. The purpose of building this prototype is to make a step towards automatic collaboration analytics.

Using the developed prototype, we move towards quantifying the quality of collaboration in **Chapter 4** with the help of field trials across 14 different CC sessions played in the context of the board game discussed earlier. We do a holistic analysis of the social (i.e., total speaking time and turn taking) and epistemic (i.e., the content of the conversations) space also considering the role based contributions and interactions and then visualize it. We define quality taking into account the convergence of the discussion (i.e., shared epistemic space as analyzed from the content of the conversations) among the group members with different roles. Finally, we visualize both the social and epistemic space using a dashboard; then discuss the stakeholders who can use such a dashboard and what future research can be done on that dashboard.

The thesis is concluded by a **General Discussion** where we summarize the findings of the studies included in the thesis and give recommendations for the community. Apart from that, the limitations are reviewed and practical implications, final thoughts are discussed.

Part I

Collaboration Indicators, Analytics and Feedback

Chapter 1

Multimodal Analytics for Real-time Feedback in Co-located Collaboration

Collaboration is an important 21st-century skill; it can take place in a remote or co-located setting. Co-located collaboration (CC) is a very complex process which involves subtle human interactions that can be described with multimodal indicators (MI) like gaze, speech and social skills. In this chapter, we first give an overview of related work that has identified indicators during CC. Then, we look into the state-of-the-art studies on feedback during CC which also make use of MI. Finally, we describe a Wizard of Oz (WOz) study where we design a privacy-preserving research prototype with the aim to facilitate real-time collaboration in-the-wild during three co-located group PhD meetings (of 3-7 members). Here, human observers stationed in another room act as a substitute for sensors to track different speech-based cues (like speaking time and turn taking); this drives a real-time visualization dashboard on a public shared display. The main purpose of this WOz study is to run a small field study to test real-time feedback during CC using a small set of MI.

This chapter is based on:

Praharaj, S., Scheffel, M., Drachsler, H., and Specht, M. (2018). Multimodal analytics for real-time feedback in co-located collaboration. In *European Conference on Technology Enhanced Learning* (pp. 187–201). Springer, Cham, doi: 10.1007/978-3-319-98572-5_15.

1.1 Introduction

Collaboration is an important skill in the 21st century (Dede, 2010). It can take place in different settings and for different purposes: collaborative meetings (Terken and Sturm, 2010; Kim et al., 2008; Stiefelhagen and Zhu, 2002), collaborative problem solving (Spikol et al., 2017b), collaborative project work (Cukurova et al., 2017a, 2018), collaborative programming (Grover et al., 2016) and collaborative brainstorming (Tausch et al., 2014). Some are in co-located and some in remote settings. "The requirement of successful collaboration is *complex, multimodal, subtle,* and learned over a lifetime. It involves *discourse, gesture, gaze, cognition, social skills, tacit practices,* etc." [emphasis added] (Stahl et al., 2013). Moreover, in each context, the indicators of collaboration vary. For instance, in collaborative programming pointing to the screen, grabbing the mouse from the partner and synchrony in body posture are relevant indicators for good collaboration (Grover et al., 2016); whereas in collaborative meetings gaze direction, body posture, speaking time of group members are more relevant indicators for good collaboration quality (Terken and Sturm, 2010; Kim et al., 2008; Stiefelhagen and Zhu, 2002). Thus, it is essential to understand what the different types of collaboration and their purpose are and what are the relevant indicators. These indicators help to formulate the intervention or feedback mechanism to facilitate collaboration (Bachour et al., 2010; Schneider et al., 2015; Bergstrom and Karahalios, 2007). Moreover, engaging in a collaborative task does not essentially build collaborative skills (Dillenbourg, 1999); rather on-time feedback encourages self-reflection (O'Donnell, 2006). The type of feedback is also dependent on the goal of the task which can be to evaluate collaboration as a process (Bachour et al., 2010) or collaboration as an outcome (indicated by learning gain) (Schneider et al., 2015) or both (Schneider et al., 2015). To understand this in-depth, we have formulated two research questions:

RQ 1: What *collaboration indicators* can be observed and are relevant for the *quality* of collaboration during CC?

RQ 2: What are the state-of-the-art *feedback* mechanisms that are used during CC?

There has been a dearth of studies on automated multimodal analysis in non-computer supported environments (Worsley and Blikstein, 2015). Considering the time and effort required to build a sensor-based automated system which can also give real-time feedback, we chose to create a WOz research prototype which can integrate human observers and existing sensor technology. This enables us to study different CC settings with a variety of multi-source multimodal indicators coming from automated sensors as well as human observers.

The remainder of the paper is structured as follows: in the related work (sec. 1.2) section we answer RQ 1 and RQ 2; it is followed by an explanation of our prototype design based on the WOz study (sec. 1.3); this is followed by a discussion (sec. 1.4) of the answers to our research questions; finally, a conclusion (sec. 1.5) is drawn and we throw some light on future work and open questions to be answered.

1.2 Related Work

In this section, we will first analyze related work according to the different indicators used during CC from multiple modalities; and secondly review the different feedback mechanisms used during CC.

1.2.1 Multimodal indicators during co-located collaboration

Different categories of verbal and non-verbal indicators have been used in the literature to measure collaboration quality ranging from tangible interaction, different speech-based cues, to gaze and eye interaction. Schneider and Blikstein (2015) used Tangible User Interface (TUI) for pairs of students to predict learning gains by analyzing data from multimodal learning environments. They tracked the gesture and posture using a Kinect Sensor[1] (Version 1) which can track the posture and gesture of a maximum of four students at a time based on their skeletal movements. They found that the *hand movements* and *posture movements* (coded as active, semi-active and passive) are correlated with learning gains. The more active a student is, the higher is the learning gain. Even the number of transitions between these three phases was a strong predictor of learning. Students who used both hands showed higher learning gains. Some of the activities that were logged by the TUI, like the frequency of opening the information box in the TUI can be correlated with learning gain. All these features were fed into a supervised machine learning framework to predict learning gain. Similarly, Martinez-Maldonado et al. (2015a) used TUI indicators for group work based on the log data generated and the gesture and posture of group members around the TUI.

Other works detected non-verbal cues during collaboration without a TUI. Stiefelhagen and Zhu (2002) tried to detect the impact of *head orientation* on the gaze direction in a group round table meeting with four members. They found that on an average 68.9 % of the time head orientation can estimate gaze direction. Moreover, attention focus of group members can be easily predicted 88.7 % of the time using head orientation as the only input. Similarly, Cukurova et al. (2017a) performed a experiment on 18 members in six groups of three members each to detect non-verbal cues of collaboration using human observation. *Hand position* (HP) and *head direction* (HD) was a good predictor of competencies in Collaborative Problem Solving (CPS). They extended this work and formed the NISPI framework (Cukurova et al., 2018) using HP and HD as non-verbal indicators. These indicators were obtained during a prototype design by students (11-20 years old) using the Arduino toolkit. Then, they were coded for each student as: 2 (*active*) if a student is interacting with the object for problem solving, 1 (*semi-active*) if the head of the student is directed towards an active peer and 0 (*passive*) for all other situations. Using this coding, different collaboration dimensions like *synchrony*, *individual accountability* (IA), *equality* and *intra-individual variability* (IIV) were formed. High competencies of CPS was detected if high levels of synchrony, IA and equality is detected in the groups.

[1] An integrated sensor tracking simultaneously infrared, depth, audio and video.

Speech-based cues are an integral part of any collaborative task. Lubold and Pon-Barry (2014) found that *proximity, convergence* and *synchrony* are different types of coordination cues obtained from the speech features (like intensity, pitch and jitter) of the pair of students collaborating. It helped them to detect rapport between group members. It was observed from correlation analysis that proximity, convergence and synchrony measured using pitch can be a good predictor of rapport between the group members during collaboration. Students also self-reported rapport which was compared and collaboration levels were determined. Bassiou et al. (2016) assessed collaboration among students solving math problems automatically. They used *non-lexical speech* features; thereby, preserving the privacy. They used a combination of manual annotation and Support Vector Machine (SVM) to predict the collaboration quality of the group. Types of collaboration marked are: Good (all 3 members are working together and contributing to the discussion), Cold (only two members are working together), Follow (one leader is not integrating the whole group) and Not (everyone is working independently). This coding was based on two types of engagement: simple (talking and paying attention) and intellectual (actively engaged in the conversation). They found that the combination of the speech-activity features (i.e., *solo duration*, *overlap duration of two persons*, *overlap duration of all three persons*) and speaker-based features (i.e., *spectral*, *temporal*, *prosodic* and *tonal* features of speech) are good predictors of collaboration. Simple indicators like the speaking time of each member can also be a good indicator of collaboration (Bachour et al., 2010; Bergstrom and Karahalios, 2007). Even a mixture of verbal and non-verbal indicators along-with *physiological signals* like skin temperature (Pijeira-Díaz et al., 2018) can be a good collaboration indicator (Madan et al., 2004; Kulyk et al., 2005).

Besides, *eye gaze* can be an indicator of collaboration quality. Some researchers (Richardson and Dale, 2005; Jermann et al., 2011; Schneider and Pea, 2013) while using eye gaze analysis found that (JVA) *Joint Visual Attention* (i.e., the proportion of times gazes of individuals are aligned by focusing on the same area in the shared object or screen) is a good predictor of the quality of collaboration of a group which is reflected by the group's performance. Moreover, Schneider and Pea (2013) showed that JVA can be used as a reflection mechanism in remote settings to show each student their partner's gaze patterns in real-time to improve collaboration. Schneider et al. (2015) got the same results by replicating the experiment in a co-located setting. The work by Schneider and Pea (2014b) used JVA, network analysis and machine learning to determine different dimensions of a good collaboration like *mutual understanding*, *dialogue management*, *division of task*, *signs of coordination* as outlined by Meier et al. (2007).

Moving on to the different purposes in which collaboration has been studied, Spikol et al. (2017b, 2018a) studied collaborative learning specifically in the context of Collaborative Problem Solving (CPS). They tracked the distance between *hand movements* and *faces* of group members. Later the recorded video streams were coded by experts with 0 (for passive), 1 (for semi-active) and 2 (for active) based on different combinations of head and hand positions for training the machine learning classifier for predicting the quality of collaboration. Recent work by Chikersal et al. (2017) dives deep into the deep structure of collaboration in dyads. They found that synchrony in facial expressions correlated with collective intelligence of the group but not significantly correlated with the synchrony of electrodermal activity of members. Another work by Grover et al. (2016) studied CPS in a pair programming context based on a pilot study. They captured data from different modalities (i.e., video, audio, clickstream and screen capture) unobtrusively using Kinect. For initial training of the classifiers using machine learning, experts coded the video recordings with three annotations (i.e., High, Medium and Low) when they found evidences of collaboration between the dyads. These evidences include *pointing to the screen*, *grabbing the mouse* from the partner and *synchrony in body position*. Later this classifier could predict the level of collaboration.

Moreover, post-hoc coding with the help of human coders has been a effective method followed for a long time to detect different indicators of collaboration. Davidsen and Ryberg (2017) videotaped the work of pairs making a collaborative discussion around a touch screen measuring "The size of one meter". The pair was trying to translate the design from graph paper to the touch screen to measure one meter. They found that *body movements, language* and *gestures* can be helpful to discover different facets of collaboration. Similarly, Scherr and Hammer (2009) observed videotaped groups and identified four clusters based on the collaborative behaviour from both verbal and non-verbal indicators (like *eye contact with peers*, *straight posture*, *clear and loud voice*, etc.). Besides, some works (Shih et al., 2009; Tausch et al., 2014) considered *epistemological* aspects of collaboration during brainstorming where the number of ideas generated by each member was the indicator of quality of collaboration. Detecting individual attention levels in classroom from the responses to questions (i.e. epistemological) is also common (Triglianos et al., 2017).

In summary, collaboration indicators can vary from non-verbal, verbal, physiological to log files obtained from shared objects like TUI or computers. It depends on the context. Table 1.1 shows the overview of the multimodal indicators detected. We can find two types of co-located collaboration indicators, i.e., *social* (verbal, non-verbal and physiological) and *epistemological* (logs, ideas).

1.2.2 Feedback during co-located collaboration

Using these multimodal indicators, different feedback mechanisms have been developed in the past to facilitate CC. Kulyk et al. (2005) designed a mechanism to give real-time feedback to participants in group meetings (with 4 members) by analyzing their speaking time and gaze behaviour. The feedback was in the form of different coloured circles representing attention from other speakers measured by eye gaze, speaking time and attention from listeners. This feedback was projected on the table in-front of where each participant was sitting using a top-down projector. They performed both quantitative and qualitative evaluation to evaluate the effect of the feedback: the feedback was accepted as a positive measure by most group members; use of feedback had a positive impact on the behaviour of group members as they

Table 1.1 Overview of studies on co-located collaboration.

References	Indicators	Goal
Schneider and Blikstein (2015)	Hand movements, posture & TUI logs	Post-hoc analysis of indicators on learning
Schneider et al. (2015)	Joint Visual Attention (JVA)	JVA indicates learning
Spikol et al. (2017b)	Distance between hands & faces	Extraction of multimodal features during collaboration
Grover et al. (2016)	Pointing, body position & grabbing mouse	Post-hoc classification of collaboration
Bachour et al. (2010)	Total speaking time	LED display to regulate audio participation in real-time
Tausch et al. (2014)	Number of ideas	Real-time metaphorical feedback to support CB
Bergstrom and Karahalios (2007)	Total speaking time	Conversation clock will regulate the equity of conversation in real-time
Cukurova et al. (2018)	Hand position and head direction	Build a non-verbal indicator framework for collaboration
Lubold and Pon-Barry (2014)	Intensity & pitch of sound, self reports	Detect collaboration levels based on rapport obtained from audio cues & self-reports
Bassiou et al. (2016)	Speech overlap duration, no overlap duration, spectral, temporal, prosodic & tonal speech features	Predict collaboration quality from audio cues
Davidsen and Ryberg (2017)	Dialogue, gesture, posture & gaze	Detect indicators of collaboration from videotaped recordings of collaboration tasks
Scherr and Hammer (2009)	Eye contact, posture & amplitude of voice	Detect indicators of collaboration from videotaped recordings of collaboration tasks

had a balanced participation and improved eye gaze. Terken and Sturm (2010) used a similar setting and feedback mechanism; they discovered that the feedback on speech increased the equity of participation in the group. But, surprisingly feedback on gaze behaviour had little effect on the interaction pattern of group members. Similarly, Madan et al. (2004) used sensors to capture nodding, speech features and galvanic skin response of dyads and built a real-time group interest index. This group interest index helped them to drive a real-time feedback. This feedback showed some group characteristics in different modes: individual PDA feedback, personal audio feedback, haptic feedback in the shoulder and public shared projected display. They studied these group characteristics in different contexts like speed dating and brainstorming sessions.

Some simpler versions of feedback which leverage the audio cues (like speaking time) during collaboration have proved effective in the past. For instance, Bachour et al. (2010) performed an experiment to measure audio participation where each group (with 3-4 members) performed a task around a smart table. It gave them real-time feedback during the task by glowing different coloured LED lights for each member. The number of LED lights that glowed for each colour denoted the total speaking time for that member. They found that a real-time feedback helped to maintain the equity of audio participation among the members. Another similar approach was used by Bergstrom and Karahalios (2007) with the help of a conversation clock. In this clock, different coloured concentric rings represented spoken participation of each member in the 4 member group. The bars and the dots in the ring denoted the length of conversation and periods of silence respectively.

Moving on to the epistemological aspect of collaboration, Tausch et al. (2014) used an intuitive metaphorical feedback moderated by human observers during *collaborative brainstorming*. Three members in each group performed the task. The group members were supposed to discuss a certain topic and their collaboration was measured by the number of ideas generated. A comparison metric for collaboration such as a baseline was calculated as the average number of ideas generated by all members. Using this baseline, each group member was marked as below average or above average depending on the number of ideas generated by each member. Then the human observers controlled the public shared display which showed a *metaphorical garden*. Each group member was represented by a flower and the group was displayed as a tree with leaves, flower and fruit. The growth of the flower and the tree symbolized the participation (measured by the contribution of ideas) of the individual and the group respectively. More balanced participation was shown by a well grown tree with leaves, fruits and flower. If a group was having unbalanced participation for a long time then lightning flashes were shown in the group garden. Another example of feedback during collaborative brainstorming was implemented by Shih et al. (2009). It supports collaborative conceptual mapping to discuss a topic and organize the ideas.

Besides the use of visual and haptic feedback was effective in some collaboration tasks around a TUI. Anastasiou and Ras (2017) gave real-time textual and haptic feedback to each group consisting of 3 members working around a TUI. The group members were needed to use different objects and find the desired power consumption using the TUI. At the end, they used a questionnaire and found that most participants of the experiment favoured the use of both visual and haptic feedback over audio feedback. Martinez-Maldonado et al. (2015a) used a TUI and gave real-time feedback on group performance for the teachers in tablets so that they can intervene when needed and can also make a post-hoc reflection after the task is over.

Use of external sensing devices to facilitate collaboration during meetings has proved its worth before. Kim et al. (2008) used a sociometric badge[2] which acted as a meeting mediator to capture audio and postures during meetings of 4 members in one group. This badge bridged the gap of dominance and increased the equity of participation among the group members using a real-time feedback on their personal mobile phones. This feedback showed a circle in the middle of a screen connected by four lines to small squares in each corner of the screen representing the individual group members. The colour and position of the circle denoted the interactivity of the group. When the group had a balanced participation then the circle was darker in colour and in the centre of the screen. The thickness of lines connecting the circle represented the speaking time of each group member. Apart from the personal mobile display to give feedback, Balaam et al. (2011) used an ambient display showing a coloured circle visualization based on the non-verbal indicator of synchrony during a collaborative task of calendar planning. DiMicco et al. (2004) used a shared group display to influence the speaking participation of each group member during a group activity.

In summary, most of these studies were in controlled conditions with small groups consisting of dyads and triads only. Table 1.2 shows the overview of feedback mechanisms used during co-located collaboration. Some real-time feedback mechanisms acted as a mere reflection for the group to self-regulate instead of an actionable feedback; while others used a post-hoc analysis for the teachers (or facilitators) to reflect on the group activity. The mode of display varied from a public display to smart phone display.

In a nutshell, most of the studies in *related work* are in controlled conditions and using specialized furniture, TUI and badges. These settings can be suitable for adhoc CC which can be difficult to adapt in a dynamic setting. They also do not cater to the privacy and fairness of individuals. Most of these studies employ human observers as post-hoc annotators for coding videos to detect traces of collaboration. To tackle these issues, we devise a human-based prototype where privacy, in-the-wild setting and dynamic design is at the centre of our WOz study.

[2]An electronic sensing device worn around the neck that can collect and analyze social dynamics

Table 1.2 Overview of studies on co-located collaboration feedback.

References	Indicators	Feedback
Kim et al. (2008)	Total speaking time & body posture	Graphical with coloured shape and lines using personal mobile screens
Tausch et al. (2014)	Number of ideas	Metaphorical as a group-garden using public shared display
Kulyk et al. (2005)	Speaking time & eye gaze	Graphical with coloured concentric circles using public table-top private projection
Terken and Sturm (2010)	Speaking time & eye gaze	Graphical with coloured concentric circles using public table-top private projection
Madan et al. (2004)	Nodding, speech features & galvanic skin response	Graphical group characteristics using audio, haptic, PDA and public shared display
Bachour et al. (2010)	Total speaking time	Coloured LED light using public shared table top LED display
Bergstrom and Karahalios (2007)	Total speaking time	Coloured concentric rings with lines and dots using public shared table top display
Balaam et al. (2011)	Pointing	Coloured circle visualization using ambient display
Martinez-Maldonado et al. (2015a)	Log data about different actions performed with the TUI	Pie chart and other statistical charts using private tablet for teachers
Anastasiou and Ras (2017)	Log data about content knowledge from TUI	Textual and haptic using public TUI display
DiMicco et al. (2004)	Total speaking time	Coloured bar charts using public shared display

1.3 A WOz study: Designing the research prototype

Based on our analysis, we aimed for creating a flexible research infrastructure that allows us to study feedback in CC making use of different indicators and combining them in different feedback instruments and media. We followed a design-based approach focusing on a specific type of meeting and evaluated different types of indicators, human-observer interfaces, as well as feedback mechanisms. The main components of our research prototype are a defined set of indicators and sensors, a user interface for CC observation managed by human observers, as well as a set of feedback components.

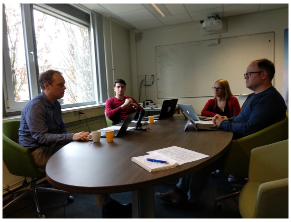


Figure 1.1 Meeting room

Figure 1.2 Annotator room

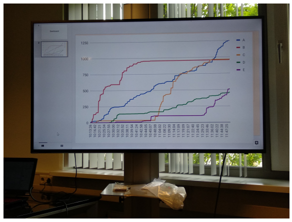


Figure 1.3 Public display

1.3.1 Experimental context

We performed the experiments during three PhD meetings with 3-7 members in each meeting in the room as shown in figure 1.1. Due to the frequent availability of these meetings and ease of not designing the task per se, we chose them. Our main focus was to execute the study *in-the-wild* and preserving the *privacy*. Thus, we used a human annotator who was present in the adjacent room separated by a one-sided transparent wall as shown in figure 1.2. Although it is difficult to see in the picture, the visibility through the wall from the side of the annotator was transparent; while the visibility from the meeting room was opaque. A microphone was used to listen to the conversation in the other room but audio was not recorded. The real-time feedback was shown on a big shared public display in the meeting room (as depicted in figure 1.3) which was managed by the annotator. The real-time feedback visualization could make use of observation data from the human observer and also visualize raw-data, e.g. the audio volume of the group work. The collaborators got a virtual sense of being tracked by a microphone automatically when they saw the changing real-time feedback of their speaking participation on the screen.

1.3.2 Data logging

For the sake of clarity in data logging, we have segregated the multimodal channel annotation into verbal and non-verbal (i.e., gestures and postures) channels and identified different non-verbal indicators as: looking at laptop or peers; looking down; looking at the feedback; typing with laptop; and making different hand gestures. The verbal indicators are: occurrence, pauses, overlaps, interruptions in speech; affirmatives in speech; and asking questions. But, to ease the logging process for the human annotator, we chose to only focus on the simpler observable audio cues which is the speaking time and turn taking of each group member in a first study. The speech-based cues are ubiquitous in any collaboration and non-verbal cues may be difficult to monitor for one annotator in a large group setting. The annotator was seeing the annotation interface embedded in a Google sheet as shown in figure 1.4. To preserve the *privacy*, we gave the annotator a coding sheet where each collaborating member was given an alias name from the English alphabet. Moreover, each participating member signed a consent form. Whenever a person starts speaking, the annotator pressed the corresponding button in the interface which automatically creates a cell in the Google sheet with the start time and name

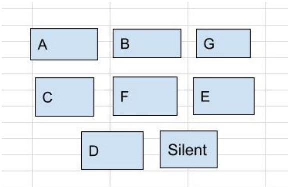


Figure 1.4 Annotation interface

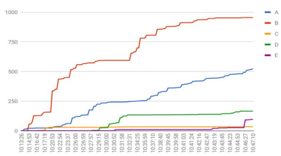


Figure 1.5 Mid feedback

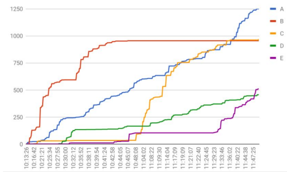


Figure 1.6 End feedback

of that person. Whenever the annotator presses another person's button, the end time of the previous person is registered in the sheet. This was possible as the buttons were coupled with a JavaScript to perform the operation. To ensure the reliability of the coding scheme, we had a provision to include multiple annotators but did not use it for our experiments as it involved only simple clicking of a button.

1.3.3 Modeling participation during collaboration

The sheet interface was connected to a chart embedded in Google Slides which was updated in real-time when a value is entered by pressing a button. The other columns in Google sheet were automatically populated based on the defined formula which calculates the cumulative speaking time of each member from the beginning of the meeting. Figure 1.5 shows the group dynamics after the first 30 minutes during a meeting using a line chart as displayed during the meeting on the big public shared display in the room. The times shown on the horizontal axis is the plot time obtained from the end time of speaking of a member. The value in vertical axis is the total speaking time in seconds from the beginning of the meeting. Figure 1.6 shows the status of the line chart at the end of the meeting. Here, the speaking time and turn taking represented the participation of each group member. We also collected oral feedback from both the annotators and the collaborators during the iterative design phase.

1.3.4 Results

From our first three iterations in the PhD meetings, we developed a first prototype for analyzing turn-taking and speaking time feedback. Our results showed that we need a higher level annotation interface. Thus, we supported human observers in that they only need to press a button when a new person starts talking. For the visualization on the public shared display, we experimented with different visualizations of the speaking time. Based on participants' feedback we altered the display format from an original pie chart to a line chart for displaying the development of the conversation over time. An example of the feedback at different times of a meeting can be seen in figure 1.5 and 1.6. We can observe that speaker B, who is a second year PhD student, dominates the conversation in the first 30 minutes; it was his turn to speak regarding his PhD project at that time. But, from that time on-wards he stops to participate in the meeting; indicated by the line parallel to horizontal axis in figure 1.6. We can also observe at the end of the meeting that speaker A, who is the promotor, has spoken the most and changed turns very often to intervene during the meeting; the turn-taking was evident from the frequent change of the shape of the line indicated by small or large spikes.

1.4 Discussion

RQ1: On the multimodal indicators during CC indicating collaboration quality — Based on the literature study, we discovered different multimodal indicators during CC in multiple contexts. They can be grouped into *social* (i.e., verbal, non-verbal and physiological) and *epistemological* (i.e., ideas and data logs) indicators. For detecting the social indicators, sensors have been used in past works. But, for detecting the epistemological indicators human help was required as it is difficult for sensors to automatically detect the number of ideas generated from speech by understanding the semantics.

RQ2: On the feedback during CC — Feedback during CC is either real-time (for reflection or guiding) or post-hoc (for the purpose of reflection). This brings into the picture two *stakeholders*: the teachers (or facilitators) and the group members. We need this distinction as it will help in designing the feedback. Some works used TUI and other electronic mediums like Interactive White Boards (IWB) and tablets during collaboration which requires a lot of preparation before a collaborative task. Therefore, it is difficult to use it in real-world dynamic settings. Besides, there is a trade-off between personalization for the group and privacy. More personalized feedback meant for the whole group is less privacy preserving. Thus, there should be a decision on the level (i.e., group, individual or both) of feedback to be shown depending on the circumstances at hand.

On the research prototype to give real-time feedback — We take a step in building an initial prototype design with the aim to facilitate real-time collaboration during meetings. We were successful in building a click-based interface for the annotator which also reduces memory overhead. This helps us to create a hybrid set up without building an actual automated sensor-based system to experiment with different types of real-time feedback mechanisms during CC. We can later use these insights to build the sensor-based or hybrid set up. Here, we can build individual components in a modular fashion to track other indicators of collaboration quality; and integrate them to a single dashboard.

1.5 Conclusions & Future Work

Collaboration being an important skill and ubiquitously present in our day to day activities, we try to look into the different collaboration indicators in various contexts in the literature. We find different types of indicators like gaze, speaking time, posture, gesture, number of ideas generated, etc. Then we look into the impact of feedback during collaboration and find that visual real-time feedback has some impact on the collaboration like improving the equity of audio participation. This feedback can range from private displays (like PDA, mobile phones) to a more public one (like TUI, shared display).

Based on this overview, we took a step further and built a real-time feedback prototype during collaboration based on a privacy-preserving WOz study in-the-wild. Here, we study collaboration during co-located PhD meetings using human observers acting as a proxy for sensors. We find that the human observers could easily track 'who spoke when and for how much time' by pressing a button.

As future work suggestions, we need to define the *goal* and *outcome* of the collaboration task and make it clear in the evaluation criteria as to whether we measure collaboration as a process, outcome or both. Then, we can focus on the feedback mechanisms for facilitating collaboration. We can also borrow some insights from the mapping of multimodal data to feedback in an individual learning context (Di Mitri et al., 2018a). The feedback can be: human based, sensor based or a hybrid of both. We need to decide the type (number of pointing gestures, speaking time, number of interruptions, number of eye contact with peers, etc.), modelling (i.e., individual, group or both) and display of feedback (i.e., personal, public or both) based on *action-based research* (Dyckhoff, 2014) where we need to take the preliminary feedback of different stakeholders like teachers (or facilitators) and the group members. Our long term goal is to do action-based research and build a sensor-based automated (or hybrid) feedback system during CC using the currently built research prototype. Here, we can include different feedback components to identify multiple indicators of collaboration and proceed towards an automated system using deep neural networks to integrate data from multiple sensors (Schneider et al., 2018).

Chapter 2

Literature Review on Co-Located Collaboration Modeling Using Multimodal Learning Analytics—Can We Go the Whole Nine Yards?

Co-located collaboration (CC) is a very complex process that involves subtle human interactions that can be described with indicators like eye gaze, speaking time, pitch, and social skills from different modalities. Continuing from Chapter 1, we do an in-depth review of the indicators of collaboration to understand how quality of collaboration is detected and measured. With the advent of sensors, multimodal learning analytics has gained momentum to detect CC quality. Indicators (or low-level events) can be used to detect CC quality with the help of measurable markers (i.e., indexes composed of one or more indicators) which give the high-level collaboration process definition. However, this understanding is incomplete without considering the scenarios (such as problem solving or meetings) of CC. The scenario of CC affects the set of indicators considered: for instance, in collaborative programming, grabbing the mouse from the partner is an indicator of collaboration; whereas in collaborative meetings, eye gaze, and audio level are indicators of collaboration. This can be a result of the differing goals and fundamental parameters (such as group behavior, interaction, or composition) in each scenario. In this review, we present our work on profiles of indicators on the basis of a scenario-driven prioritization, the parameters in different CC scenarios are mapped onto the indicators and the available indexes. This defines the conceptual model to support the design of a CC quality detection and prediction system.

This chapter is based on:

Praharaj, S., Scheffel, M., Drachsler, H., and Specht, M., "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), pp. 367–385, 1 June 2021, doi: 10.1109/TLT.2021.3097766.

2.1 Introduction

Collaboration is often mentioned as one of the important 21st-century skills (Dede, 2010) and a part of the 4Cs skill set (Kivunja, 2015) (along with critical thinking, communication, and creativity). When two or more persons work towards a common goal then collaboration occurs (Dillenbourg, 1999). Most of the works in the field of learning analytics support for collaboration have focused on the analysis of distributed (or online) collaboration (Jeong and Hmelo-Silver, 2010). However, with the pervasive use of sensors (Grover et al., 2016; Kim et al., 2008), multimodal learning analytics (MMLA) (Blikstein, 2013; Praharaj et al., 2018a; Di Mitri et al., 2018b) has picked up the pace, thus shifting the focus to the analysis of co-located collaboration (CC) (or face-to-face collaboration) with the help of sensor technology (Grover et al., 2016; Kim et al., 2008; Praharaj et al., 2021b; Tausch et al., 2014). Furthermore, sensor technology can be easily scaled up (Reilly et al., 2018) and has become affordable and reliable in the past decade (Starr et al., 2018). CC takes place in physical spaces where all group members share each other's social and epistemic space (Praharaj, 2019). "The requirement of successful collaboration is *complex, multimodal, subtle,* and learned over a lifetime. It involves *discourse, gesture, gaze, cognition, social skills, tacit practices,* etc." (Stahl et al., 2013, pp.1–2, emphasis added). According to Johnson and Johnson (2009), positive interdependence, individual accountability, promotive interaction, the appropriate use of social skills, and group processing are five variables that mediate the effectiveness of collaboration. Similarly, Meier et al. (2007) identified five aspects of collaborative process and nine dimensions of rating collaboration quality: communication (sustaining mutual understanding, dialogue management), joint information processing (information pooling, reaching consensus), coordination (task division, time management, technical coordination), interpersonal relationship (reciprocal interaction), motivation (individual task orientation). The five aspects of collaboration quality from these two works (Johnson and Johnson, 2009; Meier et al., 2007) can be matched onto each other in the following way: 1) communication/appropriate use of social skills; 2) joint information processing/group processing; 3) coordination/positive interdependence; 4) interpersonal relationship/promotive interaction; and 5) motivation/individual accountability. But, the work by Meier et al. (2007) elaborates into fine-grained subcomponents of these five aspects. Successful collaboration also depends on the focus of the assessment of collaboration (i.e., whether collaboration is assessed as a process or as an outcome (Child and Shaw, 2015)).

Quality of CC can be detected by different indicators of collaboration such as total speaking time (Bachour et al., 2010) or eye gaze (Schneider et al., 2015). These indicators can be processed and grouped together to different indexes which act as the measurable markers of CC quality. For instance, the quality of collaboration within a group can be good if there is higher equality (i.e., the index) of total speaking time (i.e., the indicator) among the group members (Bachour et al., 2010). Moreover, different scenarios of CC such as collaborative programming (Grover et al., 2016), collaborative meetings (Kim et al., 2008; Terken and Sturm, 2010), or collaborative brainstorming (Tausch et al., 2014) each has a different set of indicators denoting the quality of collaboration. For instance, in collaborative programming relevant indicators of collaboration include pointing to the screen, grabbing the mouse from the partner, and synchrony in body posture (Grover et al., 2016); whereas in collaborative meetings gaze direction, body posture, or speaking time of group members are more relevant indicators for collaboration quality (Terken and Sturm, 2010; Kim et al., 2008; Stiefelhagen and Zhu, 2002). This difference can be attributed to the goals of the tasks performed during CC and the structuring of the task (Collazos et al., 2007, 2003). In addition, the fundamental parameters of CC like team composition (such as experts or initiators), the behavior of team members (such as dominance or rapport) vary from group to group. For example, a group with fewer dominant members during CC shows a better quality of collaboration (Kim et al., 2008). Therefore, in order to get a holistic view, a scenario-driven prioritization and a mapping of the parameters of CC onto the indicators need to be done. So, the definition of collaboration and its quality varies across different research fields. It is dependent on the focus of assessment, goals, fundamental parameters (such as team composition and team behavior), the scenario in which collaboration is studied, and the way it has been operationalized in different research fields.

Furthermore, such indicators are complex interactions. These indicators cannot be detected as easily as the interactions from online data logs (or chat logs) generated during the distributed (or remote) collaboration. Thus, to understand collaboration dynamics during CC, a preliminary analysis needs to be performed to identify indicators relevant for the quality of collaboration. According to Dillenbourg et al. (1996); Dillenbourg and Traum (2006), we are in the third stage of research on collaboration (after proving the effectiveness of collaboration in the first stage and finding the conditions that predict the effects of collaboration in the second stage). In the third stage, the primary goal is to understand the interactions that take place during collaboration. To this end, the following research questions need to be answered with the help of a literature review:

RQ 1: What *collaboration indicators* have been used in research to understand the *quality* of CC?

RQ 2: What is the impact of different *scenario-based goals* and *parameters* for CC on the relevance of the different indicators?

The remainder of the paper is structured as follows: in the methodology section (Section 2.2) we describe the approach taken for this review; it is followed by an explanation of the results obtained from the review (Section 2.3); this is followed by a discussion (Section 2.4) of the results; finally, a conclusion (Section 2.5) is drawn and we throw some light on limitations, future work and open questions to be answered.

2.2 Methodology

Our broader objective was to find the CC indicators that have been detected using different modalities to understand the quality of CC. We, therefore, conducted a literature review following the guidelines of the PRISMA statement (Liberati et al., 2009). The PRISMA statement lists a step-by-step procedure to do a systematic literature survey. According to it, the information flow in a systematic literature review goes through four different phases, that is, identification, screening, eligibility, and inclusion of articles. In the identification phase, records are identified through database screening using search terms. In the screening phase, duplicate articles are removed and some other articles are removed based on quick scanning. In the eligibility phase, full-text articles are assessed based on the inclusion–exclusion criteria. Finally, articles are included based on the scope of the review. We ran our search in the following databases: ACM Digital Library, SpringerLink, Science Direct, IEEE Xplore, International Society of the Learning Sciences repository, and Google Scholar. We used the following search terms: *(multimodal indicators) AND (multimodal learning analytics) AND (collaborative) AND (quality of collaboration)*. This search term was formulated based on the scope and objectives of the review as mentioned in the research questions.

While searching, a first screening was performed by scanning the title and abstract of the articles and then removing any duplicates. The end result of this screening came to 186 articles. We then further narrowed down the number of articles based on the inclusion and exclusion criteria. The inclusion criteria is as follows.

1. The full text is in English.
2. A peer-reviewed journal article, full paper, or a workshop paper.
3. Description about both CC and use of indicators during CC.

The exclusion criteria is as follows.

1. Description about online (or remote) collaboration.
2. A demo or a poster paper.
3. Architectural details or technical implementation of a CC detection framework only.
4. Framework for assessment and evaluation of user-perceived benefits only.
5. Description about student retention, pedagogy and course design using a multimodal approach, big data engineering in CC, personality detection using MMLA, and human–machine collaboration.

Finally, 88 articles were then deemed fit for our review. We do not consider the number of groups studied during collaboration in each of these articles in the inclusion and exclusion criteria.

2.3 Results

In this section, we describe the results of our analysis. In the first round of analysis, the selected publications were classified according to the sensors, indicators, and indicator types. One or more indicator types can be tracked using the hardware device (i.e., a sensor). For instance, a microphone sensor can only track audio indicator type whereas multiple indicator types like audio, posture, gesture, and spatial can be tracked by a Kinect (i.e., an integrated sensor which can simultaneously act as an infrared, depth, audio and video sensor). This can give an idea about the sensors used in different CC studies. Each indicator type cluster is composed of multiple indicators of CC detected by the sensors. For example, audio data is composed of different indicators such as pitch, amplitude, and speaking time detected by the microphone sensor. Most articles referred to a combination of different modalities like audio and video (Viswanathan and VanLehn, 2017; Hardy and White, 2015; Andrade-Lotero et al., 2013). But, for the sake of clarity and ease of explainability, they have been reported as unimodal rows in all the tables where each indicator type belongs to only one modality. So, there will be an overlap of the references listed under each of these indicator types which do not imply that all the articles essentially were unimodal in nature. These indicators have been used to define collaboration quality with the help of high-level proxy measurements which in this review are defined as indexes obtained by aggregating one or many indicators. For example, a group which exhibits higher *equality* of total speaking time of each member during CC has a better quality of collaboration (Bachour et al., 2010).

Finally, we made a scenario-driven prioritization to choose a set of indicators depending on the particular scenario of CC. This formed the basis for modeling the collaboration detection framework by mapping the fundamental parameters in those scenarios onto the indicator types and indexes. There are different fundamental parameters in each scenario because of differing goals of different scenarios, team composition (such as roles and compulsory interaction with specific artifacts because of the task type), and varied group behavior (such as dominance or coupling). For example, some CC tasks already have preassigned roles (Hare, 1994) for each group member and in some tasks, roles emerge during collaboration (Strijbos and Weinberger, 2010). Some group members are more dominant while others are not.

Then we classified the articles based on the methodologies employed and the type of study, that is, correlational or interventionist (where feedback mechanisms had been employed to support CC) to get a high-level overview. Studies used different methodologies such as observations (e.g., Scherr and Hammer (2009); Davidsen and Ryberg (2017); Praharaj et al. (2018b); Tausch et al. (2014)), sensor-based approach (e.g., Cukurova et al. (2017a, 2018); Spikol et al. (2017b); Kim et al. (2008)), standard measurement scales like that of Meier's (e.g., Reilly et al. (2018)), self-reporting mechanism (e.g., Anastasiou and Ras (2017); Kim et al. (2008)), and indirect learning outcome performance measures (e.g., Reilly et al. (2018)). The types of study found are: correlational study (e.g., Cukurova et al. (2017a, 2018, 2017b)) and interventionist study (e.g., Praharaj et al. (2018b); Kim et al. (2008);

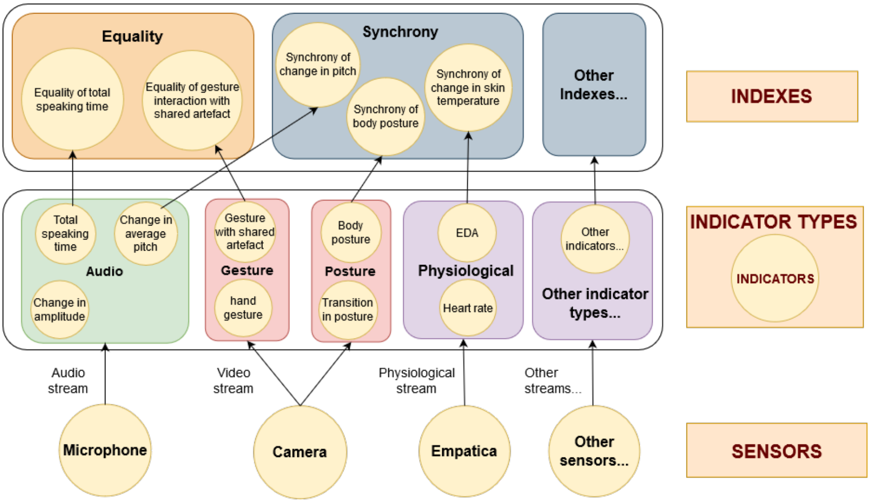


Figure 2.1 Outline for the grouping of the articles along with the terminology used in the review (i.e., sensors, indicators, indicator types, and indexes).

Tausch et al. (2014)).

2.3.1 Indicators to Assess Collaboration Quality

As a first step, all articles obtained were grouped according to the sensors, indicators, and indicator types. Fig. 2.1 (which shows some sensors) gives an outline of the grouping of the articles included in this review. First, the *sensor* data streams give rise to meaningful *indicators* of collaboration obtained after processing. Similar indicators are clustered together to different *indicator types*. For instance, the audio stream obtained from the microphone (sensor) is processed to obtain the total speaking time (indicator) which is put into the audio (indicator type) cluster. Then these indicators are aggregated and processed to form the high-level collaboration quality measure. For instance, the total speaking time (indicator) of each group member can be compared to measure the equality of total speaking time of the group (index). If the value of this equality index is high then the quality of collaboration is good. Thus, these *high-level indexes* are made up of the *low-level indicators* of collaboration (by processing and aggregation) and act as a proxy for measuring collaboration quality. The indexes (i.e., *synchrony*, *equality*, *individual accountability*, *intra-individual variability*, *information pooling*, *mutual understanding*, and *reciprocal interaction*) outlined in the results section are based on the practically detected indexes as found in the literature review. Although theoretically different indexes of collaboration quality have been outlined by Meier et al. (2007), only a few have been operationalized. Meier's scale was used by the articles considered for this review when they used a practically detected collaboration quality measure.

Sensors, indicators and indicator Types

Indicators of CC are obtained from different modalities like audio and video using different sensors like microphone and Kinect. The indicator types represent the cluster of similar indicators of collaboration detected by the sensors. One or more indicator types can be tracked by using a particular sensor. Table 2.1 gives an overview of the different sensors and their indicator types.

Indicator type—audio Most of the articles contained audio as an indicator type. *Audio* is composed of the following *indicators*: prosody of sound such as change in pitch, spectral property variation, change in tone, and intensity (Bassiou et al., 2016); nonverbal features like the total speaking time of group members (Bergstrom and Karahalios, 2007; Bachour et al., 2010), the number of interruptions (Oviatt et al., 2015), and overlap or no overlap duration of speech (Bassiou et al., 2016); the total speaking time of a member together with the attention of other group members measured by their gaze (Terken and Sturm, 2010); linguistic features such as frequency of pronouns used, length of the used sentences, and number of prepositions used (Schneider and Pea, 2014a, 2015). It has been found that a combination of both group speech-based and individual speaker-based indicators is a good predictor of the collaboration quality (Bassiou et al., 2016). The audio was captured in different settings (e.g., working around a tangible user interface (TUI) (Martinez-Maldonado et al., 2013), working with a sociometric badge worn around the neck (Kim et al., 2008), and working under camera observation in videotaped post hoc studies (Scherr and Hammer, 2009; Davidsen and Ryberg, 2017)).

To report further, Terken and Sturm (2010) gave real-time feedback to group members' in meetings by analyzing their *total speaking time* and *eye gaze*. Different colored circles were used to show the feedback by projecting in front of the group member on the table top. These colored circles represented attention to and from speakers and listeners measured by eye gaze and the total speaking time of that group member. On evaluating the effect of the feedback it was found that: the feedback was accepted as a positive measure by most group members; use of feedback promoted a balanced participation among the group members. This participation was measured in terms of the total speaking time of each member. It was found that the speaker and listener eye gaze measured to track the total attention of the listener and speaker was not a good collaboration quality indicator. According to the authors, controlling eye gaze was intuitively difficult as compared to controlling the total speaking time even though both can be consciously controlled.

Other studies also used the total speaking time as an indicator of collaboration (Bergstrom and Karahalios, 2007; Bachour et al., 2010). The group was having a conversation around a smart table. The total speaking time of each member was reflected back to them by a LED light display (Bachour et al., 2010) and concentric circle visualization (Bergstrom and Karahalios, 2007) on the table top. This mirroring feedback helped to regulate the equality of participation during the conversation. Therefore, the group that had better equality of speaking time had a better quality

Table 2.1 Overview of Sensors[1] and Indicator Types[2]

Indicator types	Sensors	References
Audio	Kinect, microphones, sociometric badge	Lubold and Pon-Barry (2014), Nakano et al. (2015), Luz (2013), Grover et al. (2016), Worsley and Blikstein (2015), Martinez-Maldonado et al. (2017a), Oviatt et al. (2015), Scherer et al. (2012), Schneider and Pea (2014a), Spikol et al. (2017b), Echeverria et al. (2017), Ochoa et al. (2013), Schneider and Pea (2015), Scherr and Hammer (2009), Terken and Sturm (2010), Kim et al. (2008), Bergstrom and Karahalios (2007), Bassiou et al. (2016), Thompson et al. (2014), Viswanathan and VanLehn (2017), Kim et al. (2015), Martinez-Maldonado et al. (2013), Bachour et al. (2010), Spikol et al. (2017b), Praharaj et al. (2018b), Worsley and Blikstein (2018), Davidsen and Ryberg (2017), Emara et al. (2017), Rodríguez et al. (2017), Dornfeld et al. (2017), Fake et al. (2017), McBride et al. (2017), Abdu (2015), Flood et al. (2015), Wise et al. (2015), Wake et al. (2015), Martin et al. (2015), Andrade (2015), Dornfeld and Puntambekar (2015), Hardy and White (2015), Andrade-Lotero et al. (2013), Thompson et al. (2013), Martinez et al. (2011), Wong et al. (2011), Johansson et al. (2011), Noel et al. (2018), Bhattacharya et al. (2018), Stewart et al. (2018), Olsen and Finkelstein (2017), Henning et al. (2009)
Posture	Kinect, camera, ceiling mounted time-of-flight sensors, sociometric badge	Grover et al. (2016), Ochoa et al. (2013), Schneider and Blikstein (2015), Scherr and Hammer (2009), Kim et al. (2008), Cukurova et al. (2018), Cukurova et al. (2017a), Cukurova et al. (2017b), Viswanathan and VanLehn (2017), Dich et al. (2018), Davidsen and Ryberg (2017), Stiefelhagen and Zhu (2002), Wise et al. (2017), Bhattacharya et al. (2018), Reilly et al. (2018)

Indicator types	Sensors	References
Gesture	Kinect, camera	Grover et al. (2016), Worsley and Blikstein (2015), Spikol et al. (2017a), Echeverria et al. (2017), Ochoa et al. (2013), Martinez-Maldonado et al. (2015a), Schneider and Blikstein (2015), Scherr and Hammer (2009), Kim et al. (2008), Anastasiou and Ras (2017), Cukurova et al. (2018), Cukurova et al. (2017a), Cukurova et al. (2017b), Viswanathan and VanLehn (2017), Martinez-Maldonado et al. (2013), Worsley and Blikstein (2018), Davidsen and Ryberg (2017), Wise et al. (2017), Emara et al. (2017), Flood et al. (2015), Wake et al. (2015), Hardy and White (2015), Andrade-Lotero et al. (2013), Martinez et al. (2011), Johansson et al. (2011)
Eye gaze	camera, eye tracker, eye tracking glass	Dierker et al. (2009), Nakano et al. (2015), Li et al. (2010), Grover et al. (2016), Schneider and Pea (2014a), Spikol et al. (2017a), Schneider et al. (2015), Scherr and Hammer (2009), Terken and Sturm (2010), Dich et al. (2018), Andrist et al. (2018), Davidsen and Ryberg (2017), Stiefelhagen and Zhu (2002), Flood et al. (2015), Martinez-Maldonado et al. (2015b), Wake et al. (2015), Andrade (2015), Andrade-Lotero et al. (2013)
Spatial	Kinect, camera	Martinez-Maldonado et al. (2017a), Healion et al. (2017), Schneider and Blikstein (2015), Kim et al. (2008), Martinez-Maldonado et al. (2017b), Spikol et al. (2017b), Spikol et al. (2018b), Wise et al. (2017), Martinez-Maldonado et al. (2015b), Reilly et al. (2018)

Indicator types	Sensors	References
Content (i.e., ideas, knowledge, or task related log data)	tangible-user-interface (TUI), human observer, tablets	Tausch et al. (2014), Kim et al. (2008), Thompson et al. (2013), Coopey et al. (2014), Harrer (2013), Fischer et al. (2002), Echeverria et al. (2017), Ochoa et al. (2013), Martinez-Maldonado et al. (2015a), Schneider and Blikstein (2015), Wong-Villacrés et al. (2016), Scherr and Hammer (2009), Anastasiou and Ras (2017), Granda et al. (2015), Thompson et al. (2014), Viswanathan and VanLehn (2017), Martinez-Maldonado et al. (2013), Ahonen et al. (2018), Dornfeld et al. (2017), Fake et al. (2017), Oshima et al. (2017), McBride et al. (2017), Abdu (2015), Flood et al. (2015), Martinez-Maldonado et al. (2015b), Manske et al. (2015), Wise et al. (2015), Martin et al. (2015), Hardy and White (2015), Hashida et al. (2013), Martinez et al. (2011), Wong et al. (2011), Olsen and Finkelstein (2017)
Writing	digital pen	Nakano et al. (2015), Zhou et al. (2014), Scherer et al. (2012), Ochoa et al. (2013), Wong-Villacrés et al. (2016), Granda et al. (2015)
Physiological	empatica	Worsley and Blikstein (2015), Pijeira-Díaz et al. (2016), Dich et al. (2018), Ahonen et al. (2018), Worsley and Blikstein (2018), Malmberg et al. (2019), Pijeira-Díaz et al. (2019), Henning et al. (2009), Henning et al. (2001), Starr et al. (2018), Elkins et al. (2009)
Self-reports	online forms, questionnaires	Wong-Villacrés et al. (2016), Kim et al. (2008), Anastasiou and Ras (2017), Pijeira-Díaz et al. (2016), Granda et al. (2015), Dierker et al. (2009)

[1] Sensors report which hardware sensors have been used to detect these indicator types in each of these referenced articles.
[2] Indicator types report the cluster of similar indicators.

of collaboration as measured by a posttest. However, this type of reflective feedback can be shallow in nature. It assumes that self-reflection will promote collaboration among the group members but doesn't drive them to actively collaborate.

To analyze other audio indicators in depth, Bassiou et al. (2016) used *non-lexical* indicators of audio. They used a combination of manual annotation and support vector machine to predict the collaboration quality of the group. Types of collaboration quality marked by expert annotators are: good (when all 3 members are working together and contributing to the discussion), cold (when only two members are working together), follow (when one member is taking the lead without integrating the whole group) and not (when everyone is working independently). This coding was based on two types of engagement: simple (i.e., talking and paying attention) and intellectual (i.e., actively engaged in the conversation). According to them, a combination of the *group speech activity* indicators (i.e., solo duration, overlap duration of two persons, overlap duration of all three persons, the ratio of the duration of speaking time of the least and most talkative person in the group, and the ratio of the duration of the speaking time of second most talkative student to the most talkative student in the group) and *individual speaker-based* indicators (i.e., spectral, temporal, prosodic and tonal) were good predictors of collaboration quality as marked by the annotators. Moreover, the group-level indicators alone were good predictors of collaboration quality. They found that it was because the individual speaker-based indicators are agnostic to the group information contrary to the group speech activity indicators. All these indicators were fed to a machine learning classifier to determine the quality of collaboration, so in the end, it was a black-box approach. They did not employ any fine-grained in-depth analysis which could have helped to find the relationship of different indicators with the quality of collaboration.

Similarly, *speaker-based* indicators like the change in intensity, pitch, and jitter were used to detect collaboration quality among working pairs (Lubold and Pon-Barry, 2014). Rapport was detected from these indicators and compared to the self-reported rapport to find the collaboration quality. The prediction gave a high-level overview of non-lexical features like pitch but missed the fine-grained semantic meaning of different non-lexical features such as turn-taking, emotional tone while speaking, cross-talk and number of interruptions. These fine-grained vocal characteristics such as turn-taking and overlap of speech are distinctive of collaboration quality; more frequent speaker changes (i.e., *turn-taking*) with overlap of speech (Kim et al., 2015) indicates a good quality of collaboration. Previous research also indicated that overlap in speech is associated with positive group performance (Çetin and Shriberg, 2006; Dong et al., 2009).

Additionally, other works on CC quality focused on expertise detection and productive problem-solving (Luz, 2013; Ochoa et al., 2013; Oviatt et al., 2015), estimation of success (Spikol et al., 2017b), collaboration detection (Viswanathan and VanLehn, 2017), and differentiating student learning strategies (Worsley and Blikstein, 2015) during CC using the audio indicator type. Oviatt et al. (2015) tracked the speech of students working in groups solving mathematical problems. They found that *overlapped speech* was an indicator of constructive problem-solving progress, expertise and collaboration. Both the *number of overlap* in speech and the *duration of the overlap* in speech were taken into account by them. Luz (2013) used the nonverbal audio indicators like presence or absence of speech, silence, pause, transition from group speech to individual speech as indicators to predict performance and expertise on a Maths dataset corpus of groups collaborating for solving mathematical problems. Using these nonverbal indicators as features, they trained a model to predict the group expertise and their performance during collaboration. They found that these features were able to predict the expertise but not the group performance. They did not do any analysis to find the valence of these individual audio indicators and how each indicator was related to the collaboration quality. Spikol et al. (2017b) used audio level and other nonverbal indicators to estimate the success of collaboration activity (i.e., measured by the human observers) while performing open-ended physical tasks around smart furniture. They found that audio level alone is sufficient to predict the quality of collaboration with high accuracy. They detected if collaboration was good or bad but didn't evaluate the contribution of how audio level was predicting in the detection of the quality of collaboration.

To summarize the audio indicator type based on different studies mentioned above: total speaking time, the number of interruptions while speaking, and overlap of speech had been found to be good indicators to predict collaboration quality across most of the articles of that cluster. The number of interruptions and overlap of speech was directly proportional to collaboration quality in some studies. Apart from these individual speaker-based indicators, the total speaking time was seen as a group level indicator when the total speaking time of individual members was compared at the group level to find the equality of participation. If a group had higher equality of total speaking time, then that group had a better quality of collaboration. Other group-level indicators (such as the ratio of the duration of speaking time of least and most talkative member, the ratio of the duration of speaking time of second most talkative member and most talkative member) had helped in the prediction of collaboration quality. Some speaker-based indicators like the change in pitch and amplitude have helped in the detection of collaboration quality, they had done so because of not losing the group level information. For example, when the change in amplitude of two or more group members were similar then they were said to be in synchrony (i.e., one of the high-level measures called indexes); thus, exhibiting a good quality of collaboration. Not all speaker-based indicators' roles in detecting the quality of CC had been successfully discerned. For instance, silence or presence of speech had been used as features to train a model to detect the collaboration quality. But, a qualitative analysis of these indicators was missing which makes it difficult to inform practitioners as to what the occurrence of single or multiple instances of silence or presence of speech can mean during CC.

Indicator type—posture This indicator type comprises *body posture* (Schneider and Blikstein, 2015; Grover et al., 2016; Kim et al., 2008), *head movements* (Cukurova et al., 2017a, 2018, 2017b), and *transitions* between these postures (Schneider and Blikstein, 2015). Schneider and Blikstein (2015) used a TUI for pairs of students to predict learning gains by analyzing data from multimodal learning environments. The task of the students was to rebuild a human auditory system on the TUI in two different conditions (i.e., the discover condition where the rebuilding takes place without instruction and the listen condition with instructions). When tracking the posture along with the gesture using a Kinect sensor (Version 1) which can track the posture and gesture of a maximum of four students at a time based on their skeletal movements, it was found that the *hand movements* and *posture movements* (coded as active, semiactive and passive) are correlated with learning gains during CC. The more active a student was, the higher the learning gain was. Even the number of transitions between these three phases was a strong predictor of learning. Students who used both hands showed higher learning gains. Some of the activities that were logged by the TUI, like the frequency of opening the information box in the TUI did not correlate significantly with learning gain. Also, other indicators like the distance between the group members and the synchrony in body posture did not prove to be effective to detect collaboration quality.

Indicator type—gesture Other works used gestures of group members in open-ended CC scenarios such as building prototypes (Spikol et al., 2017a; Cukurova et al., 2018, 2017a). Gesture is comprised of *hand movements* (Cukurova et al., 2017a, 2018), *hand gestures* like pointing (Grover et al., 2016), *hand interactions with an object* (Spikol et al., 2017a; Cukurova et al., 2018), *hand interactions around touch screens* (Martinez-Maldonado et al., 2015a; Echeverria et al., 2017) or special interaction devices like a TUI. To elaborate further, Spikol et al. (2017a) and Cukurova et al. (2017a, 2018) studied collaborative learning specifically in the context of collaborative problem solving (CPS). They tracked the combination of *hand movements, head direction*, and physical engagement using customized smart furniture. The videos were then coded by experts with 0 (for passive), 1 (for semiactive) and 2 (for active) based on different combinations of head and hand positions. These codes helped to determine synchronization and physical engagement. A group in which all group members were coded as active for most of the time had a higher value of synchrony. Hence the group had a good quality of collaboration. It will be elaborated in detail in the next part where we discuss the indexes. Another CPS context was studied by Grover et al. (2016) in pair programming. They captured data from different modalities (i.e., video, audio, clickstream, and screen capture) unobtrusively using Kinect. For initial training of the classifiers using machine learning, experts coded the video recordings with three annotations (i.e., High, Medium and, Low) when they found evidence of collaboration between the dyads. The indicators of collaboration detected are: *pointing to the screen*, *grabbing the mouse from the partner*, and synchrony in *body position*. This classifier then later predicted the level of collaboration. Further qualitative analysis was not done. The problem with capturing the gestures is that sometimes the view of the hand movements of the students gets obfuscated or overlapped; it is solely dependent on the positioning of cameras (i.e., the angle from which the camera can capture the frontal view or top view). Consequently, considering different indicator types like audio and eye gaze along with gesture is preferred.

Summarizing the gesture and posture indicator type, most of the tasks were open-ended using TUIs. It was found that if group members used both their hands, spent more time in an active engaging posture, and the majority of the members were in the active posture, then they had a good quality of collaboration. However, synchrony in body posture is not always a good marker of collaboration.

Indicator type—eye gaze This indicator type comprises the *joint visual attention* (JVA) (i.e., the proportion of times gazes of individuals are aligned by focusing on the same area of the shared object or screen). JVA is a good predictor of the quality of collaboration for a group which is reflected by their performance. Schneider et al. (2015) showed that JVA can be used as a reflection mechanism in co-located settings; they showed each student their partner's gaze patterns in real time to improve collaboration. The higher the JVA was, the better was the quality of collaboration. Similar to JVA, Dierker et al. (2009) used an augmented reality (AR) set up during a collaborative object choice task; here, they established joint attention by assigning different roles to the group members working in pairs. One member was the gazer who had to observe an object on the head-mounted display and fixate it on the table; then the other member who was the searcher had to find that object on the table. One group received real-time augmented visual and acoustic feedback with the help of AR goggles to facilitate their collaboration, whereas the other group did not receive any feedback. It was found that the group receiving feedback had a shorter reaction time and lower error rates during the task.

Most of the other studies on eye gaze focused on: the attention of other group members on their peers (Terken and Sturm, 2010); determining the social context from gaze (Li et al., 2010) during group work; observing gaze patterns in post hoc studies (Scherr and Hammer, 2009) from the videotaped collaboration recordings; coding the activity index (i.e., 2 for active, 1 for semiactive, and 0 for passive) of group members interacting with an object based on eye gaze and other nonverbal features (Cukurova et al., 2017a, 2018). Some studies (Terken and Sturm, 2010) did not find any effect of the eye gaze of group members on the quality of collaboration. The experiments linked with the use of eye gaze were sometimes dependent exclusively on the shared artifacts which needed to be properly set up in the room to get the correct measure of JVA.

Indicator type—spatial This indicator type is a mix of the proximity indicator (i.e., the distance between the group members) (Martinez-Maldonado et al., 2017b;

Schneider and Blikstein, 2015; Spikol et al., 2017b,a) and the positioning of the members (i.e., their mobility) (Martinez-Maldonado et al., 2017b; Healion et al., 2017). Some collaboration scenarios like medical simulations need the collaborating members to move around the room (or occupy particular positions) while performing the operation or other medical tasks. These studies did not find any relationship between positioning in the room and the collaboration quality. However, the lesser the distance between the group members is, the better is the quality of collaboration (Spikol et al., 2017b,a). Other studies, however, did not find any correlation between the distance of group members and the quality of collaboration (Schneider and Blikstein, 2015).

Indicator type—content Apart from the indicator types discussed above, this indicator type is a combination of ideas (Tausch et al., 2014; Kim et al., 2008) and knowledge (i.e., the content-related knowledge obtained from the interactive devices or the task itself) (Martinez-Maldonado et al., 2015a; Anastasiou and Ras, 2017). Tausch et al. (2014) used human observers during collaborative brainstorming to monitor the number of ideas generated by each member. Three members in each group performed the task. The group members were supposed to discuss a certain topic and their collaboration quality was measured by the number of ideas generated. A comparison metric for collaboration such as a baseline was calculated using the average number of ideas generated by the group. Using this baseline, each group member was marked as below average or above average depending on the number of ideas generated by each member. Then the feedback was shown as a metaphorical group garden moderated by human observers. It was found that the groups who received real-time feedback had a better quality of collaboration because of a nearly equal number of ideas contributed by each group member in the group without any dominance from one member. Similarly, self-reports have been used to monitor the number of ideas generated by each member during collaborative brainstorming (Kim et al., 2008). Content of interaction during an activity (around a TUI) was tracked and communicated back to the group members using textual and haptic feedback (Anastasiou and Ras, 2017) on the tabletop. In addition to this, the actions of students around a TUI also helped in detecting the quality of CC. Martinez-Maldonado et al. (2015a) tracked these actions and communicated back to teachers inside a classroom in real time. In this study, the teachers gave the students a task to work collaboratively around a TUI to build conceptual maps, perform collaborative brainstorming and take part in scripted group meetings. The teachers received feedback about the performance of a group both on individual and group levels with colored visualizations, statistical displays, and notifications on personal tablets. This enabled them to intervene immediately when they find misconceptions or problems in any group's performance. Most of these works employ human observers. This is because of the semantic nature of the discussion where automated understanding of the content is difficult by using a machine.

Many other works used TUIs or multi-tabletop touch interfaces to track the content of the collaborative task and activity (Echeverria et al., 2017; Granda et al., 2015; Wong-Villacrés et al., 2016). Echeverria et al. (2017) used a combination of a TUI-based tool called DBCollab, a personal tablet, and a Kinect sensor to track the activity of students during a database design session in the classroom. They gave real-time feedback to facilitate the database design task. The teacher's solution was stored and compared with the solution of the students; this helped to drive the real-time feedback by comparing the similarity between both solutions. This feedback improved collaboration. Granda et al. (2015) used a multi-tabletop TUI for database design. They gave feedback on students' activity to the teachers in colored symbols. Basically, they tracked the database entity-related actions like creating, editing, and deleting the objects. Wong-Villacrés et al. (2016) tracked content-related activities by comparing a TUI-based set up and a paper-based set up. They found that the group in the TUI set up had more respect for their peers, better communication, and in turn better collaboration as compared to the group in the TUI set up. It is due to the reason that they received continuous feedback from the TUI about their contribution and their peers' contribution which improved their awareness.

Indicator types—writing, physiological, and self-reports Writing includes different indicators derived from the interactions using the digital pen like the pen stroke analysis (Zhou et al., 2014). Physiological indicator type has skin temperature (Pijeira-Díaz et al., 2016) and heart rate (Henning et al., 2009) as indicators. Pijeira-Díaz et al. (2016) used electrodermal measures obtained from one wrist using empatica (i.e., a smartwatch to measure different physiological signals like heart rate and skin temperature) and tried to relate it to three aspects measured by a test and self-reports. The three aspects are: collaborative will measured by a self-report questionnaire before the collaboration task; collaborative product measured by a self-report questionnaire after the collaboration task; and dual learning gain measured by the difference between the posttest and pretest scores. If in a group the direction of arousal pattern of electrodermal activation was synchronous among the group members, then that group showed a good quality of collaboration measured in terms of learning gain. Other uses of self-reports are in the form of a satisfaction survey given to the participants involved in group work (Anastasiou and Ras, 2017; Dierker et al., 2009) or some extra information related to the collaboration task (like information about the self-perceived levels of rapport (Lubold and Pon-Barry, 2014)). The higher rapport between the group members results in a better quality of collaboration.

Now, we summarize the indicator types discussed above (eye gaze, spatial, content, and physiological). All the articles which use eye gaze as an indicator type to detect collaboration quality conclude that the more often JVA occurs, the better is the quality of collaboration. Some closed collaborative tasks which had predefined specific mobility and position requirements in the room tracked the distance between the group members. It was found that the lesser the distance between the members is, the better is the quality of collaboration. However, this was not consistent.

Summarizing the content indicator type, the content of the discussion during CC gives rise to idea generation and it was found that if all group members equally contribute to the number of ideas generated, then that group had a good quality of collaboration. In some other works, the content of CC was dependent on the task requirements and the closeness it has to the designed solution. This indicated the quality of collaboration. Some CC tasks tracked the physiological signals and found that if the patterns of the skin temperature of the group members are in sync then those groups exhibit good collaboration quality.

Combined indicator types Some works (e.g., Martinez-Maldonado et al. (2013); Worsley and Blikstein (2015); Kim et al. (2008)) used a combination of multiple indicator types. For instance, Martinez-Maldonado et al. (2013) used a TUI, microphone array, and Kinect to detect the indicators of CC. They performed a task with two phases (i.e., brainstorming and linking). Then their aim was to differentiate different collaboration levels by taking the help of a combination of the captured *audio* and the physical tabletop actions like the *touch* on the TUI, *frequency* of opening of different task-related information shown in the TUI. A microphone array was used to capture the audio; for touch, they used Kinect to differentiate touch and other interactive actions of each person. During the post hoc analysis, they found that the more collaborative groups had higher verbal interactions as compared to the less collaborative groups during the brainstorming phase. They also exhibited less concurrency and parallel work. Additionally, the more collaborative group also had more verbal responses after someone spoke.

Worsley and Blikstein (2015) used *human annotations, speech, electrodermal activation (EDA) data*, and *gestures* to differentiate student learning strategies while working in groups. The groups were assigned to two different conditions: principle-based reasoning and example-based reasoning. They found that students in the principle-based reasoning condition showed more flow (i.e., near or below average audio, hand/wrist movement, and electrodermal activation) and action (i.e., above average hand/wrist movement) behavior compared to their counterparts in the example-based reasoning group; flow behavior also positively correlated with learning (i.e., the outcome of collaboration).

Kim et al. (2008) used a sociometric badge (i.e., an electronic sensing device worn around the neck that can collect and analyze social dynamics), which acted as a meeting mediator to capture *audio* and *postures* during meetings of four members in one group. This badge bridged the gap of dominance and increased the equality of participation among the group members using real-time feedback on their personal mobile phones. Dominance was primarily measured by the total speaking time and equality of turn taking of the group members. If these are more or less equal then there is less dominance and the quality of collaboration is good. However, the use of more indicator types may not always help in maximizing the CC detection potential, rather can be a requirement of that particular scenario (Ahonen et al., 2018).

Indexes

Indexes are the high-level quality markers of collaboration. They can act as a proxy to understand, measure, and predict collaboration quality. They are composed by aggregating the low-level indicators of collaboration such as pointing, head orientation, hand movement, eye gaze, etc. Table 2.2 shows the overview of the indexes that have been detected practically from different indicators of CC.

Synchrony It means a situation where two or more group members are in sync with each other based on some criteria. For example, if two members in a group are speaking at different amplitude but exhibiting the same pattern of their speech (e.g., the rise and fall of the pitch of both members are similar to each other) then they are showing a high level of synchrony (Lubold and Pon-Barry, 2014). Synchrony has been detected using audio indicator type (Lubold and Pon-Barry, 2014; Nakano et al., 2015) and writing indicator type (Nakano et al., 2015). Lubold and Pon-Barry (2014) found a positive correlation between synchrony and rapport (generated by comparing perceptual rapport from annotators and self-reported rapport) during collaborative interactions. A good rapport between group members can enhance the collaboration (Chapman et al., 2005). Nakano et al. (2015) used writing (i.e., timestamped duration of writing notes or not writing obtained from the pressure and contact features of a digital pen) as an indicator type to detect synchrony. They found different participation styles like passive participation, receptive participation, conversation management, and proactive participation among the group members using binary (i.e., present or absent) behavior labels obtained from writing and gaze indicator type (e.g., group member x is gazing at group member y, group member x is gazing at group member y's note, group member x is writing a note). The co-occurrence patterns (i.e., number of times one or more behaviors occur in a time window) of these behaviors can be used to predict the participation styles during collaboration. Participation styles change during a collaborative task because of role swapping to promote positive interdependence leading to effective collaboration (Soller, 2001). Similarly, synchrony can also be defined using *nonverbal indicators* such as activity during group work (i.e., all members in the group are either active, semiactive, or in passive posture (Cukurova et al., 2018)). Synchrony was detected there by using number coded activity indexes (i.e., 2 for active, 1 for semiactive and 0 for passive) derived from different indicators like *hand position* and *head orientation* (Cukurova et al., 2017a, 2018, 2017b). They designed a task in which each group member was interacting with an object in a group. The group members are said to be in synchrony when all the members are in the same state (i.e., all active, semiactive and passive). Here the valence of synchrony was determined based on whether the synchrony is positive because of all active group members or negative due to all passive ones. It was found that groups with high competence university students' (as assessed by expert teachers) had more instances of positive synchrony during CC. So, the groups that showed higher instances of active or positive synchrony had better

Table 2.2 Overview of Studies on Practically Detected Collaboration Indexes and Indicators

Indexes[a]	**Indicator types**	**Indicators**	**References**
Synchrony ⥣	Audio	rise and fall of average pitch, intensity	Lubold and Pon-Barry (2014)
		rise and fall of average amplitude	Nakano et al. (2015), Spikol et al. (2017b)
	Posture	body position, leaning forward	Grover et al. (2016)
		relaxed or active body posture	Schneider and Blikstein (2015)
		head direction	Cukurova et al. (2017a), Cukurova et al. (2018)
	Gesture	pointing	Barthelmess et al. (2005)
		hand movement	Spikol et al. (2017b)
		using both hands	Schneider and Blikstein (2015)
		hand position	Cukurova et al. (2017a), Cukurova et al. (2018)
	Eye gaze	gaze at speaker, non-speaker/note	Nakano et al. (2015)
		joint visual attention (JVA)	Dierker et al. (2009), Li et al. (2010), Schneider and Pea (2014a), Schneider and Pea (2015), Andrist et al. (2018)
	Writing	presence or absence of writing	Nakano et al. (2015)
	Physiological	electrodermal activation (EDA)	Dich et al. (2018), Pijeira-Díaz et al. (2019), Elkins et al. (2009), Starr et al. (2018)
		heart rate	Henning et al. (2009), Henning et al. (2001)

Indexes[a]	**Indicator types**	**Indicators**	**References**
Equality ⇑	Audio	jitter	Lubold and Pon-Barry (2014)
		total speaking time	Kim et al. (2008), Bergstrom and Karahalios (2007), Bachour et al. (2010), Praharaj et al. (2018b)
	Posture	body movement, sitting, walking	Kim et al. (2008)
		head direction	Cukurova et al. (2018), Cukurova et al. (2017b)
	Gesture	all types (i.e., pointing, nodding)	Kim et al. (2008)
		hand interactions	Cukurova et al. (2018), Cukurova et al. (2017b)
	Content	identifying patterns between data	Coopey et al. (2014)
		number of ideas and questions	Kim et al. (2008)
	Writing	proportion of participation in database modelling	Granda et al. (2015)
Individual Accountability ∗	Posture	head direction	Cukurova et al. (2018), Cukurova et al. (2017a)
	Gesture	hand position	Cukurova et al. (2018), Cukurova et al. (2017a)
Intra-individual Variability ⇓	Posture	head direction	Cukurova et al. (2018), Cukurova et al. (2017b)
	Gesture	hand position	Cukurova et al. (2018), Cukurova et al. (2017b)
Information Pooling ⇑	Content	web search	Hashida et al. (2013)

Indexes[a]	**Indicator types**	**Indicators**	**References**
Mutual Understanding ⇑	Audio	dialogues, verbal discourse, statements, questions	Rodríguez et al. (2017), Abdu (2015), Fischer et al. (2002), Dornfeld et al. (2017), McBride et al. (2017), Johansson et al. (2011)
	Content	task related content, knowledge construction, quantitative and conceptual discourse, idea flow	Abdu (2015), Fischer et al. (2002), Dornfeld et al. (2017), McBride et al. (2017), Martinez-Maldonado et al. (2015b), Fake et al. (2017), Scherr and Hammer (2009), Oshima et al. (2017), Thompson et al. (2013)
	Gesture	touch actions on tabletop, hand movement	Martinez-Maldonado et al. (2015a), Johansson et al. (2011)
	Posture	head orientation	Stiefelhagen and Zhu (2002)
	Eye gaze	eye gaze on peers, shared devices	Martinez-Maldonado et al. (2015b)
	Spatial	position in the room	Martinez-Maldonado et al. (2015b)
Reciprocal Interaction ⇑	Gesture	hand movement on tabletop	Wise et al. (2017)
	Content	explanation, initiation and arguments	Wise et al. (2017)

[a] Indexes report the aggregated collaboration indicators and indicator types report the cluster of similar collaboration indicators. ⇑ denotes that if the value of the index is high then the quality of the collaboration is better and vice versa, and ⇓ denotes that if the value of the index is low then the CC quality is better and vice versa. ∗ denotes that the index's role in determining CC quality is unclear.

quality of collaboration. Other indicator types like eye gaze (Schneider and Pea, 2013) (i.e., JVA or synchronization in eye gaze) have helped to detect synchrony and the findings suggest that it can help in the detection of effective collaboration whereas synchrony may not reflect collaboration in some indicator types like posture (Schneider and Blikstein, 2015). Higher level of physiological synchrony of the skin temperature as seen by Pijeira-Díaz et al. (2019); Dich et al. (2018) can also indicate good quality of collaboration.

Equality In the work by Lubold and Pon-Barry (2014), they explained that if the amplitude of the speakers is the same during their *speech* then they exhibit equality (or convergence). Equality has been defined using *nonverbal postures* with the help of statistical formulas like the sum of the squared difference between the number of coded activities of each group members, standard deviation, and mean difference (Cukurova et al., 2018). Some of these works (Cukurova et al., 2018, 2017b) computed equality among the group members by using different statistical measures like the sum of the squared differences between the activity index (i.e., number coded based on the activity of the group members: 2 for active when a member is interacting with an object; 1 for semiactive when a member is paying attention to the peer but not interacting with the object; and 0 for passive in all other situations) of each group member, the standard deviation of the activity index among the group members and the average mean of the activity index among the group members. The high competence groups (as detected by expert teachers) had all group members with higher physical interaction with the object, in turn showing higher equality for the group. Equality has also been detected using audio as an indicator type (Bachour et al., 2010; Bergstrom and Karahalios, 2007). Here, they used reflective visualization to show the group members the total speaking time of their conversations. This helped them to regulate the equality of participation. So, the over participators (i.e., the group members with a higher percentage of speaking time in the group) reduced their speaking time and the under participators improved their speaking time towards the end of the group session. The groups with higher equality of participation showed better quality of collaboration as evaluated by a posttest. Other examples of equality index computation are by Kim et al. (2008) who used a meeting mediator (or a sociometric badge based real-time feedback) to reduce the gap between dominant and nondominant members during collaborative brainstorming and other tasks. As per their hypothesis, groups who used the meeting mediator had balanced participation and became more collaborative. Tausch et al. (2014) used human observers to monitor the group conversations during collaborative brainstorming. These observers helped to maintain the equality of number of ideas generated by the members by moderating a metaphorical feedback that resembled a groupgarden. The groups with had higher equality of participation measured in terms of ideas generated by each member had also better quality of collaboration.

Individual accountability (IA) IA has been used as another index to measure collaboration quality. It means that at least one of the members in the group is paying attention to the activity of other members; there is not a single member who ignores the activity of other members (Cukurova et al., 2018, 2017a). They used the activity indexes as marked by numbered coding (i.e., 0, 1, and 2 as described earlier) to measure individual accountability. Conceptually, it means that every member in the group is undertaking their share of work and also acknowledging the contribution of the other members. In these works (Cukurova et al., 2018, 2017a), IA was less effective to predict the quality of CPS even though they had a hypothesis that the groups with a higher quality of collaboration will have a higher value of IA. According to the authors IA might not be properly coded to capture the complex collaborative processes in CPS scenario.

Intra-indivdual variability (IVA) IVA for a particular group member is detected by the difference in behavioral activity (i.e., number coded as 2 for active, 1 for semiactive, and 0 for passive) of the member in two sequential time windows (Cukurova et al., 2018). High competence CPS groups (as rated by expert teachers) had a similar frequency of changes in their physical interactions as compared to the low competence groups. Therefore, in other words, a lower value of IVA indicated that the quality of collaboration for the group was good. This may be attributed to a higher shared understanding between the group members (Barkley et al., 2014).

Information pooling It is the accumulation of information measured from the content of the conversation in a particular CC task (Hashida et al., 2013). In this study, the group members try to gather as much information as possible regarding the shared web search to move towards the collaborative web search. So, they help each other out to do the common objective of web search. The groups have a good quality of collaboration if they are good in information pooling.

Mutual understanding It denotes the level of understanding between the group members which is detected mostly by the content of their conversation from the audio indicator type (Rodríguez et al., 2017; Abdu, 2015). Other indicator types like posture, gesture, eye gaze and spatial also help in the detection of mutual understanding based on how each member makes eye contact with the others, the comfort level among them based on their positioning and distance in the group and how they back-channel their conversations (Martinez-Maldonado et al., 2015a,b). Higher level of mutual understanding indicates higher quality of collaboration.

Reciprocal interaction It is measured by the gesture and content indicator type which denotes how group members reciprocate to each other during the CC. This can be a reply given to an initiated question or defending one's position with suitable arguments within the group. The groups which had group members with preassigned roles performed better and had better reciprocal interaction as compared to the groups without any preassigned roles.

To summarize the results, we found different indicators of collaboration which were grouped into different indicator types. Then these indicators were processed and aggregated to form the *indexes* in some works which form the high-level collaboration quality definition. Some group-level indicators in the audio indicator type such as total speaking time, the ratio of the speaking time duration of the most talkative member, and the least talkative member along with the individual speaker-based indicators such as spectral, temporal audio features are indicative of collaboration quality. But, the same is not true for the individual speaker-based indicators alone. Other group level indicators like overlapped speech, interruptions are also indicative of the CC quality; the higher their number, the better is the quality. Duration of speaker and listener eye gaze combined with total speaking time was not useful to detect CC quality. But, JVA measured from eye gaze was useful and indicated that the quality of collaboration is good if there is a higher occurrence of JVA. Similarly, joint posture movements were not indicative of collaboration quality, rather specific postures like active posture indicated better collaboration quality. Similarly, specific gestures such as using both hands indicated better quality of CC. Joint arousal measured from the EDA in physiological indicator type was indicative of CC quality; higher the occurrence, the better is the quality.

We found that the detection mechanisms for these indicators varied from human-based, sensor-based to hybrid event detection. Furthermore, different practically detected indexes, that is, *synchrony, equality, individual accountability*, *IVA*, *mutual understanding*, *information pooling*, and *reciprocal interaction* have been aggregated from a different set of indicator types. For instance, synchrony has been detected using audio, posture, gesture, eye gaze, and writing as indicator types. However, synchrony has not been detected using the content indicator type. It is because detecting synchrony from the CC task content requires understanding the semantics and intent of what is spoken. It is difficult to detect that automatically and laborious for human observers to detect it in a post hoc manner. Unlike these indexes, *equality* has been detected from content indicator type. Similarly, other indexes have been detected from selected indicator types as seen in Table 2.2. Thus, we need to understand in depth what the different collaboration indicators and their sources in different scenarios are before deciding on the design of the suitable conceptual framework model.

2.3.2 Scenario-Driven Prioritization of CC

To map the low-level indicators and the balancing between these indicators on the index level into useful feedback for collaboration, we analyzed the literature on different forms of collaboration and classified these according to the collaboration targets. We found 13 different scenarios of CC: problem-solving, planning, learning, programming, database design, healthcare simulation, gaming, engineering design, design, concept mapping, brainstorming, meetings, and browsing. Table 2.3 gives an

Table 2.3 Overview of Collaboration Scenarios With the Indicator Types, Indicators, Indexes[1] and Their Valence[2] With Reference to Collaboration Quality

Scenarios	**Indicator types**	**Indicators, *Indexes* and Valence**	**References**
Problem-solving	Audio	*synchrony* in rise and fall of amplitude $\Uparrow$, *synchrony* in rise and fall of pitch $\Uparrow$, number of syllables used per second $\Downarrow$, pause duration $\Uparrow$	Nakano et al. (2015), Lubold and Pon-Barry (2014), Scherer et al. (2012)
		verbal discourse such as statement, questions, numbers, context of problem, dialogues $*$, misconception in problem solving $*$	Ochoa et al. (2013), Thompson et al. (2014), Rodríguez et al. (2017), Stewart et al. (2018), Abdu (2015)
		nonverbal audio cues such as silence $*$, number and duration of overlap of speech $\Uparrow$, uninterrupted speech length $*$, frequency of turn taking $\Uparrow$, *equality* of total speaking time $\Uparrow$, *equality* of total speaking time and turn taking $\Uparrow$	Luz (2013), Oviatt et al. (2015), Viswanathan and VanLehn (2017), Kim et al. (2015), Bachour et al. (2010), Kim et al. (2008)
		linguistic features: pronouns, prepositions $*$, number of anaphoras (anybody, anyone, all, another) used $\Uparrow$	Schneider and Pea (2014a), Schneider and Pea (2015)
	Writing	handwriting signal $*$, area used to sketch geometric representation $\Uparrow$, digital pen stroke, writing speed and pressure $*$	Zhou et al. (2014), Scherer et al. (2012), Ochoa et al. (2013)
	Spatial	mobility in the room $*$, distance between individuals $*$	Healion et al. (2017), Schneider and Blikstein (2015)
	Eye gaze	JVA (*synchrony*) $\Uparrow$	Schneider and Pea (2014a), Schneider and Pea (2015)
	Gesture	using both hands $\Uparrow$, touch actions on tabletop $*$	Schneider and Blikstein (2015), Isenberg et al. (2010), Ochoa et al. (2013)
	Posture	*synchrony* in active or relaxed body posture $*$, specific body movements $*$	Schneider and Blikstein (2015), Kim et al. (2008), Ochoa et al. (2013)
	Content	task related $\Uparrow$, tabletop content logs matching near to the solution $\Uparrow$, type of contribution (i.e., symmetric$\Uparrow$, asymmetric$\Downarrow$ or individual$\Downarrow$)	Abdu (2015), Olsen and Finkelstein (2017), Thompson et al. (2014), Anastasiou and Ras (2017), Viswanathan and VanLehn (2017), Isenberg et al. (2010)
	Self-reports	closeness in group $\Uparrow$, tools used for CC $*$	Isenberg et al. (2010), Anastasiou and Ras (2017)
Design	Content	*equality* between patterns of data $\Uparrow$, *mutual understanding* from content logs $\Uparrow$, topic of discussion $\Uparrow$, idea flow $*$, task related content $\Uparrow$	Coopey et al. (2014), Fischer et al. (2002), Dornfeld et al. (2017), McBride et al. (2017), Martinez-Maldonado et al. (2015b), Fake et al. (2017), Flood et al. (2015)
	Eye gaze	JVA (*synchrony*) $\Uparrow$, eye gaze on peers $*$, eye gaze on shared devices $*$	Schneider et al. (2015), Martinez-Maldonado et al. (2015b), Flood et al. (2015)
	Audio	small group talk $*$, dialogues $*$, verbal discourse $*$	Wake et al. (2015), Dornfeld et al. (2017), Emara et al. (2017), McBride et al. (2017), Fake et al. (2017), Flood et al. (2015)
	Gesture	task related touch actions on the TUI $\Uparrow$, facial expression $\Uparrow$, prompts, patterns of gesture $*$	Evans et al. (2016), Malmberg et al. (2019), Emara et al. (2017)
	Physiological	simultaneous arousal by electrodermal activation (EDA) $\Downarrow$	Malmberg et al. (2019)
	Posture	body motion $*$	Flood et al. (2015)
	Spatial	space usage $*$	Martinez-Maldonado et al. (2015b)

Scenarios	Indicator types	Indicators, *Indexes* and Valence	References
Learning	Audio	speech intonation ∗, initiator or listener, verbal discourse ∗	Scherr and Hammer (2009), Davidsen and Ryberg (2017), Andrist et al. (2018), Wong et al. (2011), Thompson et al. (2013), Oshima et al. (2017), Hardy and White (2015), Andrade (2015), Andrade-Lotero et al. (2013), Martin et al. (2015)
	Eye gaze	gaze at peers, JVA (*synchrony*) ⇑	Scherr and Hammer (2009), Andrist et al. (2018), Davidsen and Ryberg (2017)
	Gesture	average hand movement, number of pointing ⇑	Scherr and Hammer (2009), Davidsen and Ryberg (2017), Thompson et al. (2013)
	Physiological	*synchrony* in arousal by electrodermal activation ⇑	Pijeira-Díaz et al. (2019)
	Posture	standing or sitting, bending body ∗	Davidsen and Ryberg (2017), Scherr and Hammer (2009)
	Content	topic of discussion related to the task ⇑, *mutual understanding* from knowledge construction ⇑	Scherr and Hammer (2009), Davidsen and Ryberg (2017), Wong et al. (2011), Manske et al. (2015), Thompson et al. (2013), Oshima et al. (2017), Hardy and White (2015), Andrade (2015), Andrade-Lotero et al. (2013), Martin et al. (2015)
Engineering design	Spatial	distance between group members ∗	Schneider and Blikstein (2015)
		distance between group members ⇓	Cukurova et al. (2017b), Spikol et al. (2017a), Cukurova et al. (2017a), Cukurova et al. (2018), Spikol et al. (2017b), Spikol et al. (2018b)
	Physiological	near or below average arousal by EDA ⇑	Worsley and Blikstein (2015), Worsley and Blikstein (2018)
	Gesture	*synchrony* in hand movement, wrist movement ⇑, *synchrony* in pointing ⇑	Schneider and Blikstein (2015), Cukurova et al. (2017b), Spikol et al. (2017a), Cukurova et al. (2017a), Cukurova et al. (2018), Spikol et al. (2017b), Worsley and Blikstein (2015), Spikol et al. (2018b), Worsley and Blikstein (2018)
	Posture	*synchrony* in posture ∗, *equality* of posture ⇑, synchrony in posture ⇑, *Intra-individual variability* of active or passive posture ⇓, *Individual accountability* of active or passive posture ∗	Schneider and Blikstein (2015), Cukurova et al. (2017b), Spikol et al. (2017a), Cukurova et al. (2017a), Cukurova et al. (2018), Spikol et al. (2017b), Spikol et al. (2018b)
	Audio	amplitude ∗	Spikol et al. (2017b), Spikol et al. (2018b)

Scenarios	**Indicator types**	**Indicators, *Indexes* and Valence**	**References**
Meetings	Audio	*equality* of total speaking time ⇑, average speech segment length ∗, presence of speech ∗	Terken and Sturm (2010), Kim et al. (2008), Bergstrom and Karahalios (2007), Praharaj et al. (2018b), Bhattacharya et al. (2018), Henning et al. (2009)
Meetings	Audio	prosody such as average amplitude and average energy variation in speech ∗	Kim et al. (2008), Bergstrom and Karahalios (2007)
Meetings	Content	*equality* of number of ideas ⇑, content logs matching solution ⇑	Kim et al. (2008), Martinez-Maldonado et al. (2015a)
Meetings	Eye gaze	speaker and listener eye gaze ∗	Terken and Sturm (2010), Stiefelhagen and Zhu (2002)
Meetings	Gesture	*equality* of average gesture variation ⇑	Kim et al. (2008)
Meetings	Posture	*mutual understanding* from head orientation ⇑	Stiefelhagen and Zhu (2002), Bhattacharya et al. (2018)
Meetings	Self-reports	self-report on dominance ∗	Kim et al. (2008)
Meetings	Physiological	*synchrony* in heart rate ⇑	Henning et al. (2009)
Brainstorming	Content	*equality* of number of ideas ⇑, *equality* of total speaking time ⇑, TUI content logs matching the solution ⇑	Tausch et al. (2014), Kim et al. (2008), Martinez-Maldonado et al. (2013)
Brainstorming	Gesture	*mutual understanding* from touch actions with the TUI ⇑	Martinez-Maldonado et al. (2015a), Martinez-Maldonado et al. (2013)
Brainstorming	Posture	*equality* of body movement like sitting and walking ⇑	Kim et al. (2008)
Brainstorming	Audio	nonverbal audio features like presence or absence, length, verbal interactions ∗	Martinez-Maldonado et al. (2015a), Kim et al. (2008), Martinez-Maldonado et al. (2013)
Healthcare simulation	Spatial	correct mobility and positioning in the room around the patient manikin ⇑	Martinez-Maldonado et al. (2017a), Martinez-Maldonado et al. (2017b)
Browsing	Content	content of search (*information pooling*) ⇑	Hashida et al. (2013)
Gaming	Eye gaze	JVA (*synchrony*) ⇑	Dierker et al. (2009), Li et al. (2010)
Gaming	Gesture	*reciprocal interaction* by hand movement on tabletop ⇑	Wise et al. (2017)
Gaming	Posture	body movement ∗	Wise et al. (2017)
Gaming	Spatial	space usage around the tabletop ∗	Wise et al. (2017)
Gaming	Audio	verbal discourse ∗	Johansson et al. (2011)

Scenarios	**Indicator types**	**Indicators, *Indexes* and Valence**	**References**
Planning	Physiological	*synchrony* in direction of arousal ⇑	Pijeira-Díaz et al. (2016)
	Gesture	*synchrony* in pointing ⇑	Barthelmess et al. (2005)
	Content	content of the task like planning, editing, modifying ∗	Harrer (2013)
Programming	Physiological	*synchrony* in arousal from EDA ⇑	Ahonen et al. (2018), Dich et al. (2018), Starr et al. (2018)
	Spatial	proximity ∗	Reilly et al. (2018)
	Gesture	grabbing mouse from the partner ∗	Grover et al. (2016)
	Posture	body position, leaning forward ∗, spending more time in iterating (actively involved in programming a solution) pose as compared to planning, tinkering ⇑	Grover et al. (2016), Reilly et al. (2018)
Database design	Gesture	number of touch actions on the interactive surface ∗	Wong-Villacrés et al. (2016), Echeverria et al. (2017), Granda et al. (2015)
	Content	content logs from TUI matching the correctness of solution ⇑	Wong-Villacrés et al. (2016), Echeverria et al. (2017), Granda et al. (2015)
	Self-reports	evaluation of own and peers' group work skills ∗	Wong-Villacrés et al. (2016), Echeverria et al. (2017), Granda et al. (2015)
Concept mapping	Content	content logs from TUI matching the correctness of solution ⇑	Martinez-Maldonado et al. (2015a)
	Audio	verbal interactions, verbal response to peers ⇑	Martinez-Maldonado et al. (2013), Martinez et al. (2011)
	Gesture	concurrent and parallel touch actions with the TUI ⇓	Martinez-Maldonado et al. (2013), Martinez et al. (2011)

[1] marked in italics when practically detected. [2] positive ⇑, negative ⇓, unclear or no effect ∗.

overview of the studies on different scenarios of collaboration and relevant indicators and indicator types found in those scenarios in detail. In all these scenarios the group size ranged from 2–4 members. *Problem-solving* includes scenarios of solving a complex problem like maths or physics problems or solving a puzzle. *Engineering design* deals with designing a prototype while *design* can cover multiple tasks such as designing coursework, a website, a course, or a game. *Collaborative learning* scenario specifically implies that the goal of the task is learning. *Meetings* are gatherings of members to discuss and brainstorm about a task. Thus, there is an overlap between *brainstorming* and meetings as some meeting scenarios had brainstorming as a sub-scenario phase. Some other scenarios also contain overlapping articles because some collaboration scenarios include other scenarios as a part of different subphases in that scenario. For instance, Kim et al. (2008) had problem solving as the scenario which had two different subphases of brainstorming and meetings. This means that the scenario separation is solely based on the end collaboration target where each scenario need not be mutually exclusive but rather serve as a guide to distinguish the indicators of collaboration based on the collaboration end goal. *Gaming* mostly involves dyads (or pairs) who interact with their partner on a shared artifact. *Planning* is a session where group members plan a diet plan or some other day-to-day planning activity is undertaken. *Database design* uses interactive tabletops to design the database schemas. *Concept mapping* is the linking of similar concepts. *Healthcare simulation* involves surgeons or nurses during group operations or medical practice training. *Programming* involves working on code mostly in dyads. *Browsing* refers to a group who share information with each other to browse a website or other information.

Contextualization of different indicators

Considering the indicators detected in different scenarios, there are two broad categories of indicators. First, the verbal indicators grouped in content indicator type. Second, the nonverbal indicators grouped in gesture, posture, spatial, and eye gaze indicator type. The scenarios such as problem-solving, design, learning, meetings, brainstorming, planning, database design, concept mapping, and browsing use both the verbal and nonverbal category of indicator type to detect collaboration quality. But, the other nonverbal-heavy scenarios such as engineering design, gaming, programming, and healthcare simulation are action-based or require considerable interactions with a shared artifact along with the interaction among group members. So, depending on the goal of the task, context (such as the use of specialized furniture, TUI, other shared artifacts like a prototype, or patient manikins in case of medical simulation), the type of indicators detected changes. The relevance of these indicator types can be peeled further to determine whether they are always dependent on the context or they are independent of the context. For example, eye gaze indicator type for CC quality detection during a meeting scenario of a 3–4-member group is computed as the time of listener and speaker eye gaze while same eye gaze during a problem-solving, design task is computed as the JVA. Although speaker and listener eye gaze is not a good indicator of collaboration quality (Terken and Sturm, 2010), higher instances of synchrony in eye gaze (or JVA) is indicative of the better quality of CC (Dierker et al., 2009; Schneider et al., 2015). So, eye gaze is dependent on the context. However, synchrony in posture may not be a good predictor of collaboration quality (Schneider and Blikstein, 2015).

Besides, less distance between the collaborating group members means that they have a higher level of comfort and the quality of collaboration is better (Spikol et al., 2017b,a) for that group. However, this was that not consistent with all the previous works (Schneider and Blikstein, 2015). Schneider and Blikstein (2015) did not find any significant correlation between the group members' distance and collaboration quality.

Audio is a commonly occurring indicator type across most of the scenarios. Total speaking time (Bachour et al., 2010; Kim et al., 2008), interruptions (Oviatt et al., 2015), and overlap in the speech (Dong et al., 2009; Zhou et al., 2014; Çetin and Shriberg, 2006) were good predictors of collaboration quality. Some indexes such as synchrony of rise and fall in pitch, and equality of the amplitude are directly proportional to the quality of CC (Lubold and Pon-Barry, 2014). Silence has been used as a feature to train a machine learning model to detect the CC quality (Luz, 2013) in problem-solving. But, a qualitative analysis of these indicators was not done by the authors. One can interpret silence as a thinking or reflection stage when group members start thinking about the problem. Thus, it is difficult to inform the practitioners as to what the occurrence of single or multiple instances of silence can mean about the quality of CC. However, a balanced speaking time is desirable during meetings (Terken and Sturm, 2010; Kim et al., 2008) when every group member needs to speak in a discussion or contribute some ideas.

Besides, gestures detected by the hand interactions with the TUI showed that groups showed better quality of collaboration when they were focused on a particular purpose and had more occurrence of both task unrelated touches and unrelated overlapping sequences (Evans et al., 2016). It meant that they are working collaboratively towards the task objective instead of working individually. In general groups where members used both their hands (Schneider and Blikstein, 2015) and had near average hand movement (Worsley and Blikstein, 2015) showed better quality of collaboration. So, the gesture is dependent on the context. For example, a planning scenario (Barthelmess et al., 2005) has pointing gestures as an indicator whose higher number indicates better collaboration quality while grabbing mouse from the partner combined with active posture in case of programming scenario (Grover et al., 2016) is a sign of good collaboration. The valence of individual indicators' contribution for detecting CC quality was not discussed because it was operationalized using a machine learning classifier without any qualitative analysis.

Groups whose members' direction of arousal pattern of electrodermal activation was synchronous showed the good quality of collaboration measured in terms of learning gain (Pijeira-Díaz et al., 2016). However, it was not true all the time (Pijeira-Díaz et al., 2019). During collaborative learning, using electrodermal activation (EDA) in a group of 3 people collaborating, it was found that instances of arousal and relaxed states among the group members (or directional agreement) are not reached at the same time window (Pijeira-Díaz et al., 2019) even though they have a good collaboration quality. This directional agreement is context-independent. However, on contextualizing it with other modalities like the video data, it was found that the group members presented the most negative facial expressions during the simultaneous arousal episodes (Malmberg et al., 2019). Simultaneous arousal episodes (as measured by the EDA) occurred during different phases of CC. The quality of collaboration was poor during most of the instances of simultaneous arousal with a low level of interactions in the group. According to them (Malmberg et al., 2019), this arousal can be because of rising stress levels or confusion levels leading to unproductive collaboration.

As mentioned earlier, different tools and methods have been used for coding and analysis of these indicators. Some scenarios like brainstorming favor the use of human observers (or a human-based set up only) for detecting some of the indicators of CC while other scenarios like engineering design are better suited for the use of a sensor-based set up only or a hybrid set up for detecting CC indicators. Some scenarios such as *programming*, *planning*, and *database design* do not use indicators from audio indicator type while most other scenarios use indicators from audio indicator type. Connecting these indicators and indicator types in different scenarios with the sensors in Table 2.1, it is clear that these indicators define the type of set up needed for collaboration detection. In addition, these indicators vary a lot in different scenarios because of the differing goals and parameters.

Fundamental parameters during collaboration

To understand the scenarios further, we need to take into account the fundamental *parameters* of CC. The parameters of collaboration are primary aspects such as *team composition* (e.g., experts, initiators, or roles of being initiators), the *behavior of team members* (e.g., dominance, coupling, or conflict), *types of interaction* (e.g., active or passive, or critique), *behavior during collaboration* (e.g., knowledge co-construction, reflection, coherence, misconception, or uncertainty). To elaborate the parameters, *dominance* (Kim et al., 2008; Schneider and Blikstein, 2015) includes the dominance and leadership parameter. The *Coupling* (Olsen and Finkelstein, 2017; Lubold and Pon-Barry, 2014) includes the comfort level, coupling, coordination, and rapport between the group members. *Coherence* (Schneider and Pea, 2014a, 2015) includes verbal coherence where group members build upon each other,Äôs ideas and verbal-discourse coherence. *Engagement* (Cukurova et al., 2017b, 2018; Spikol et al., 2017b) includes engagement, participation, and interaction. *Learning strategy* (Worsley and Blikstein, 2015, 2018) reports the strategy adopted by the group during CC. *Heterogeneity* (Manske et al., 2015; Flood et al., 2015; Wong et al., 2011) refers to the difference in previous knowledge or the difference in capabilities among the group members. Fischer (2000); Fischer and Ostwald (2005) proposed that heterogeneous collaborating teams possess a symmetry of ignorance (Rittel, 1984) which makes them more interesting, wherein neither team possesses the full breadth of knowledge to solve the problem independently, and thus collaborating with each other can help to resolve the problem. *Roles* (another parameter of CC) (Wise et al., 2015; Martin et al., 2015), when self-assigned, evokes a sense of responsibility within the group that are designed to facilitate group progress towards a goal (Hare, 1994). Some roles are preassigned and some emerge during collaboration (Strijbos and Weinberger, 2010). Moving on to another parameter, *misconception*, it arises when group members share a common understanding among themselves without any refutation and reflection (Abdu, 2015). *Uncertainty* is more common among group members with less mutual understanding (Rodríguez et al., 2017). During design scenarios of CC, *critique* (McBride et al., 2017) is a parameter that surfaces for the first time where individual group members criticize each other's work to develop a shared understanding or to reach consensus. *Knowledge co-construction* (Scherr and Hammer, 2009; Oshima et al., 2017; Thompson et al., 2013) is sharing each other's ideas and using them to build the shared understanding of the situation or the task at hand.

2.3.3 Confluence of Both Approaches to Assess Quality of Collaboration

We focused on modeling the conceptual framework for the most dominant scenario found: CPS. We plan to model the quality of CC in some of the other scenarios in the future, but proceed for this one at the moment. It is because of the well-defined goals and objectives in this scenario based on the number of studies analyzed. Based on the scenarios we found for ideal collaboration and its parameters like team composition (such as experts, initiators), the behavior of team members (such as dominance, coupling), types of interaction (such as active or passive), we mapped these parameters onto the indicator types and indexes. This mapping defines a conceptual framework for the chosen CPS scenario. Table 2.4 gives an overview of this mapping.

Now, we drill down further into the CPS scenario in Table 2.4. If we consider one parameter dominance when taking into account audio as an indicator type, it can be detected by using the equality index in the group. Contrary to that, the same parameter can be mapped onto synchrony as a measurable index when posture is considered. In this case, as shown in the table with an upward or downward arrow (indicating the direct or inverse relationship with indexes and collaboration quality), lower dominance means higher synchrony or higher equality resulting in better collaboration quality. Similarly, less uncertainty among group members can be measured by better mutual understanding in the group resulting in a higher quality of collaboration. So, the fundamental characteristics of the group in one scenario (i.e., the parameters) are made visible by the proxy measurable property of the aggregated indicators (i.e., the indexes) to give an idea about the collaboration quality.

To summarize, first, in our review, we started with a bottom-up analysis. In that, we grouped different articles based on the sensors used, indicators of collaboration derived from these sensors, indexes formed by aggregating these indicators, and finally detecting the quality of collaboration. Next, we formed a scenario-driven

Table 2.4 Modeling a CC Scenario—Collaborative Problem-Solving

Parameters[a]	Indicator types	Indexes[b]	References
Expertise ⇑	Audio	Dialogue management∗	Luz (2013), Oviatt et al. (2015), Ochoa et al. (2013)
	Posture	Synchrony∗, IA∗	Ochoa et al. (2013)
	Gesture	Synchrony∗, IA∗	Ochoa et al. (2013)
	Writing	Synchrony∗, IA∗	Zhou et al. (2014), Ochoa et al. (2013)
Dominance ⇓	Audio	Equality	Kim et al. (2008), Bachour et al. (2010)
	Posture	Synchrony, Equality	Schneider and Blikstein (2015), Kim et al. (2008)
	Gesture	Synchrony, Equality	Schneider and Blikstein (2015), Kim et al. (2008)
	Writing	Synchrony∗, Equality∗	Zhou et al. (2014)
	Self-reports	Equality	Kim et al. (2008)
Coupling ⇑	Audio	Synchrony, Equality	Lubold and Pon-Barry (2014), Olsen and Finkelstein (2017)
	Posture	Equality	Schneider and Blikstein (2015)
	Gesture	Mutual understanding∗, Synchrony	Isenberg et al. (2010), Schneider and Blikstein (2015)
	Content	Mutual understanding∗, Synchrony	Isenberg et al. (2010), Schneider and Blikstein (2015)
	Self-reports	Mutual understanding∗	Isenberg et al. (2010)
Reflection ⇑	Self-reports	Information pooling∗	Anastasiou and Ras (2017)
	Audio	Information pooling∗	Thompson et al. (2014)
	Content	Information pooling∗	Thompson et al. (2014)
Roles ⇑	Audio	Mutual understanding∗, Task division∗	Kim et al. (2015)
Coherence ⇑	Audio	Reaching consensus∗	Schneider and Pea (2014a), Schneider and Pea (2015), Olsen and Finkelstein (2017)
	Eye gaze	Reaching consensus∗	Schneider and Pea (2014a), Schneider and Pea (2015)
	Content	Reaching consensus∗	Olsen and Finkelstein (2017)
Uncertainty ⇓	Audio	Mutual understanding	Rodríguez et al. (2017)
Misconception ⇓	Audio	Mutual understanding	Abdu (2015)
	Content	Mutual understanding	Abdu (2015)
Engagement ⇑	Audio	Equality∗, Synchrony	Bassiou et al. (2016), Viswanathan and VanLehn (2017), Nakano et al. (2015)
	Posture	Mutual understanding∗	Viswanathan and VanLehn (2017)
	Gesture	Mutual understanding∗	Viswanathan and VanLehn (2017)
	Content	Mutual understanding∗	Viswanathan and VanLehn (2017)
	Eye gaze	Synchrony	Nakano et al. (2015)
	Writing	Synchrony	Nakano et al. (2015)
	Spatial	IVA∗	Healion et al. (2017)

[a] ⇑ denotes that if the value of the parameter is high then the quality of the collaboration is better and vice versa, and ⇓ denotes that if the value of the parameter is low then the CC quality is better and vice versa. [b]Some indexes reported here have been detected practically, while some indexes marked with a ∗ have been reported by us based on our understanding of indexes from the article by Meier et al. (2007) and the practically detected ones.

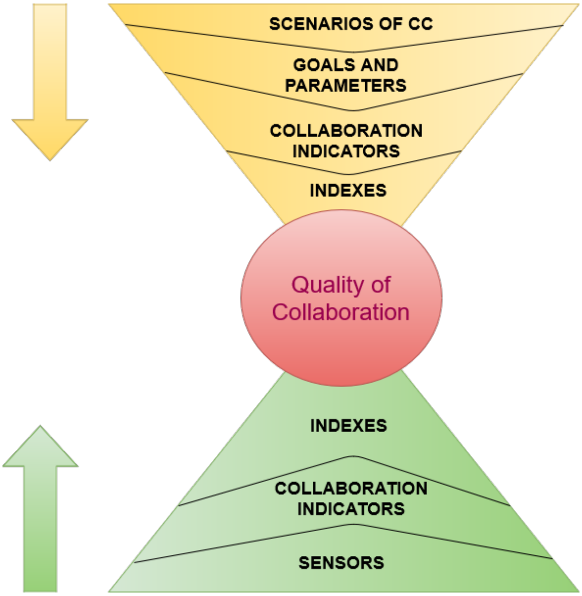


Figure 2.2 Confluence of both approaches of CC quality detection.

prioritization because of the differing goals of different types of CC scenarios, task requirements, and fundamental group parameters (such as dominance and coupling). From those scenarios, we mapped the parameters of CC onto the collaboration indicators and indexes for each of these scenarios. Fig. 2.2 shows the confluence of the analysis from both approaches. For the scope of the review, we restrict it to CPS which was the dominant scenario with well-defined goals and objectives as found in most articles. In CPS, if the group members are less dominant, then there is equality of total speaking time among the group members and their participation is almost equal, resulting in a good quality of collaboration (Kim et al., 2008; Bachour et al., 2010).

2.4 Discussion

Regarding the first research question ("What *collaboration indicators* have been used in research to understand the *quality* of CC?"), we have identified indicators for the quality of collaboration on two different levels. In the first part, we have identified categories of *low-level* sensor-based, human or hybrid events in collaboration that have been observed in different studies. We have collected indicators of collaboration that have been used in studies to identify relevant activities of users for the collaboration quality. In the second part, we have started from *high-level* indexes that have been used to identify collaboration quality in research. These indexes are composed of one or more indicators obtained from multiple indicator types and act as a proxy to detect the quality of the collaboration process. For instance, counting the number of ideas during a brainstorming scenario in CC is obtained from the events grouped in the content indicator type; while a high-level process definition, that is, equality of the number of ideas generated by each member in the group, measures the quality of collaboration. Thus, the *event–process* conceptual framework provides a holistic picture of the quality of collaboration observed during CC using MMLA. This conceptualization is an essential foundation stone for building different types of collaboration detection, monitoring, and prediction systems. We find that some of the indicators like total speaking time (Bachour et al., 2010; Kim et al., 2008), and number and duration of overlap of audio (Zhou et al., 2014) are consistently indicative of collaboration quality across different studies but the same is not true for other indicators such as distance between group members. The distance between group members gives a mixed indication of the quality of CC; that is, sometimes it is inversely proportional (Spikol et al., 2017a,b) or sometimes there is no relation (Schneider and Blikstein, 2015) with CC quality. The comprehensive overview of the indicators will help practitioners to choose the sensors and indicators according to their set up. If they see that certain indicators (such as writing speed, pressure from the digital pen, distance between group members, and space usage in the room during group work) from past studies are not having any relation with CC quality then they can focus on the indicators (such as total speaking time and JVA) that worked in most settings in their preliminary experiments.

The operationalization of these indexes has suffered from multiple limitations. Sometimes it is challenging to code the indicators to compute the indexes (Cukurova et al., 2018, 2017a) as in the case of individual accountability; thus failing to detect CC quality. Another limitation is the use of machine learning approaches (Grover et al., 2016; Luz, 2013; Stewart et al., 2018; Viswanathan and VanLehn, 2017) which use one or more indicators to detect CC quality but fail to address the qualitative aspect of these indicators. For example, silence and pause are good indicators of collaboration combined with other indicators (Luz, 2013) but it is not clear if more or less occurrence of silence in itself indicates anything about the quality of CC. This tension between the transparency of the learning analytics models and the accuracy was highlighted by Cukurova et al. (2020) and is still an open question. Some machine learning models which are like a black box have higher accuracy even though they are not transparent in terms of the role of each of the indicators of CC. Moreover, we find that some indexes have been detected from certain indicator types but not from others. For instance, synchrony has not been detected using the content indicator type. This may be because of the difficulty involved in detecting and analyzing the content of a discussion (or the semantic nature of the discussion itself) during collaboration. This also highlights the importance of choosing the right sensing mechanisms (or sensors) in the respective CC scenario. However, equality has been easily detected using the content indicator type (number of ideas as an indicator) as it is easier to measure a quantitative value (i.e., the number of ideas generated by each member during collaboration). This brings to light the need for scenario-driven prioritization and modeling.

Considering the second research question ("What is the impact of different *scenario-*

based goals and *parameters* for CC on the relevance of the different indicators?"), we found that the scenario of CC chosen has a huge impact on the indicators of collaboration obtained. Some scenarios have a stark contrast in terms of the collaboration indicators observed; for instance, collaborative brainstorming and collaborative gaming. However, some scenarios have certain overlapping collaboration indicators; for instance, collaborative design and collaborative concept mapping. This detection of scenario-based indicator types is also dependent on the use of external objects (e.g., patient manikins or shared artifacts). The scenarios which use these external objects tend to be inclined towards nonverbal indicator types (such as engineering design, gaming, and healthcare simulation). Moreover, some indicator types like eye gaze, gesture, and audio are dependent on context while some others like physiological ones are not. We find that higher occurrence of JVA (Schneider et al., 2015) measured from the eye gaze indicates better CC quality while the same is not true when individual eye gaze of speaker and listener is considered (Terken and Sturm, 2010). This indicates that CC is *scenario-dependent* and the collaboration indicators can vary depending on the scenario, its goal, and context. But, when we consider physiological indicator type then we find that instances of aroused and relaxed states are context-independent and can be misleading unless contextualized with other modalities like audio (Malmberg et al., 2019). Apart from the variation in the scenarios, groups also vary in their fundamental parameters like team composition (such as experts, initiators) or the behavior of team members (such as dominance, rapport) in CC. To understand the impact of these parameters on the indicators of collaboration in each scenario, we create a parameter-based listing and proceed for modeling the conceptual framework in some of these scenarios.

We have modeled a conceptual framework for one of the dominant CC scenarios which had well-defined task objectives (i.e., CPS). In this framework, we mapped the CC parameters (such as behavior, composition, interaction, etc., of group members) onto the indicator types and the indexes. We found that *mapping* the parameters helped in furthering the *semantic enrichment* of the parameters, highlighting the relevance of the indicators, and thereby defines a measurable complete set up. For instance, *dominance* as a parameter of CC can be mapped onto audio as an indicator type (taking into account the total speaking time indicator) to measure the equality index in the group; whereas the same parameter can be mapped onto synchrony as a measurable index when posture is considered. So, the same fundamental parameter, that is, dominance in this case, can be measured differently depending on the indicator type and the indexes considered for measuring the quality of collaboration. If a group has higher dominance then specific members are more dominant than others. This is measured by synchrony or equality. So, the higher the dominance, the lesser is the synchrony or equality and the worse is the quality of collaboration. Therefore, this conceptual framework is similar to a data dictionary which can act as a roadmap for future research and evaluation on CC quality. It gives a high-level overview of the current state to inform practitioners.

However, this mapping is incomplete. We find a scarcity in the operationalization of the indexes and a lack of well-defined task goals. This limited our conceptual framework design to only one of the dominant scenarios. To overcome this scarcity, we make use of the expected indexes that can be substituted based on our understanding from the theory and practice. Thus, there is an urgent need for practitioners (or teachers) to act upon the other theoretical indexes when monitoring collaboration quality in CC using multiple modalities. This can make more indexes from the theory visible in practice and lead us to define a measurable set up for each scenario. Nevertheless, the framework is a starting point for making design-based decisions of a particular scenario of CC so that more indexes can be added up to make it complete and strengthen the CC quality detection.

2.5 Conclusions and Future work

CC has acquired significant importance due to the ease of detecting collaboration from the universal use of sensors. In this study, we performed a literature review to look into the indicators that indicate the quality of collaboration from two different perspectives (i.e., from the sensors used to detect the indicators, then indexes, and thus the quality of CC, and from the different scenario-driven prioritization of CC to contextualize the quality indicators, indexes of CC). Our *goal* for this review was to use these quality indicators of CC from past studies and create a conceptual framework (or data dictionary) for practitioners and researchers to which they can refer whenever needed. To this end, we found different low-level *indicators* like hand movements, head movements, eye gaze, posture, and number of ideas. These can be grouped into different indicator types such as audio, posture, and gesture. Some indicators (such as total speaking time, and overlap in speech) are consistently indicative of CC quality while some others (distance between group members, synchrony in posture movements) are not. Next, we looked at the high-level *indexes* (comprising of synchrony, equality, IA, IVA, mutual understanding, information pooling, and reciprocal interaction) as the aggregated result obtained from the indicators of CC. Indexes describe the relationship between the different indicators considering the distribution of the collaborating group and act as proxy measurement criteria to detect and predict the collaboration quality. Moreover, the indexes of collaboration can be linked to some particular indicator types for detecting the quality of collaboration.

However, this understanding is incomplete unless we uncover the role of scenarios in detecting the indicators of collaboration and modeling CC. We find this in our scenario-driven prioritization mapping of CC parameters (such as behavior, interaction, etc.) onto the indicator types and the indexes to move towards designing a conceptual framework for modeling CC. This final confluence of both approaches of modeling collaboration quality (i.e., sensor-based and scenario-based) gives us a holistic picture of CC quality detection in a particular scenario. We find that when we analyze the indicators further in terms of the scenario-based goals and task context. Moreover, we find different limitations in the previous works such as inconsistent evidence provided by some indicators, coding complexity in open-ended tasks, and inconclusive evidence provided by some of the indicators because of the use of machine learning black-box approaches.

There are some limitations to this review. The conceptual model that we explain at the end by mapping the parameters in scenarios to the indicators and indexes is not complete, rather is only the starting point. We have plugged in some of the indexes marked with an asterisk in the Table 2.4 although we do not know if they will remain the same once they are operationalized. We did not model the conceptual framework for other scenarios due to a lack of sufficient operationalized indexes and task-based goals in those scenarios. This opens up future avenues of research if we can borrow from research on collaboration indexes in an online setting. For instance, previous works have detected different indexes during *remote or online collaboration* (from the eye gaze as an indicator) like reaching consensus, information pooling, and time management (Schneider and Pea, 2014b) (as outlined in Meier et al. (2007)) with the help of network analysis and graph theory. These works can provide us a fertile ground in a CC setting to uncover other indexes of collaboration that can drive the modeling. Moreover, some indexes have been operationalized in a handful of studies which brings into question their role in detecting the quality of collaboration on a larger scale.

Another limitation is that we did not look into different types of study (i.e., correlation vs interventionist) keeping in mind the scope of the review. This can open doors for another direction of future work. Our goal in the future will be to use the model of CC in the designed scenarios and then look into different feedback mechanisms that have been built using these indicators to facilitate collaboration. This combined with the indicators of collaboration quality can help us to derive the conceptual and implementation model to discover other indexes of collaboration. As a result of which, it will pave the way to form the feedback mechanism to facilitate collaboration in real time for a particular collaboration task.

Finally, we did not consider the number of groups used by different studies. We think this will be a good direction of future research even though it will be difficult to determine a threshold as to how many groups considered in a study will make it worthy of inclusion in the review. As per the title of the review article, we do not think we are there yet (i.e., the whole nine yards) because CC modeling is dependent on various factors as we have mentioned in the introduction, that is, the definition of collaboration and its quality is dependent on many factors like how it is operationalized, in what context, and the impact of culture. Thus, we have a made a starting step to model CC in one of the scenarios taking into account the indicators, indexes, and parameters but not considering the number of groups, or type of algorithms used.

Part II

Prototyping Automatic Collaboration Analytics Setup

Chapter 3

Towards Automatic Collaboration Analytics for Group Speech Data Using Learning Analytics

After the literature review on the indicators of collaboration quality and defining the conceptual model for collaboration quality detection and prediction system in Chapter 2, we move towards the prototyping of a technical set up for automated collaboration analytics. For doing this, we are using group speech data in this chapter taking the help of the definition of CC quality in Chapter 2. Most of the past works have used the audio modality to detect the quality of CC. The CC quality can be detected from simple indicators of collaboration such as total speaking time or complex indicators like synchrony in the rise and fall of the average pitch. Most studies in the past focused on "how group members talk" (i.e., spectral, temporal features of audio like pitch) and not "what they talk". The "what" of the conversations is more overt contrary to the "how" of the conversations. Very few studies studied "what" group members talk about, and these studies were lab based showing a representative overview of specific words as topic clusters instead of analysing the richness of the content of the conversations by understanding the linkage between these words. To overcome this, we made a starting step in this chapter based on field trials to prototype, design a technical set up to collect, process and visualize audio data automatically. The data collection took place while a board game was played among the university staff with pre-assigned roles to create awareness of the connection between learning analytics and learning design. We not only did a word-level analysis of the conversations, but also analysed the richness of these conversations by visualizing the strength of the linkage between these words and phrases interactively. In this visualization, we used a network graph to visualize turn taking exchange between different roles along with the word-level and phrase-level analysis. We also used centrality measures to understand the network graph further based on how much words have hold over the network of words and how influential are certain words. Finally, we found that this approach had certain limitations in terms of automation in speaker diarization (i.e., who spoke when) and text data pre-processing. Therefore, we concluded that even though the technical set up was partially automated, it is a way forward to understand the richness of the conversations between different roles and makes a significant step towards automatic collaboration analytics.

This chapter is based on:

Praharaj, S., Scheffel, M., Schmitz, M., Specht, M., and Drachsler, H. (2021). Towards Automatic Collaboration Analytics for Group Speech Data Using Learning Analytics. *Sensors*, 21(9), 3156, doi: 10.3390/s21093156.

3.1 Introduction

Collaboration is an important 21st Century skill (Dede, 2010). Basically, collaboration occurs when two or more persons work towards a common goal (Dillenbourg, 1999). The majority of the works in the field of learning analytics to generate collaboration insights have focused on the analysis of distributed (or online) collaboration (Jeong and Hmelo-Silver, 2010). However, with the ubiquity of the use of sensors (Grover et al., 2016; Kim et al., 2008), multimodal learning analytics (Blikstein, 2013; Praharaj et al., 2018a; Di Mitri et al., 2018b) has picked up pace, thus shifting the focus to the analysis of co-located collaboration (CC) (or face-to-face collaboration) with the help of sensor technology (Praharaj et al., 2018b; Kim et al., 2008; Praharaj et al., 2019; Tausch et al., 2014). Moreover, sensor technology is scalable (Reilly et al., 2018) and has become affordable and reliable in the past decade (Starr et al., 2018). CC takes place in physical spaces where all group members share each other's social and epistemic space (Praharaj, 2019). "The requirement of successful collaboration is *complex, multimodal, subtle,* and learned over a lifetime. It involves *discourse, gesture, gaze, cognition, social skills, tacit practices,* etc." (Stahl et al. (2013) pp. 1–2, emphasis added). Therefore, the quality of collaboration can be detected from one or more of the different modalities like audio, video and data logs. Audio is a commonly occurring modality during collaboration (Bassiou et al., 2016; Bachour et al., 2010; Bergstrom and Karahalios, 2007; Terken and Sturm, 2010).

The quality of co-located collaboration from group speech data alone has been detected in the past using indicators of collaboration derived from audio. The indicators of collaboration can be as simple as total speaking time of a group member (Bachour et al., 2010) or as complex as synchrony in the rise and fall of the pitch (Lubold and Pon-Barry, 2014). For example, Bachour et al. (2010) mirrored the total speaking time of each group member in the form of the number of glowing coloured LED lights for each group member on a tabletop display. Then, they found that group members who spoke more reduced their speaking time and group members who spoke less improved their total speaking time. In the other example, two members in a group were speaking at different amplitudes, but exhibiting the same pattern of their speech (e.g., the rise and fall of the average pitch of both members were similar to each other), then they showed a high level of synchrony (Lubold and Pon-Barry, 2014), which resulted in a better rapport and better quality of collaboration. Therefore, the collaboration indicators are context dependent. This is due 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), the *behaviour of team members* (e.g., dominance, rapport, conflict), the types of interaction (e.g., active or passive) and *behaviour during collaboration* (e.g., knowledge co-construction, reflection, coherence, misconception, uncertainty).

Most studies on CC in the past focused on automated analysis using temporal (time domain features like the energy of the signal, amplitude), spectral indicators of speech (frequency-based features like pitch, rhythm) (Lubold and Pon-Barry, 2014; Bassiou et al., 2016) and other *non-verbal indicators* like total speaking time (Bergstrom and Karahalios, 2007; Bachour et al., 2010), frequency of turn taking (Kim et al., 2015) or using machine learning classifiers to analyse these features of speech (Luz, 2013; Bassiou et al., 2016). Therefore, most of these studies focused on the analysis of the non-verbal indicators of audio instead of looking at the verbal audio indicators such as the content of the conversation, actual keywords used, dialogues and the main themes of conversation. These non-verbal audio indicators do not convey true meaning because most used black-box machine learning methods and some studies reported the indicators (e.g., silence is an indicator for collaboration quality (Luz, 2013)) without informing about the valence, i.e., how good or bad these indicators are. Moreover, the non-verbal audio indicators are less overt as compared to verbal audio indicators. For example, higher or lower total speaking time may be a good or bad indicator of collaboration quality, while "yes" or "no" will most of the time convey the same semantic meaning in any conversation. Few other studies have focused on the non-automated (or semi-automated) coding and analysis of the content of speech, which is laborious (Bassiou et al., 2016; Lubold and Pon-Barry, 2014).

Apart from the majority of studies focusing on the analysis of *non-verbal audio indicators*, very few studies used the *verbal audio indicators* or the content of the audio for the analysis of CC quality. For example, for "talk traces" (Chandrasegaran et al., 2019) and "meeter" (Huber et al., 2019), verbal audio indicators of collaboration were used for the analysis. In "talk traces", Chandrasegaran et al. (2019) did topic modelling during the meeting and then showed the topic clusters as a visualization feedback by comparing with the meeting agenda, which was fixed before the meeting. Moreover, topic modelling shows a surface-level 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, which can lead to the loss of the holistic meaning of the conversations and a possible under-representation of certain topics. The other "meeter" study (Huber et al., 2019) classified the dialogues of the group members based on a lab 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 min. Their performance (or quality of collaboration) was measured based on the number of ideas they wrote down on the cards, which was quality controlled before counting the total ideas to weed out bad ideas. They did not find significant effects of information sharing and shared understanding on the quality of collaboration. Therefore, these studies on verbal audio indicators of collaboration were too abstract in either choosing representative keyword clusters as topics or classifying dialogues into a few selected categories that do not affect the collaboration quality. They did not show the linkage of the conversation between different group members. Furthermore, these studies were performed in controlled settings. Therefore, to overcome these limitations, we conducted a field trial to build a technical set up and then prototyped it in real-world settings to advance towards automatic collaboration analytics from group speech data. To this end, we have the following overarching research question:

RQ: To what extent can co-located collaboration analytics from group speech data be done automatically?

To answer this primary research question, we sub-divided it into two sub-research questions:

RQ1 What co-located collaboration indicators have been detected from group speech data in past studies?

RQ2 What collaboration analytics can be employed to analyse group speech data from co-located collaboration automatically?

To answer *RQ1*, we look at the already available literature in Section 3.2. To answer *RQ2*, we designed a technical set up and report about the materials and methods used in Section 3.3. Our objective of building this technical set up was to analyse the "what" of the conversation in an automatic manner. We collected, processed and visualized audio data automatically. The data collection took place while a board game was played among the university staff with pre-assigned roles to create awareness of the connection between learning analytics and learning design. This game was also helpful to collect indicators for measuring student and teacher behaviour. We not only did a word-level analysis of the conversations, but also analysed the richness of these conversations by visualizing the strength of the linkage between these words and phrases interactively. Our main goal was to use this technical set up to do role-based profiling based on the exchange of conversation turns taking into account the content of the conversation. To analyse the content of the conversation, we generate meaningful visualizations and interpretations in Section 3.4. In this visualization, we used a network graph to visualize turn taking exchange between different roles (such as teacher, student and study coach) along with the word- and phrase-level analysis. We also used centrality measures to understand the network graph further based on how much words have hold over the network of words and how influential are certain words. Then, we discuss our findings in Section 3.5 and the challenges and limitations in Section 3.6. We found certain limitations of our technical set up in terms of automation in speaker diarization (i.e., who spoke when) and data pre-processing. Finally, we conclude with a highlight of the implications of this work and future work in Section 3.7. Therefore, the main reason for building this technical set up was to make a starting step towards automatic collaboration analytics, which can assist different stakeholders in a university to understand the group conversations in depth, do role-based profiling and analyse how each group member contributes to the discussion.

3.2 Indicators of Co-Located Collaboration from Audio

Audio is a commonly occurring modality during collaboration (Bassiou et al., 2016; Bachour et al., 2010; Bergstrom and Karahalios, 2007; Terken and Sturm, 2010). Indicators of collaboration derived from audio are: prosody of sound such as pitch, spectral property, tone and intensity (Bassiou et al., 2016); non-verbal features like total speaking time of group members (Bergstrom and Karahalios, 2007; Bachour et al., 2010), interruptions (Oviatt et al., 2015) and overlap or no overlap duration (Bassiou et al., 2016); speaking time of a group member combined with the attention of other group members measured by their eye gaze (Terken and Sturm, 2010); linguistic features such as pronouns, sentence length and prepositions (Schneider and Pea, 2014a, 2015); verbal features like the keywords used, topics covered (Chandrasegaran et al., 2019) and dialogues (Huber et al., 2019). It has been found that a combination of both group speech-based and individual speaker-based indicators is a good predictor of the collaboration quality (Bassiou et al., 2016). As seen from the examples in different past studies, these indicators of collaboration are dependent on the context. This is due 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), the *behaviour of team members* (e.g., dominance, rapport, conflict), the types of interaction (e.g., active or passive) and *behaviour during collaboration* (e.g., knowledge co-construction, reflection, coherence, misconception, uncertainty).

To elaborate further, Terken and Sturm (2010) designed a mechanism to give real-time feedback to participants in group meetings by analysing their *speaking time* and *eye gaze* behaviour. Feedback was given in the form of different coloured circles representing attention to and from speakers and listeners measured by eye gaze and the total speaking time of that member. This feedback was projected on top of the table in front of where each participant was sitting using a top-down projector. They performed both quantitative and qualitative evaluation of the effect of the feedback: the feedback was accepted as a positive measure by most group members; the use of feedback had a positive impact on the behaviour of group members as they had a balanced participation. There was a balanced participation in terms of the speaking time of each group member. It was found that the eye gaze measured to track the total attention of the listener and speaker was not a good predictor of the quality of collaboration. As per the authors, this was because of the difficulty in intuitively controlling gaze behaviour as compared to controlling the speaking behaviour even though both can be consciously controlled.

Some other works also used total speaking time as an indicator of collaboration (Bergstrom and Karahalios, 2007; Bachour et al., 2010). The participants were having a group conversation around a smart table. The total speaking time of each member was reflected back to them by a coloured LED light display (Bachour et al., 2010) and concentric circle visualization (Bergstrom and Karahalios, 2007) on the table. They found that this helped to regulate the equality of participation during a group conversation. The group members who spoke most of the time (or were dominant) started to speak less than usual, and the members who spoke less started speaking more, thereby promoting equality of participation among the group members. Therefore, the group that had better equality of speaking time had better quality of collaboration as measured by a post-test.

To analyse other audio indicators in depth, Bassiou et al. (2016) used non-verbal features as collaboration indicators. They used a combination of manual annotation and a support vector machine to predict the collaboration quality of the group. The types of collaboration quality marked by expert annotators were: good (when all three members in the group were working together and contributing to the discussion), cold (when only two members were working together), follow (when one member was taking the lead without integrating the whole group) and not (when everyone was working independently). This coding was based on two types of engagement: simple (i.e., talking and paying attention) and intellectual (i.e., actively engaged in the conversation). It was found that a combination of the *group speech activity* indicators (i.e., solo duration, overlap duration of two persons, overlap duration of all three persons, the ratio of the duration of the speaking time of the least and most talkative person in the group, the ratio of the duration of the speaking time of the second most talkative student to the most talkative student in the group) and *individual speaker-based* indicators (i.e., spectral, temporal, prosodic and tonal) were good predictors of collaboration quality as marked by the annotators. Moreover, the group-level indicators alone were good predictors of collaboration quality. According to the authors, this was because the individual speaker-based indicators were agnostic to the group information, contrary to the group speech activity indicators. All these indicators were fed to a machine learning classifier to get the measurements, so in the end, it was a black-box approach. They did not employ any fine-grained analysis, which could help to uncover the degree of contribution of different indicators to the prediction of good or bad collaboration quality.

Similarly, *speaker-based* indicators like the intensity, pitch and jitter were used to detect collaboration quality among working pairs (Lubold and Pon-Barry, 2014). 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 the average pitch of both members are similar to each other), then they are showing a high level of synchrony (Lubold and Pon-Barry, 2014). Lubold and Pon-Barry (2014) found a positive correlation between synchrony and rapport (generated by comparing perceptual rapport from annotators and self-reported rapport) during collaborative interactions. A good rapport between group members can enhance the collaboration (Chapman et al., 2005). The prediction gave a high-level overview of non-verbal features like pitch, but missed the fine-grained semantic meaning of different non-verbal features such as turn taking, emotional tone while speaking, cross-talk and number of interruptions. These fine-grained vocal characteristics such as turn taking and overlap of speech are distinctive of collaboration quality; more frequent speaker changes (i.e., *turn taking*) with overlap of speech (Kim et al., 2015) indicates a good quality of collaboration. Previous research also indicated that overlap in speech is associated with positive group performance (Çetin and Shriberg, 2006; Dong et al., 2009).

Additionally, other works focused on expertise detection and productive problem solving (Luz, 2013; Ochoa et al., 2013; Oviatt et al., 2015), estimation of success (Spikol et al., 2017b), collaboration detection (Viswanathan and VanLehn, 2017) and differentiating student learning strategies (Worsley and Blikstein, 2015) during collaboration. Oviatt et al. (2015) tracked the speech of students working in groups solving math problems. They found that *overlapped speech* was an indicator of constructive problem-solving progress, expertise and collaboration. They used both the *number of overlaps* in speech and the *duration of overlap* in speech when tracking the interruptions during speaking. Luz (2013) used the non-verbal audio indicators like speech, silence, pause and transition from group speech to individual speech as indicators to predict performance and expertise on a math dataset corpus of groups collaborating in solving math problems. 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. They did not do any analysis to find the valence of these individual audio indicators. Spikol et al. (2017b) 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 smart furniture. They found that audio level alone was sufficient to predict the quality of collaboration with high accuracy. A binary coding classification for collaboration quality was used instead of a richer set of fine-grained level of coding. Again, a deep qualitative analysis of how audio level contributed to the detection of collaboration quality was missing. Table 3.1 gives an overview of some of the studies on detecting the indicators of collaboration from audio and their operationalization.

All the above studies analysed the non-verbal audio indicators (such as total speaking time, number of interruptions while speaking, overlap of speech) instead of the verbal audio indicators of collaboration. Non-verbal audio indicators of collaboration are less overt as compared to verbal audio indicators. Semantically, the content of the conversation, i.e., the verbal audio indicators of collaboration, have the same meaning most of the times.

With the rise of automatic speech recognition techniques, few studies (for example, "talk traces" (Chandrasegaran et al., 2019), "meeter" (Huber et al., 2019)) took into account verbal audio indicators of collaboration. In "talk traces", Chandrasegaran et al. (2019) did topic modelling, then showed the combination of words as topic clusters and also compared it with the meeting agenda. Although topic modelling shows a representative overview of the different word clusters and their evolution during collaboration, it does not show the link between these words and the rest of the conversation, which makes it hard to understand the meeting as a whole. In "meeter" (Huber et al., 2019), the dialogues of the group members were categorized

Table 3.1 Indicators of collaboration and its operationalization.

Parameters	Indicators	Operationalizing Collaboration Quality	References
Dominance	Total speaking time	If all group members speak for almost equal total time, then there is less dominance in the group and better quality of collaboration	Kim et al. (2008); Bachour et al. (2010); Bergstrom and Karahalios (2007)
Active participation	Frequency of turn taking	More frequent turn changes indicate higher active participation and better quality of collaboration	Kim et al. (2015)
Roles (one leader and other non-leaders)	Keywords used, topics covered	Closeness of the topics generated in real-time to the topics on the meeting agenda	Chandrasegaran et al. (2019)
Rapport	Synchrony in the rise and fall of the average pitch	Higher synchrony in the rise and fall of the average pitch indicates higher rapport and better collaboration quality	Lubold and Pon-Barry (2014)
Expertise	Overlapped speech	Overlap in speech is an indicator of constructive problem solving, expertise and good CC quality	Oviatt et al. (2015); Zhou et al. (2014)

based on a collaborative task (i.e., brainstorming and sharing ideas on open-ended fixed topics in short 10 minute sessions) in a lab setting to measure information sharing and shared understanding. The number of ideas generated was an indicator of collaboration quality. They did not find any significant effects of information sharing and shared understanding on the quality of collaboration. Therefore, these controlled lab studies analysing the content of the conversation provided a high level representative overview of few topics or categories without showing the relationship between the different group members based on their conversations. Therefore, to overcome this, we describe the prototyping of our technical set up with the help of field trials in a real world setting.

3.3 The Technical Set up

In this section, we describe the technical set up: tasks undertaken, their context, architecture of the set up, data collection, pre-processing, processing and methods used for the data analysis.

3.3.1 Task Context

The collaboration task that we used as the basis for our audio recordings was to design a learning activity using the Fellowship of Learning Activity ((FOLA)^2) (`http://www.fola2.com/`, last accessed on 30th April 2021) game. It is a board game (Schmitz et al., 2019) (e.g., an online version of the game (`https://game.fola2.com/`, last accessed on 30th April 2021) currently under development) played face-to-face with different themed cards and roles that is used in workshops to create awareness of the connection between learning analytics and the learning design. It also can 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 meetings (with each meeting spanning between 60 and 90 min) among different teaching staff and other staff of a university. This task had different phases, which were colour-coded based on the cards supposed to be used in that phase (as *blue, red* and *yellow*) (the phases and cards have the same meaning and are being used interchangeably henceforth) with different roles assigned to each member. The *blue* (card) phase defines the steps in the learning activity. Each learning activity consists of a sequence of interactions such as learner to teacher, learner with learning environment, material to learner and so on. The *red* phase 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, etc. The *yellow* phase defines what we want to know about the interaction or within the learning activity. For example, it can be engagement, social interaction or how students take initiative. The yellow cards can be used to get input on what teachers do in classrooms to value their design choices or actions. Each card also had some prompts to steer the group conversation.

We recorded the conversations during these meetings. The conversations were in Dutch. Each group member was pre-assigned roles during the conversation: *study coach*, *student*, *technology-enhanced learning learning analytics (TEL LA) advisor*, *game master*, *educational advisor* and *teacher*. These roles had the same meaning as a real-life student, teacher or advisor, while the game master was the main moderator of the game who also helped to steer the conversation during the task. Each group member had a clip-on microphone attached along with the respective audio recorder, which recorded and stored the conversation locally in that recorder. Next, we outline the architecture used for data collection, processing and analysis.

3.3.2 Architecture

First, the audio files from each group member were saved into a storage space in the respective local device, i.e., the audio recorder. Then, after the meeting, these files were immediately transferred to the central storage space, which was the long-term storage. For the pre-processing and subsequent operations on the data, we did not disturb the original data collected, rather we took a copy of the files in the storage space for the pre-processing and processing unit. Here, we pre-processed and transcribed these audio files using Google speech-to-text. Finally, the data were processed and analysed to generate meaningful insights and passed on to the visualization unit to generate the visualizations. These visualizations were generated in a post hoc manner after the group meetings. Figure 3.1 shows the outline of the current architecture for collecting and analysing audio data during CC. In the subsequent sections, we describe the pre-processing, processing and analysis.

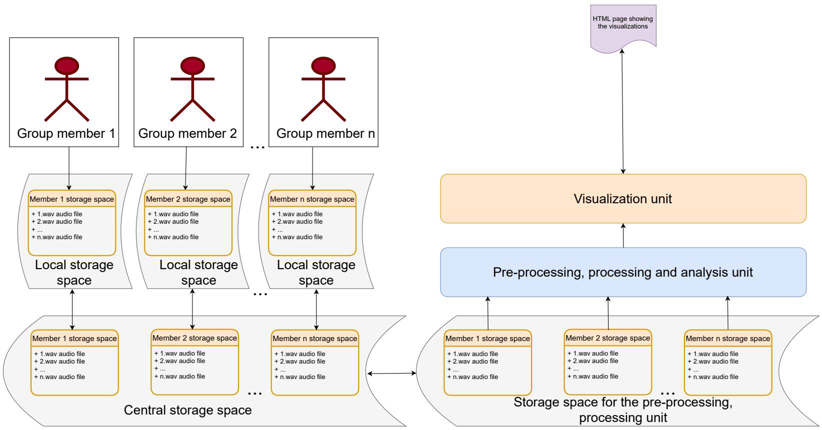


Figure 3.1 Architecture for collecting and analysing audio data during CC.

3.3.3 Data Pre-Processing and Wrangling

The data pre-processing, processing, analysis and visualizations were done in Python using different openly available libraries. We pre-processed the stored audio files for each group member by extracting the timestamp of the audio file, which denoted the exact end time of that audio file (in .wav audio file format) and the duration of the audio file. Then, we derived the start time of the audio file using the end time and the duration of the file. Next, we associated the conversations in each audio file with the group member playing a certain role, thereby associating with the group member who was speaking using *VoxSort Diarization* (`https://www.voice-sort.com/`, last accessed on 30th April 2021) software. This is known as speaker diarization, which helped us figure out "who spoke when?". The group members playing specific roles were anonymized by mapping their names to roles, and these audio files were combined into one file. Finally, we transcribed this audio file using Google speech-to-text (free version) in Python. For this transcription, we split the audio file into 5 s window chunks (with a small overlap between the adjacent windows), which worked the best in our case. This helped us to increase the accuracy of the transcription, which was otherwise not good enough when we used long files of 1 to 30 min in duration. This process was repeated, and finally, these chunks were combined, sorted by timestamps from the beginning till the end of the meeting. While performing these subsequent steps, the information like timestamp and speaker role was put into the data tables in a .csv file format. Figure 3.2 shows the schematic overview.

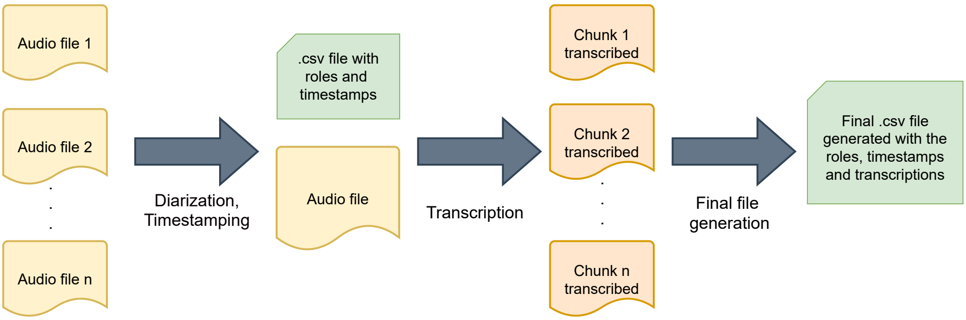


Figure 3.2 Data pre-processing schematic overview.

Then, the final extracted data along with the timestamps were stored in data tables in a .csv file format, as shown in the Figure 3.3. Basically, the data table had the *start time*, the *end time*, the roles of the group member under *names* (256 under names denotes noise) and the utterance as *text_y*. This data table represented the ordered conversation from the beginning till the end of the meeting. If Google speech-to-text failed to transcribe an audio file, then the corresponding text entry was left blank. Normally, this happened when a part of the audio was of a really short duration or had some random sounds like a click sound, um, claps or laughter. Now, the wrangled (cleaned, structured and enriched raw data in a desired format) data were ready for processing and analysis.

A	B	C	D	E
	Start time in millisecond	End time in milliseconds	Names	Text_y
0	0	1759		
1	1760	1919	256	
2	1919	2079	256	
3	2080	23599	Game Master	Nou de voorbereidingen gedaan dan Ja zijn allemaal willekeurig Ja ik heb ook niet zoveel bij iemand iets neergelegd Ik heb even gezorgd dat de student en docent dat die al
4	23600	23679	256	
5	23680	29759	Game Master	hebben ze dat makkelijk voor jullie liggen ze daar ook ja gewoon
6	29760	29839	256	
7	29840	30639	Game Master	
8	30640	30719	256	
9	30720	37279	Game Master	Google dahua wel een hoesje omheen zit mag ook echt opschrijven en stiften
10	37280	37359	256	
11	37360	54319	Game Master	watervaste en whiteboard Marker stift okÃ© stop doe ik eerst even met je moet ik eerst even met jullie over praten hebben we het vandaag over discussie College van we
12	54320	55599	TELI LA Advisor	onderuit
13	55600	56879	Educational Advisor	Varsseveld
14	56880	59039	Game Master	misten
15	59040	61599	TELI LA Advisor	ja ja
16	61600	61679	Game Master	
17	61680	61759	256	
18	61760	70079	Game Master	individueel activiteiten
19	70080	74479	Teacher	en wat
20	74480	76319	Game Master	staat ook achter op het bord
21	76320	79759	TELI LA Advisor	de vier onderwerpen die in dat
22	79760	112559	Game Master	College langskomen zijn de onderwerpen 10 Focus teamrollen zijn is het onderwerp je eigen piemel verder ervaring in het kader van de belbin rollen en belde rollen belde is
23	112560	112639	256	
24	112640	134479	Game Master	beschrijving Bel Mauritsschool zoveel op dat onderdeel voor het onderdeel zo voordat onderdelen onderdelen dus dat is eigenlijk het tweede onderwerp ervaring wordt nc
25	134480	134559	256	
26	134560	135199	Game Master	

Figure 3.3 A sample from the data table in .csv format.

3.3.4 Data Processing

The data table stored as a .csv file was processed. Our primary focus was to analyse the content of the conversations to generate meaningful insights. To do this, we proceeded with text cleaning, processing and analysis, which come under the umbrella term of *natural language processing*. The usual approach in natural language processing is to first to clean the text. Next, we built the text model from the conversation corpus. We had to make sure that our text model could understand similarities and also understand when two different words meant similar things. Therefore, the following steps were taken by us in order to achieve this cleaning:

- Tokenization—The process of splitting sentences into good words or tokens. It lays the foundation for the next steps of cleansing.
- Elimination of stop words—The process of removing words that mean little; these are usually words that occur very frequently. Apart from using the libraries in Python for stop word removal, we also defined our list of contextual stop words that were considered unimportant for this model.
- Lemmatization and stemming—Lemmatization and stemming convert a word into its root form. For example, for the words running and runs, the stem of both words is run. Thus, after we stemmed, these words would be grouped together and retain same meaning for the model even though they had different forms.
- Sentence segmentation—We split the unstructured spoken text into different sentences, which helped the model understand the boundaries of the long text to make it more semantically distinct.
- Vectorization—Since we cannot input plain words into a model and expect it to learn from it, we had to vectorize the words. This basically means creating unit vectors for all words. As the machine can understand numbers only, so the vectorized version of words will create a dictionary for the model, which would be useful later while generating bigrams (two word combinations appearing together), trigrams (three word combinations appearing together) and topic modelling based on the keywords.

After cleaning, we proceeded with the analysis and visualizations. For the analysis, we first visualized an exploratory view of the keywords used in different phases of the CC task by different roles with the help of the cleaned text model to understand the content of the conversations. Then, we proceeded with the detailed analysis with the help of richer visualizations to understand the content and context of the conversation further with the help of our technical set up. In the subsequent sub-section, we describe the materials and methods used for the data analysis.

3.3.5 Data Analysis

For the scope of this article, where we describe a proof-of-concept to prototype the development of our technical set up towards automatic collaboration analytics for group speech data, we restricted our analysis to only the first out of the 14 meetings. First, we visualized in an exploratory manner to see the frequently used keywords in the text model by different group members playing different roles by using frequency analysis. To make sense of the visualizations and understand the context of the conversations, we took the help of the game master to generate summarized annotations for each phase in English (for example, a sample annotation can be seen in Figure 3.4).

Blue	List of students and their role card (Leaner -> Mat	**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 dis**
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.
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. Smar
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 asl
Red	Mobile phone is added, Concept mapping tool tak	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 empt
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

Figure 3.4 A sample annotation of the CC task.

Then, to get a representative overview of these keywords, we examined the topical clusters obtained by using LDA (latent Dirichlet allocation) and LSI (latent semantic indexing) in different phases of the meeting session. LSI helped us to identify the coherence score based on which we decided the ideal number of topics in that phase, and then, we used LDA (which is a probabilistic approach of topic modelling) in multiple iterations to find these topical clusters. LDA and LSI just show the representative keyword collection and are unsupervised algorithms for topic modelling, which can cluster semantically similar words without the need for user labelling (or input).

To go in depth into the representative overview of these words in relation to the word exchange between different roles, we looked at the different bigrams (consecutive two-word phrases) and also ranked them based on the tf-idf (term frequency-inverse document frequency) ranking. tf-idf ranking of the bigrams helped to give an overview about the frequently (with a lower tf-idf ranking) and rarely (with a higher tf-idf ranking) used bigrams. Next, we wanted to see the relationship between these words and phrases with the help of parts-of-speech tagging and construction of knowledge graphs. Knowledge graphs show the relationship between the subject, the object with the verb or the verb phrase linking them.

Knowledge graphs (as shown in the visualizations section) are sometimes difficult to interpret because of the inaccurate sentence segmentation of unstructured data. Moreover, they also do not show the strength of different words (i.e., how often these words have been used) and the strength of linkage between these words. To show this, a co-occurrence matrix was made. A co-occurrence matrix shows how many times the words co-occur in the same sentence. For example, in the two sentences: "I love riding bike" and "Bike ride is loved by many", the co-occurrence matrix after doing text pre-processing (where we tokenized the sentences, removed stop words like "is", "by" and lemmatized and stemmed "loved", "riding") would be as in Figure 3.5. Therefore, all the tokenized words in the text corpus are listed in rows once and again in columns, and then, the value in the co-occurrence matrix shows the number of times each word co-occurs with the other word in one sentence. Therefore, in this example, “love bike ride” is a strong combination, which is evident from the co-occurrence matrix. Machines understand this co-occurrence matrix as it shows the strength between words with the help of numbers.

	I	love	ride	bike	many
I	1	1	1	1	0
love	1	1	2	2	1
ride	1	2	1	2	1
bike	1	2	2	1	1
many	0	1	1	1	1

Figure 3.5 A sample co-occurrence matrix.

Then, we visualized this co-occurrence matrix using social network analysis or the network graph. In this network graph (as shown in the visualizations section), each word from the text corpus can be shown as a node, and the edges between these nodes denote the strength between these words like how often they co-occur in the same sentence. To make the network graph visualization easier and intuitive, we built an interactive feature, which helped to highlight a specific node and its neighbours in the graph by selecting that specific node. To analyse the network graph in depth, we also 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 number of times a node lies on the shortest path between other nodes. For explainability, 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 (or word) is eigenvector centrality. Therefore, a node with a high eigenvector centrality score must be connected to many other nodes who themselves have high scores. In the next section, we describe the visualizations generated by using these data analysis methods in the context of the first session.

3.4 Visualizations

First, we did an exploratory visualization using this technical set up to see the frequently used keywords in different phases by different roles. As described in the Task Context Sub-section above, in the blue phase, the main objective was to discuss the steps in the learning activity. Each learning activity consisted of a sequence of interactions such as learner to teacher, learner with learning environment, material to learner and so on. Figure 3.6 shows the frequency of the words used in different utterances (or spoken segments) with the roles. “Team”, “groep” and “groepjes” in Dutch mean “team”, “group” and “groups”, respectively, in English. The lemmatizer for Dutch language did not work as expected for all the words, so “groep” and “groepjes” were separate, and there were a few words like that that needed manual tweaking. The main conversation in this phase was about groups or centred on groups. The teacher and TEL LA advisor spoke mostly about groups. “Vraag” and “test” actually comprise a card “vraag test” played for the interaction between teacher and learner, which means “question test”. “Belbin” roles comprise a card for the interaction between material and learner. Belbin team roles are actually nine different team role behaviours that make a high performing team. “Blok”, “1”, “2” actually refer to the Block 1 and Block 2 cards played for the interaction between learner and learner.

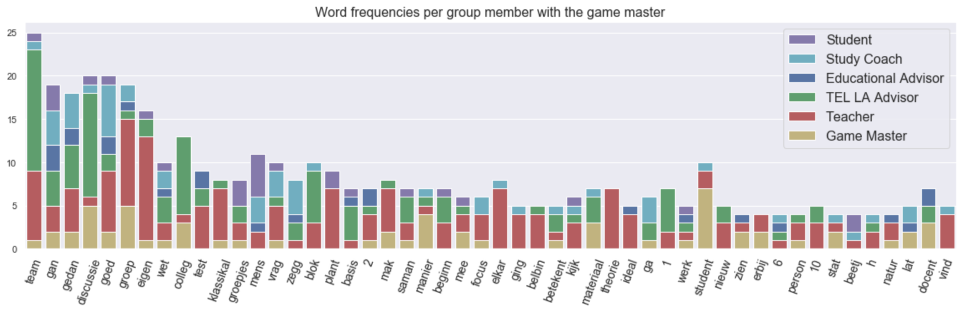


Figure 3.6 Top 50 word utterance frequency in the blue phase with roles.

The red phase as defined earlier was supposed to be a conversation about learning enhancing technology. Figure 3.7 shows the frequency of the words used in different utterances (or spoken segments) with the roles. The Dutch term “technologie” refers to the use of technology as was desired to be found in this phase. Furthermore, TEL LA advisor fulfilled the role quite well by being the sole speaker on technology apart from the game master, who was always present in most of the discussions because of his moderating nature. Some words that related to technology or its usage were: “moodl”, “poster”, “concept”, “mapping”, “mobil” and “shakespeak”. “Moodl” refers to moodle for the assignment. The concept mapping tool was referred to by the study coach and the educational advisor, and “shakespeak” was an interaction polling system used for interactive lectures in the classroom, which can act as an interaction booster. This was referred to only by the TEL LA advisor. The use of the mobile (“mobil”) phone to take a picture of the post-its (i.e., a paper sticky note) was discussed in this phase.

The yellow phase as defined earlier was supposed to be a conversation on the interaction within the learning activity and aspects a teacher might want to know about them. For example, it can be engagement, social interaction or how students take

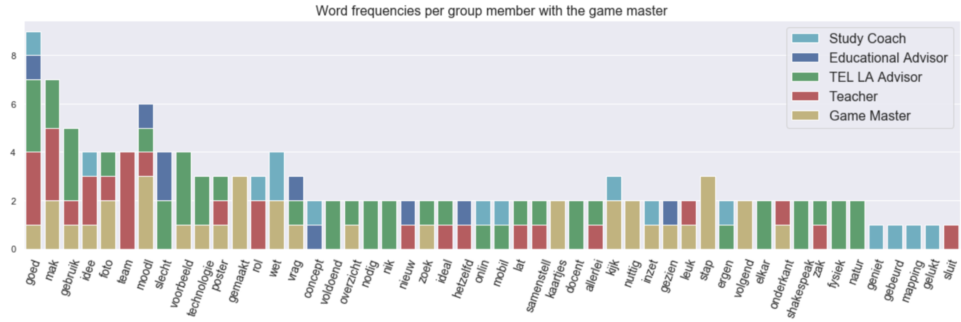


Figure 3.7 Top 50 word utterance frequency in the red phase with roles.

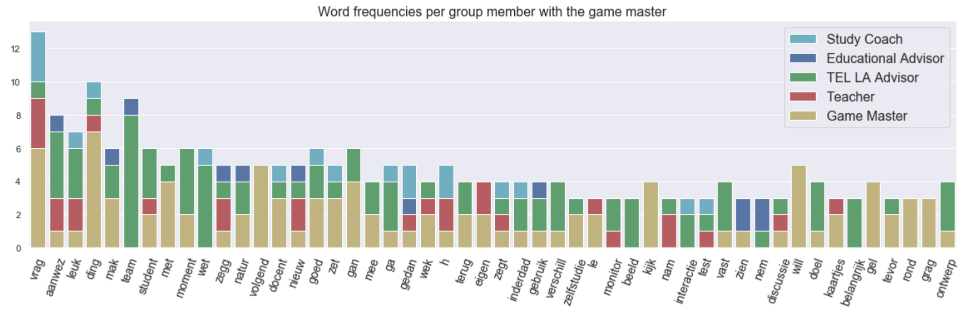
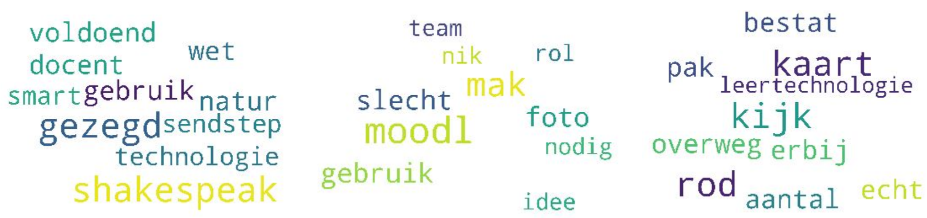


Figure 3.8 Top 50 word utterance frequency in the yellow phase with roles.

initiative. Figure 3.8 shows the frequency of the words used in different utterances (or spoken segments) with the roles. “Interactie” refers to the “interaction”, and “aanwez” means “presence”, which was at the top of the word utterance frequency because there was a specific discussion on presence and having fun, as we can see in the annotations. “Zelfstudie” and “monitor” mean “self-study” and “monitor”, respectively. They were referred to by the TEL LA advisor and the teacher during the conversation about student to material interaction.

To get a representative overview of these keywords, we examined the topical clusters obtained by using LDA and LSI in the red phase, which had a technological underpinning. We chose to go deeper into this phase because of our inclination towards technology. Figure 3.9 shows the overview of the three topical cluster word clouds (with each word cloud consisting of the top 10 probable words) obtained in the red phase. As LDA and LSI just show the representative keyword collection and are unsupervised algorithms for topic modelling, we needed to label it to assign a meaning out of each cluster. Upon examining the probabilistic inclination of the topics, we found that TEL LA advisor had a higher probabilistic likelihood of getting Topic 1 as compared to other roles. Topic 1 dealt with the use of different types of interaction technology as discussed in this phase. These were mainly evident from the words: “technologie”, “shakespeak”, “sendstep” and “smart”. These technologies were to be used by the teacher while interacting with the learner, which was evident from the word “docent”, which means “teacher” in English. Therefore, some of the words in the cluster when compared with the annotations gave the meaning of the topical theme. Similarly, Topic 2 can be elaborated based on “team”, “foto”, “rol”, “moodl” and “slecht”. Topic 2 refers to the use of moodle for assignments, making a photo of the post-its using the phone. This topic cluster also captured bad (“slecht”) teams, ideas and overview roles (“rol”) per student. The last topical cluster, Topic 3, mostly focused on the use of red cards (“rod”, “kaart”) and learning technology (“leertechnologie”).

When we analysed the turn taking of different roles during the red phase, we found that the TEL LA advisor and the teacher had the most exchange of turns between them. Therefore, to further explore their roles and the usage of the words, we analysed the bigrams (two consecutive word combinations) of the words. Tables 3.2 and 3.3 show the bigrams ranked from high to low tf-idf ranks and low to high tf-idf ranks, respectively. The tf-idf ranking tended to give a higher rank to bigrams that were used rarely and low ranks to bigrams that were used often. Therefore, from the tables, we can observe that “smart shakespeak”, “fysiek elkaar” and “goed powerpoint” were some of the top-ranked bigrams because they occurred rarely, and likewise, we also observed the low-ranked bigrams, which occurred frequently. This summarized the technology-related topics that were supposed to be discussed and also looked similar to the above topics computed by LDA. Similarly, Tables 3.4 and 3.5 show the top-ranked and bottom-ranked bigrams respectively based on the tf-idf ranking for the teacher.



(a) Topic 1 **(b)** Topic 2 **(c)** Topic 3

Figure 3.9 Topic clusters as word clouds in the red phase.

Table 3.2 Bigrams of the TEL LA advisor with high tf-idf ranking (i.e., bigrams rarely used).

Phrases (Original in Dutch)	Translated into English
smart shakespeak	smart shakespeak
fysiek elkaar	physically each other
goed powerpoint	good powerpoint

Table 3.3 Bigrams of the TEL LA advisor with low tf-idf ranking (i.e., bigrams frequently used).

Phrases	Translated into English
mobile phone	mobile phone
poster dieter	poster dieter
phone gebruiken	phone use
maken poster	make poster
gebruiken foto	use photo

Table 3.4 Bigrams of the teacher with high tf-idf ranking (i.e., bigrams rarely used).

Phrases	Translated into English
zekering interaction	certain interaction
foto maken	make photo
blok boos	block angry

Table 3.5 Bigrams of the teacher with low tf-idf ranking (i.e., bigrams frequently used).

Phrases	Translated into English
mindmap maken	make mindmap
maken posters	make posters
posters rol	posters role
samen denkt	think together

We wanted to see the relationship between these words and phrases used by the two speakers (with the help of knowledge graphs as in Figures 3.10 and 3.11) between whom most turn taking happened, i.e., the TEL LA advisor and the teacher. The green nodes show the subject and object, and the red links are the verbs or verb phrases. Although this is an interesting way to show the relationship between spoken text, it sometimes was difficult to understand the knowledge graph because of the accuracy of the sentence segmentation in the text corpus. Furthermore, we did not necessarily see the strength of the words (i.e., how often the words had been used) and the strength of the links (i.e., how often the two- or three-word phrases had been used) between these words. Therefore, we moved to the construction of a co-occurrence matrix that showed the strength of the words and also the links between words.

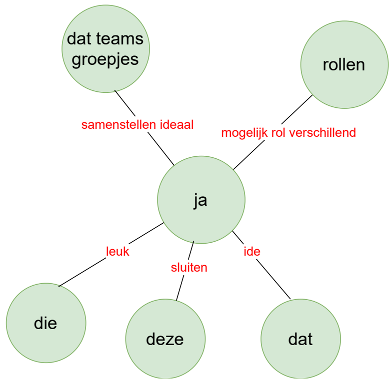


Figure 3.10 Part of the knowledge graph of the teacher in the red phase (zoomed in).

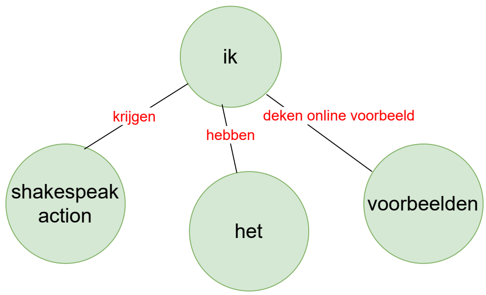


Figure 3.11 Part of the knowledge graph of the TEL LA advisor in the red phase (zoomed in).

When there is a huge text corpus, then visualizing these relationships from a co-occurrence matrix is easier when it is displayed as a social network by using graphs with nodes and edges (as in Figure 3.12), where each node shows the word and its frequency reflected by the node size and the link between the nodes, i.e., the edges show the strength of the words co-occurring in the same sentence as the edge thickness. This visualization is interactive where we can select a node in the graph and highlight that node along with its neighbours. This will be helpful to get an overview of the richness of the conversations and the interaction patterns of different roles.

Moving a step further, we show a portion of the graph where the node size is proportional to the betweenness centrality, which is a better measure than the frequency of the word. Betweenness centrality shows how often a node (or word) acts as a bridge node (or a node that has more control over the network). Another centrality measure that can be a good indicator of the influence of a node (or word) is eigenvector centrality. Therefore, a node with a high eigenvector centrality score must be connected to many other nodes who themselves have high scores. Table 3.6 shows an overview of the comparison of the word frequency in each utterance and different centrality measures for the red phase. Figure 3.13 shows the connection between some words in the network graph where the node size is proportional to betweenness centrality value of that node. “Goed” had the highest betweenness centrality, and upon highlighting it, it can be seen that it is connected to “team”, which has the second highest betweenness centrality. The connection value between them is one, and the connection value between “goed” and “poster” is two. Therefore, good and poster co-occurred more in a sentence than good and team in the red phase. Out of that, “good” and “poster” as words were used by the TEL LA advisor, and “team” was not used by the TEL LA advisor at all. From Figure 3.7, it is clear that “team” was used by the teacher only in the red phase and by no other roles. Therefore, the purpose of these examples was to show that these graph networks can be a useful way to visualize the word importance, strength and usage by different roles during collaboration.

Table 3.6 Top 5 words with frequency-wise ordering, betweenness centrality (BC)-wise ordering and eigenvector centrality (EC)-wise ordering in the red phase in decreasing order. The English translation of the Dutch processed words is in the brackets.

Frequency	**BC**	**EC**
goed (good)	goed (good)	mak (make)
mak (make)	team (team)	poster (poster)
moodl (moodle)	gebruik (use)	goed (good)
gebruik (use)	technologie (technology)	rol (role)
idee (idea)	rol (role)	allerlei (all kinds of)

If EC is seen in Table 3.6, then “mak” and “poster” are the two words with the highest values of EC, which means that they are influential. If we refer back to the frequently occurring bigrams in Tables 3.3 and 3.5, then “maken poster(s)” was one of the common ones for both roles who occupied most of the conversations in the red phase as computed from the frequent exchange of turns between them. This could be one of the reason for the high EC value. “Technologie” (“technology”) is in the top five of the betweenness centrality values in Table 3.6 even though it was not in the top five of frequency or EC, which was not surprising. This was because this red phase was about technology, and it was certain that the keyword technology would have more control over the network of words. Therefore, viewing the words, connecting words and the strength between them from the perspective of centrality could be interesting to discover latent relationships in the spoken conversations.

Thus, in this study, we took a computational approach for prototyping our technical set up and advancing towards automatic collaboration analytics.

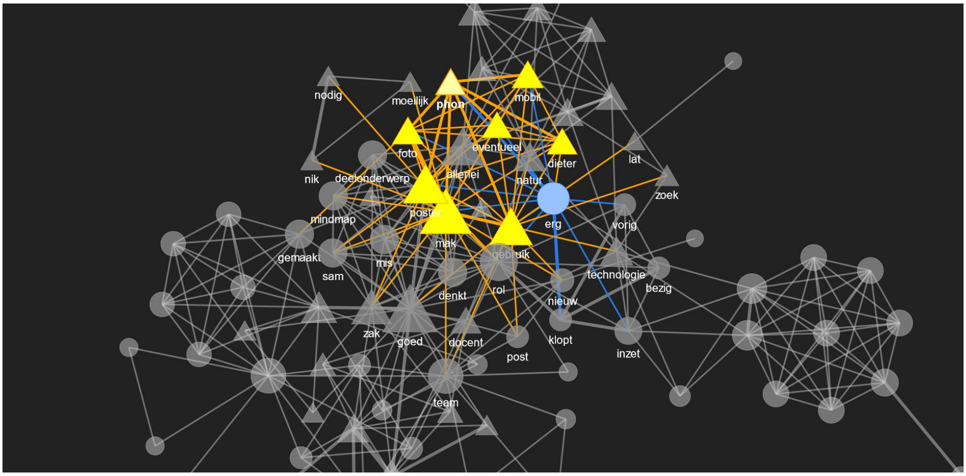


Figure 3.12 A sample social network (or network graph) of the words of the TEL LA advisor (shown as rectangles in yellow when highlighted) along with the whole red phase conversation (all other roles are shown as circles in blue when highlighted).

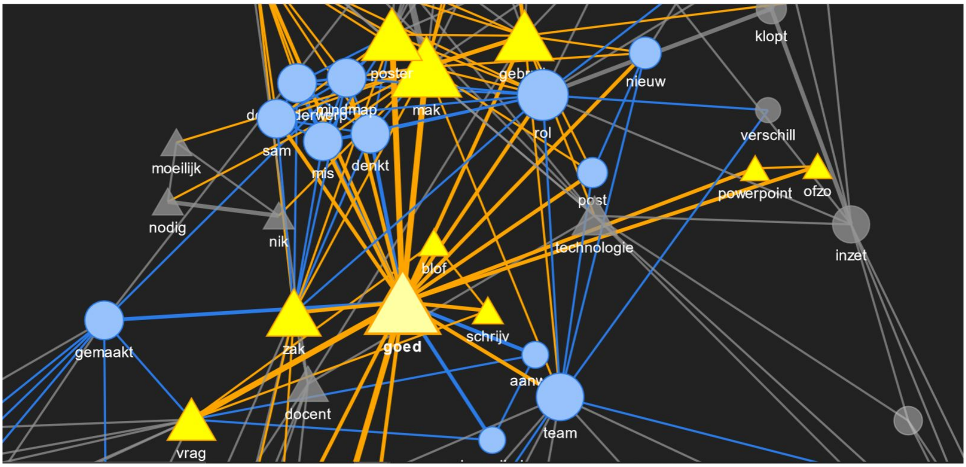


Figure 3.13 A part of the network graph with the highest betweenness centrality node ("goed") highlighted along with its neighbouring nodes.

3.5 Discussion

First, with the help of *RQ 1: "What co-located collaboration indicators have been detected from group speech data in past studies?"*, we found that most studies are on non-verbal audio indicators (e.g., total speaking time, pitch). Most previous studies on co-located collaboration (CC) focused on spectral and temporal audio indicators (Bassiou et al., 2016; Lubold and Pon-Barry, 2014). These indicators are less obvious to understand or denote what is happening during collaboration as compared to verbal audio indicators (or content of the conversation). Sometimes, depending on the cultural background of a group member, the tone of the voice can vary. Therefore, the tone of voice can be a good or bad indicator of collaboration depending on the background. On the other hand, taking the example of "what" of the conversations, it is more overt in most of the circumstances irrespective of the background. The reason for most studies being heavily inclined to the non-verbal audio indicators of collaboration could be due to the lesser maturity of automatic speech recognition (ASR) systems, making it difficult to transcribe the conversations. Of late, with the advent of different ASR systems like Google speech-to-text, it has become much easier to convert speech to text with high accuracy. Upon expanding the literature, we found very few studies (e.g., Chandrasegaran et al. (2019); Huber et al. (2019)) on verbal audio indicators of collaboration in a CC context. These studies were lab based and mostly focused on getting a representative high-level overview of the conversations as topical word clusters instead of examining the richness of the linkages among the words in the conversation.

To address this, we answered *RQ 2: "What collaboration analytics can be employed to analyse group speech data from co-located collaboration automatically?"*. For this, we built a technical set up and conducted a field trial where we recorded sessions of a board game where people had to collaboratively design learning activities and each player was assigned a specific role to play beforehand. With the help of this collaborative task, first, we had an exploratory understanding of the collected audio data set. Then, we visualized the relationship between these words, apart from the representative topic modelling (which are certain representative word clusters) done in the past (Chandrasegaran et al., 2019). First, we understood the different bigram (or two consecutive) word phrases and made a distinction between some of the most occurring bigrams and least occurring bigrams. We did this in one phase of the CC task, which was more inclined towards the technology for two roles, i.e., the TEL LA advisor and teacher, because they had the most exchange of turns during that phase. We thought this could help us uncover the main bigrams to understand the contribution of dominant turn-taking roles in that phase. Even though bigrams showed a representative collection of two consecutive word phrases, it was still difficult to understand how the conversations happened and what were the influential words (measured by eigenvector centrality), as well as what were the most controlling words (measured by betweenness centrality), which could be shown using different centrality measures, as also done in the past (Das et al., 2018). To understand this further, in addition to visualizing the strength of the bigrams and longer word phrases, we plotted the social network graph (as done earlier in online settings (Xie et al., 2018)), which was interactive and made it easy to select a particular node and highlight its neighbours. This network graph made it easier to understand the contribution of individual roles in the group conversation. The strength of the link between words (measured by how often they co-occurred in a sentence) along with their use by different roles helped to capture both word-level and role-level interaction, which can be a simplification approach to understand a huge text corpus. The highlighting, interactive feature of the tool was also purposefully built to reduce the information overload and only focus on that particular selected node (or word) and the neighbours automatically highlighted. Besides the visualization and its design choices, which were fully automated, during pre-processing, some degree of human involvement was necessary for sanity checks, which is explained in detail in the next section (under Challenges and Limitation).

Now, the next obvious question is: *"What is the way forward once we have this technical set up ready and a tool ready to be used in different settings?"* The first step for us is to use this across the other sessions for which we collected audio data and see if we can find some recurring themes. We also want to add further enhancements (as additional modules) and refine the visualizations by involving the two main stakeholders (i.e., group members and the person managing the CC task) for whom this was made. We want to understand the turn-taking patterns between roles further and how the conversation evolves around these turn-taking patterns. For now, it is a post hoc group collaboration analytics tool. To take it a step further, it would be interesting to add a module to detect the quality of collaboration. One possible step can be to compute the cosine similarity distance by comparing the text vectors of each role as computed by this technical set up to the expected contribution of the roles. This can give an estimate of how aligned the conversations are to the expected contribution for that role and an estimate of the collaboration quality if we can quantify the cosine similarity distance as a quality measure. Therefore, to proceed in this direction, we need to bootstrap the results of all these sessions to build models of different roles along with a human to form some expected standards (or words they would use) for each role during the CC task.

Therefore, for this article, our main contribution was twofold: identifying the gaps of the current co-located collaboration (CC) quality detection approaches from audio data and making a starting step towards an automatic holistic collaboration quality detection technical set up with the prototyping of our set up in the context of a CC task. This was a technical article stressing the different technological approaches (the coding details of which can be found on GitHub (`https://bit.ly/autocollabanalytics`, last accessed on 30th April 2021)) to move towards automatic collaboration analytics right from audio data collection to generating meaningful visualizations.

3.6 Challenges and Limitations

There are many challenges. First, architectural challenges are full automation, the accuracy of speaker diarization and the accuracy of speech to text. During speaker diarization, sometimes, labels of roles were misplaced, which were manually corrected. Next, there are challenges in processing and analysing the data, which are largely dependent on the accuracy of the speech to text, which we will explain below. The unstructured text data obtained from audio are much different than the data obtained from any online forums. Therefore, unstructured text data generates much noise, which to some extent can be structured by sentence segmentation. However, sentence segmentation working on only spoken text without punctuation marks or delimiters can cause sentence boundary detection problems. Another challenge in text processing is to correct the names, which were most of the time wrongly transcribed. For example, “moodle” was wrongly transcribed to “moeder”, and we had to manually fix this in the corpus. Therefore, when studies are in-the-wild without a controlled lab environment, then there are more chances for natural, unstructured conversations, which will need cleaning and structuring before analysis can yield meaningful results. The stop word corpus available to the algorithm did not remove all the contextual stop words that were not relevant for this discussion. We also needed to manually remove some contextual stop words like some action verbs depending on their importance in our context. When we lemmatized and stemmed the words, then the lemmatizer for Dutch text was not accurate enough because of its lesser usage and popularity compared to English. Therefore, we needed to manually correct some words, which could be seen in blue phase when “groep” and “groepjes” were not reduced to the same lemma as “groep”. The annotation process was time consuming.

The limitations in terms of automation can be summarized from the challenges. We needed the help of a human to pre-process to some extent for cleaning the corpus, the sanity check on the names transcribed and to make sense of the visualizations with the help of annotations. Although we are advancing towards automatic collaboration analytics, we need to eliminate other bottlenecks, especially to reduce the dependence on humans to as little as possible.

3.7 Conclusions and Future Work

First, we listed the indicators of collaboration obtained from the audio modality in the literature. We found two broad categories: non-verbal audio indicators (such as temporal, spectral audio features, total speaking time, overlap of speech) and verbal audio indicators (such as the content of the conversation, i.e., the spoken words). There have been many studies on the first category, but very few studies on the second category of analysing the content of the conversation. We found that with the maturity of automatic speech recognition systems, recently, analysis of the content of the conversations has picked up pace. Most studies analysing the content of the conversation looked at the high level topics and were lab based.

Therefore, we took a step further to build a technical set up and conducted a field trial to analyse the richness of natural unstructured conversations and the strength of the links between these words and phrases used in the conversation context, and thus, we prototyped the tool to move towards automatic collaboration analytics. Here, we analysed the conversations during a board game while designing a learning activity where group members with different roles (such as student, teacher, TEL LA advisor, study coach) interacted with each other. We found different interaction patterns between the teacher and the TEL LA advisor by analysing the word-level and phrase-level interaction during the technology related discussion phase of one collaboration session. Even though we were moving towards automated collaboration analytics, we found limitations in terms of automation with speaker diarization and data pre-processing.

As mentioned in the Discussion, we want to enhance the technical set up further by understanding the conversations around these turn-taking patterns between different roles. The major outlook for the future will be to measure the quality of collaboration and give feedback. Due to COVID-19, we are also looking into adapting our approach to a remote (or online) setting from a face-to-face setting. Because of our modular approach, it will be easier to adapt everything in the technical set up except the speaker diarization. We will not need speaker diarization in an online setting, and it will be much easier to get different clean audio streams from each group member in an online setting.

Part III

Towards Measuring Quality of Collaboration

Chapter 4

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

Colocated collaboration (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 Chapter, 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 move towards quantifying collaboration quality using the set up in Chapter 3. First, we define collaboration quality using audio-based indicators in a short literature review. Then we conduct field trials by collecting audio recordings in 14 different sessions in an university setting while the university staff collaborates over playing a board game to design a learning activity. This collaboration task consists of different phases with each collaborating member assigned a pre-fixed role. We analyze the collected group speech data using the set up built in the previous chapter to do role-based profiling and visualize the collaboration analytics with the help of a dashboard.

This chapter is based on:

Praharaj, S., Scheffel, M., Schmitz, M., Specht, M., and Drachsler, H. (2022). Towards Collaborative Convergence: Quantifying Collaboration Quality with Automated Co-located Collaboration Analytics. In *Learning Analytics and Knowledge Conference*, ACM, doi: 10.1145/3506860.3506922.

4.1 Introduction

Collaboration is one of the four important 21st-century skills (Kivunja, 2015). Collaboration is said to occur when two or more persons work towards a common goal (Dillenbourg, 1999). The recent interest on co-located (or face-to-face) collaboration (CC) is because of the ubiquity of the use of sensors and rise of multimodal learning analytics (Praharaj et al., 2021a; Di Mitri et al., 2018b). "CC takes place in physical spaces where the group members share each other's *social* and *epistemic* space." (Praharaj, 2019, 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 (Kim et al., 2015), total speaking time (Praharaj et al., 2018b) 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 (Praharaj et al., 2021b), log data about content of discussion if any) (Praharaj et al., 2018b). 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 (Praharaj et al., 2021a, 2018a). This can be attributed to the differing goals of collaboration and the group characteristics (which both can be collectively grouped under the parameters of collaboration) (Praharaj et al., 2021a). Majority of studies on CC in the past focused on the audio indicator type (Praharaj et al., 2021a).

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 (Bassiou et al., 2016), total speaking time of group members (Bergstrom and Karahalios, 2007; Bachour et al., 2010), interruptions (Oviatt et al., 2015), overlap or no overlap duration of speech (Bassiou et al., 2016); speaking time of a group member combined with the attention of other group members measured by their eye gaze (Terken and Sturm, 2010); linguistic features such as pronouns, sentence length and prepositions (Schneider and Pea, 2014a, 2015); verbal features like the keywords used, topics covered (Chandrasegaran et al., 2019), dialogues (Huber et al., 2019). Contrary to a majority of studies focusing on non-verbal audio indicators, a handful of studies used the 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 (Praharaj et al., 2021b). For example, "talk traces" (Chandrasegaran et al., 2019), "meeter" (Huber et al., 2019) used verbal audio indicators of CC for the analysis. In "talk traces", Chandrasegaran et al. (2019) did topic modeling during the meeting and then showed the topic cluster visualizations as feedback by comparing with 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 (Huber et al., 2019) made dialogue classification of the 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.

Similar to the partially automated analysis of epistemic space during CC, there has been a battery of works (Jeong and Chi, 2007; Teasley et al., 2008) 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) (Jeong and Chi, 2007). Main goal of knowledge convergence is learning together (Teasley et al., 2008). 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 (Teasley et al., 2008).

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 4.2 to get a background overview. Then, to answer **RQ2** and **RQ3**, we design an experimental set up where we collect audio data of university staff (pre-assigned with different roles) collaborating while designing learning activity using a board game in the university in 14 different sessions. 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 4.3). 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 (Jeong and Chi, 2007; Teasley et al., 2008). 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 4.4) and then discuss the future research that can be done on this dashboard (in Section 4.5). Finally, we discuss our findings (in Section 4.6) with the limitations and conclude in Section 4.7.

4.2 Co-located Collaboration Indicators from Audio

Majority of past works used audio indicator type to determine the quality of collaboration (Praharaj et al., 2021a). The simplest audio indicator of collaboration was *total speaking time* (Bergstrom and Karahalios, 2007; Bachour et al., 2010). The total speaking time of each group member was reflected back to them by a coloured LED light display (Bachour et al., 2010) and concentric circles visualization (Bergstrom and Karahalios, 2007) 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 had better quality of collaboration as measured by a post-test. Besides, more frequent speaker changes (i.e., *turn taking*) with overlap of speech (Kim et al., 2015) indicates a good quality of collaboration. Previous research also indicates that overlap in speech is associated with positive group performance (Çetin and Shriberg, 2006; Dong et al., 2009).

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) (Bassiou et al., 2016). 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 (Lubold and Pon-Barry, 2014). When two members in a group are speaking at different amplitude 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 (Lubold and Pon-Barry, 2014). Lubold and Pon-Barry (2014) 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 (Chapman et al., 2005). 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)” (Praharaj et al., 2021b, p. 4).

Additionally, Oviatt et al. (2015) 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 (2013) 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. (2017b) 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" (Chandrasegaran et al., 2019), “meeter” (Huber et al., 2019)) took into account the content of collaboration. In “talk traces”, Chandrasegaran et al. (2019) 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” (Huber et al., 2019) 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

Table 4.1 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	Chandrasegaran et al. (2019), Praharaj et al. (2021b)
Dominance	Total speaking time	Higher equity of total speaking time means less dominance in the group and higher quality of collaboration	Social	Kim et al. (2008), Bachour et al. (2010), Bergstrom and Karahalios (2007), Praharaj et al. (2019)
Active participation	Turn taking frequency	More frequent turn taking changes mean higher active participation and better quality of collaboration	Social	Kim et al. (2015)
Expertise	Overlapped speech	Overlap in speech is an indicator of constructive problem solving, expertise and good CC quality	Social	Zhou et al. (2014), Oviatt et al. (2015)
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	Lubold and Pon-Barry (2014)
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	Jeong and Chi (2007), Teasley et al. (2008)

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 (Jeong and Chi, 2007) and cognitive convergence (Teasley et al., 2008) (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 (Teasley et al., 2008). This effectively improves collaboration.

Table 4.1 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.

4.3 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.

4.3.1 Experimental Context

The collaboration task was to design a learning activity using the Fellowship of Learning Activity and Analytics ((FOLA)^2) [1] game. We used this task to collect the audio recordings. It is a board game (Schmitz et al., 2019) (e.g., of an online version of the game [2] currently under active development) played face-to-face with different themed cards and roles that is 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. The collaboration task in each session had different phases which were colour-coded based on the cards supposed to be used in 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 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, etc. The *yellow* phase (varied in length from 10 to 30 minutes across the sessions) defines what we want to know on the interaction or within the learning activity. For instance, it can be engagement or how students take initiative. To steer the group conversations, there were also prompts on each card. A demo from one of the game

[1] http://www.fola2.com/
[2] https://game.fola2.com/
[3] The phases and cards mean the same and have been used interchangeably henceforth

Figure 4.1 A game session demo

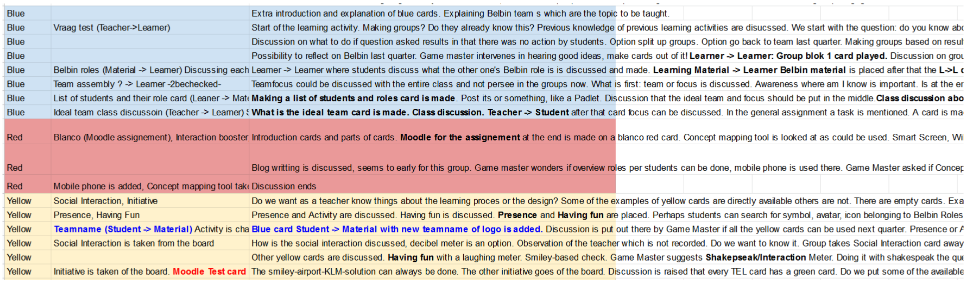
session is shown in figure 4.1.

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 real life advisors (age varied from 36–64 years, experience varied from 1–20 years and gender consisted of 10 males and 8 females), teachers (age varied from 28–64 years, experience varied from 1–40 years and gender consisted of 13 males and 1 female), learners (age varied from 19–27 years, experience varied from 0.16–9 years and gender consisted of 13 males and 1 female) and study coaches (age varied from 28–56 years, experience varied from 0–14 years and gender consisted of 12 males and 2 females).

4.3.2 Methods

Our architecture for data collection, processing and analysis was based on our previous work (Praharaj et al., 2021b). We followed a similar data collection, pre-processing and processing approach (as in Praharaj et al. (2021b)) 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 a much cleaner user interface to play with the transcribed data and make minor modifications when needed.

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



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 ab
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**
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 -> Mat	**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)	**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 tak	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 4.2 A sample annotation

After data processing (where we clean the dataset and make it machine understandable), we find that 4th and 13th session needed to be removed because of poor quality of recordings and transcriptions. Then we move to the data analysis and visualizations with the 12 sessions. First, we visualize 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 take the help of 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 4.2).

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 4.3 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 these 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 use this co-occurrence matrix to build an interactive network graph to visualize (as shown in figure 4.12) 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 have 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 have combined both these social and epistemic components into a single dashboard to get a better understanding of the collaboration sessions. Further, we also explore 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 do 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.

4.4 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 4.3) which shows the speaking time and turn taking of the group members and then we explore the epistemic space (as in Figures 4.7 and 4.12). Here, we describe our findings in three different sub-sections.

4.4.1 Convergence within a phase in a session

We have defined collaborative convergence in the Introduction in 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 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 (blue phase in this case) in a session (1st session in this case). The blue phase was a conversation about the steps in the learning activity. If we compare Figure 4.7 and 4.8, when we observe "Belbin" 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 4.7) only advisor, learner and study coach uttered Belbin. But, in the first 20 minutes of blue phase (i.e., in Figure 4.8), 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 4.7 and 4.8 with Figure 4.9 or 4.10). 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 vs splitting in groups. Similar to epistemic convergence, social convergence can be observed based on the participation of different roles across every 10 more minutes slices of conversation

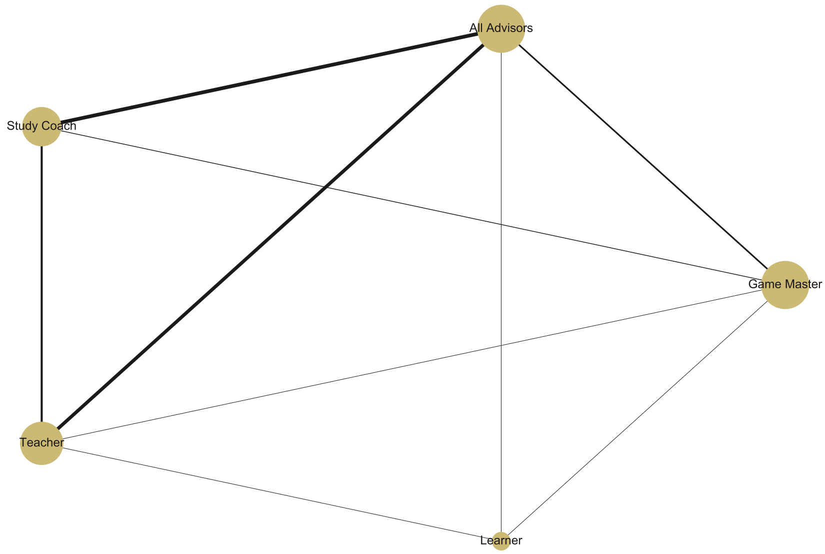


Figure 4.3 1st 10 minutes social space (1st session blue phase)

in Figures 4.3, 4.4, 4.5 and 4.6 of the blue phase. The learner in 1st 10 minutes did not interact with the study coach and spoke the lowest in terms of speaking time (shown as the smallest node). Towards the end it developed a link with the study coach because of turn taking exchange between them. So, both social and epistemic space gives us a holistic understanding of the evolving conversations' patterns within the phase of a session.

Regarding individual convergence, 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 improved a lot from first 10 minutes to the end of the blue phase (as can be seen in Figures 4.7, 4.8, 4.9 and 4.10) 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. 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.

4.4.2 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 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

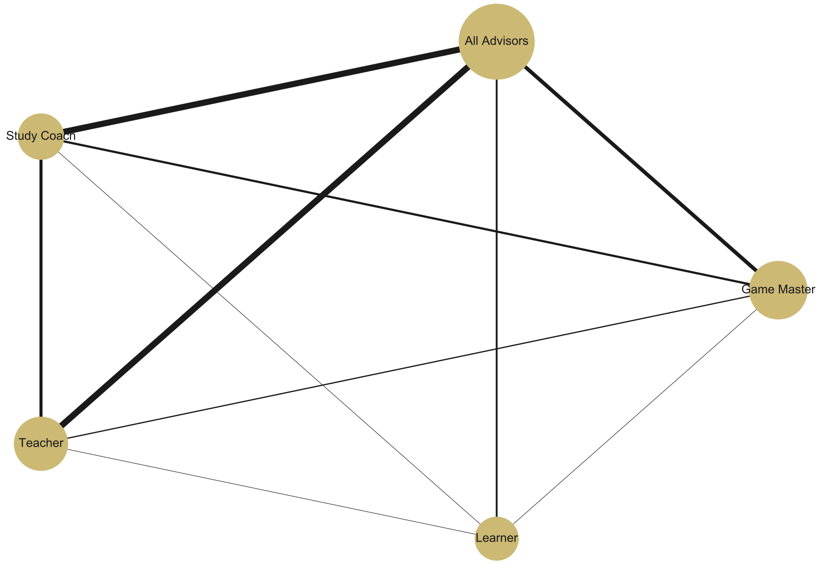


Figure 4.4 1st 20 minutes social space (1st session blue phase)

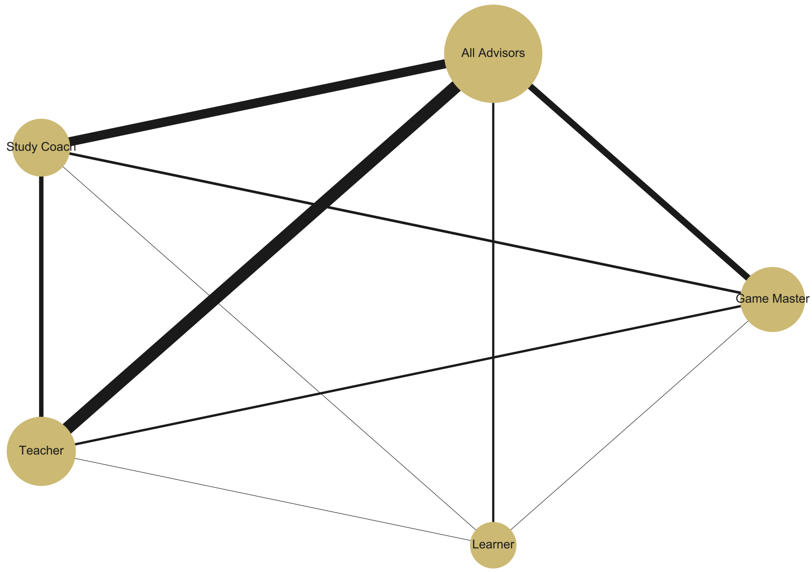


Figure 4.5 1st 30 minutes social space (1st session blue phase)

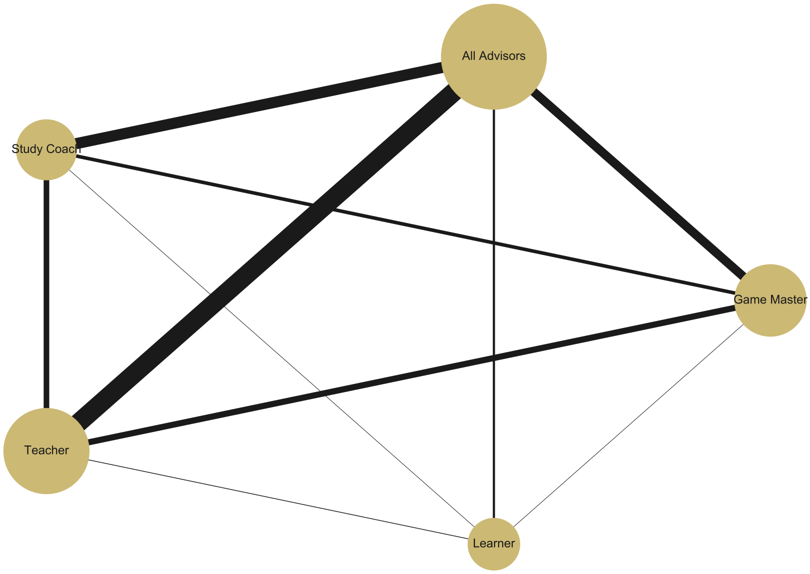


Figure 4.6 Full social space (1st session blue phase)

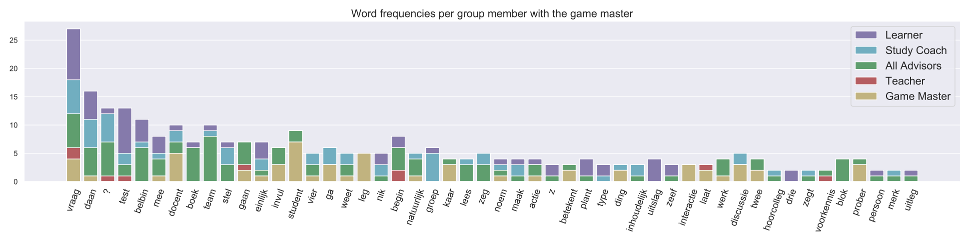


Figure 4.7 Top 50 word utterance frequency in the 1st session blue phase in 1st 10 minutes with roles

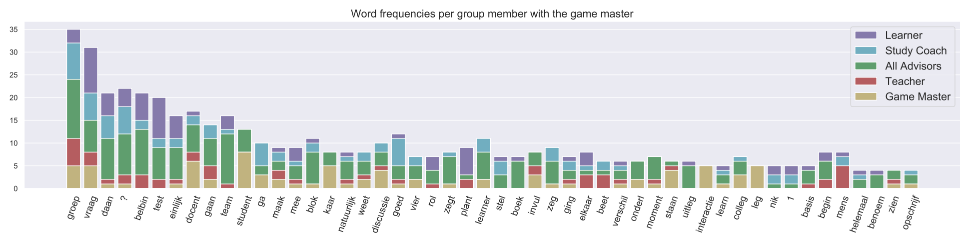


Figure 4.8 Top 50 word utterance frequency in the 1st session blue phase in 1st 20 minutes with roles

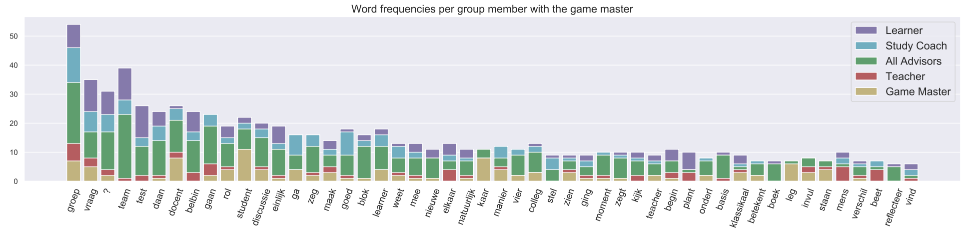


Figure 4.9 Top 50 word utterance frequency in the 1st session blue phase in 1st 30 minutes with roles

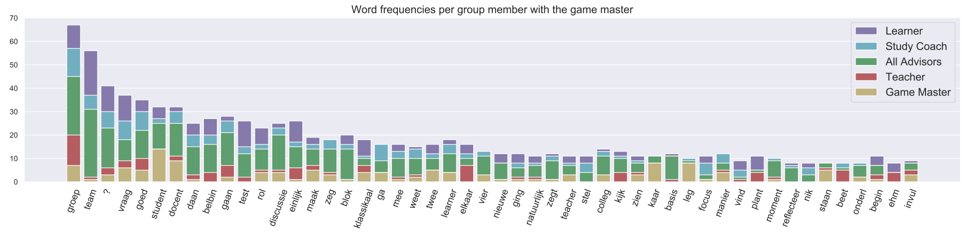


Figure 4.10 Top 50 word utterance frequency in the full 1st session blue phase with roles

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 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.

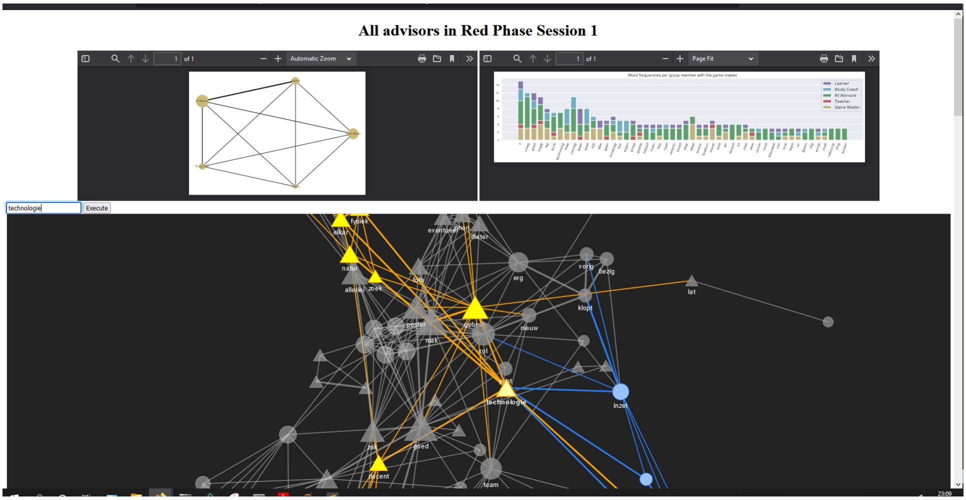


Figure 4.11 Screenshot of the dashboard with social and epistemic components

4.4.3 Dashboard encompassing the social and epistemic components

Figure 4.11 shows the dashboard highlighting a node for all advisors in the red phase in session 1. It has four main components. 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 4.12 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 analyse 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 number of times a node lies on the shortest path between other

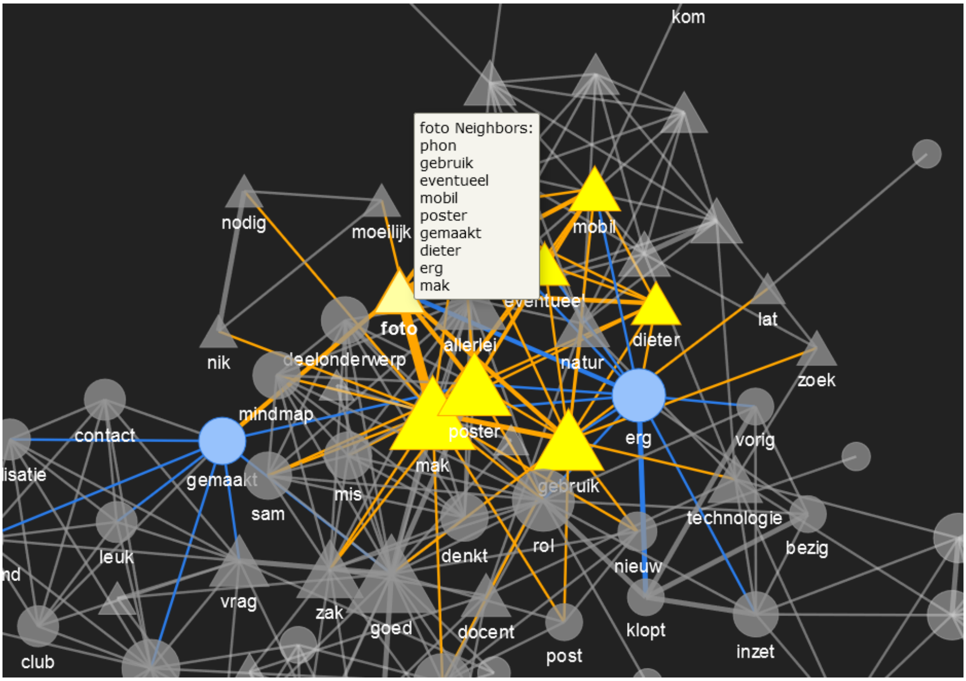


Figure 4.12 Zoomed in network graph highlighting a node of the advisor in yellow rectangles and rest others in blue circles in red phase

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 proves 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.

4.5 Future research about the dashboard

So, 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. 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) who will be using it. 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.

4.6 Discussion

To answer "RQ1: What co-located collaboration indicators have been identified from group speech data in the related literature?", we do a short literature review where we identify different indicators of CC quality from group speech data and define how they have been operationalized contextually to measure the quality of CC. We find that most studies (Bachour et al., 2010; Kim et al., 2008) in the past focused on the analysis of the social space of collaboration and few studies (Chandrasegaran et al., 2019; Huber et al., 2019) that focused on the epistemic space were abstract in nature. 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 conduct field trials in 14 different sessions (only 12 of which are later used for data analysis) where we collect the audio recordings. The collaboration task was to design a learning activity with each group member 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 take the help of the already defined indicators of collaboration quality in RQ1. We analyzed the collaboration convergence (ie., 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 (Jeong and Chi, 2007; Teasley et al., 2008). We find that specific keywords utterance frequency analysis 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 in 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 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 build a dashboard encompassing both the social and epistemic components. We use 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 build 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-occurrence (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 (Xie et al., 2018) and better than the abstract representation of topics (Chandrasegaran et al., 2019). We also computed the centrality measures (Das et al., 2018) 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 in the “future research about the dashboard”. 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 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 view of the collaboration patterns, it is not suited for a specific stakeholder group which can be modified 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.

4.7 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. We find this is a starting step in the direction of automated collaboration analytics and feedback 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.

General Discussion

The main objectives of the thesis are: 1) To describe the components used to define and measure the quality of co-located collaboration (CC) including indicators (i.e., low level events like the total speaking time), indexes (i.e., the processes like the equality of speaking time), task goals and parameters (i.e., the group composition and behaviour during CC) that define the CC quality; 2) To develop a technical prototype and test it with field trials to move towards automated collaboration analytics; 3) To visualize the CC analytics and move towards quantifying CC quality.

The thesis approached these objectives in three distinct parts. **Part I** described the definition for CC quality, and how quality is contextualised in certain scenarios of collaboration. It also connects it to suitable feedback mechanisms to support learners in collaboration. **Part II** described the prototyping of an automatic CC analytics set up (using the definitions in Part I) where we built an architecture for data collection, processing, analysis, visualizations and then tested it in a field study based on a specific CC task in a university setting. Here, we primarily focused on the audio-based indicators of CC. In **Part III**, we used the set up built in PART II to move towards quantifying the quality of collaboration based on the content of the conversations and how group members speak and then visualized this analytics using a dashboard.

This concluding chapter first summarises and discusses the main outcomes from these three parts as a mix of observations and recommendations; then addresses the limitations of the presented research as well as implications for the field and future research.

Observations and Recommendations

Based on our findings in three parts of the thesis, we derive these observations and recommendations for the CC analytics and the CSCL community as three primary clusters. The clusters are based on different phases of research: in the early phase, the goal is to *1) Define and understand the theory of CC quality, analytics*; in the mid phase, the goal is to *2) Prototype the automatic CC analytics set up*; and in the later phase, the goal is to *3) Visualize the CC analytics to move towards quantifying the quality of CC*.

1. Define and understand CC quality, analytics

    a) Define CC quality

        - CC quality can be defined using an event-process framework of indicators-indexes

- Scenario, context and group characteristics (i.e., parameters) need to be considered when defining CC quality
- Mapping of the parameters on indicators and indexes can define the measurable conceptual framework for CC quality detection

b) Operationalization of CC quality

- Operationalization of CC quality suffers from coding complexity and use of opaque inexplainable algorithms
- The gap between the theoretical measurable markers (i.e., the indexes) of CC quality and the practically detected measurable markers needs to be bridged

c) Understand CC analytics and feedback design

- All indicators of CC quality can be grouped into two categories (i.e., social and epistemic)
- A group of epistemic indicators cannot be detected and analyzed by sensors alone
- CC analytics and feedback design is dependent on the stakeholders
- Humans in the prototyping and feedback loop help to speed up the design

2. Prototype an automatic CC analytics set up

a) Content of the conversations can provide rich information about CC quality

b) Design and prototype of a CC analytics set up is difficult to fully automate

c) Interactive network graphs can provide rich insights into the interconnected conversation patterns between group members

3. Visualize the CC analytics to move towards quantifying the quality of CC

a) There needs to be a focus shift from the social space to the epistemic space of CC analytics

b) Collaborative convergence (i.e., an increase in shared knowledge among group members) measured from the epistemic space helps to quantify the quality of CC

c) Both social and epistemic space can complement each other when visualized on a dashboard (using interactive network graphs and bar graphs) to understand CC analytics holistically

1. Define and understand CC quality, analytics

In Part I (i.e., the early phase) the CC quality needs to be defined along with its operationalization and usage to understand CC analytics and feedback design.

a) Define CC quality

CC quality can be defined using an event-process framework of indicators-indexes

We did an in-depth literature review of the indicators of CC quality. Here, we first focused on the indicators that are relevant to understand the quality of CC from the related literature. Then, we defined the indexes of CC using these indicators. The low-level sensor-based, human or hybrid *events* in collaboration after processing and aggregation form the indicators of collaboration which help to detect the quality of CC. The high-level indexes consist of one or more indicators obtained from multiple indicator types. They act as a proxy to detect the quality of CC. For example, counting the number of ideas during a brainstorming scenario in CC is obtained from the events grouped in the content indicator type; while a high-level *process* definition, that is, equality of the number of ideas generated by each member in the group, measures the quality of collaboration. Here, higher equality in a group denotes a better quality of collaboration. Thus, the *event–process* conceptual framework provides a holistic overview of the quality of collaboration based on the practical studies.

This conceptualization is a precursor for building different types of collaboration detection, monitoring, and prediction systems. Some of the indicators like total speaking time (Bachour et al., 2010; Kim et al., 2008), and number and duration of overlap of audio (Zhou et al., 2014) are consistently indicative of collaboration quality across different studies. But, the same is not true for other indicators such as distance between group members. The distance between group members gives a mixed indication of the quality of CC; that is, sometimes it is inversely proportional (Spikol et al., 2017a,b) or sometimes there is no relation (Schneider and Blikstein, 2015) with CC quality. Thus, the comprehensive overview of the indicators will help practitioners to choose the relevant types of sensors, sensing mechanisms and indicators to detect CC quality according to their setup. So, first they can focus on the indicators (such as total speaking time and joint visual attention) that worked in most settings instead of focusing on the indicators (such as writing speed, pressure from the digital pen, distance between group members, and space usage in the room during group work) that have no relation with CC quality during preliminary studies.

Scenario, context and group characteristics (i.e., parameters) need to be considered when defining CC quality

The indicators of CC quality vary across scenarios, their goals and contexts, and the fundamental characteristics of the group. Therefore, we also looked at the impact of different *scenario-based goals* and *parameters* for CC on the relevance of the different indicators in the literature review. We found that the scenario of CC (i.e., task context) chosen has a huge impact on the indicators of collaboration obtained. The parameters of CC (i.e., group composition and group behaviour such as dominance, rapport) also affects the detection of quality. Some scenarios have a stark contrast in terms of the collaboration indicators observed; for instance, collaborative brainstorming and collaborative gaming. However, some scenarios have certain overlapping collaboration indicators; for instance, collaborative design and collaborative concept mapping. This detection of scenario-based indicator types is also dependent on the use of the shared artifacts (e.g., patient manikins or smart touch table). The scenarios heavily dependent on shared artifacts tend to be inclined towards nonverbal indicator types (such as engineering design, gaming and healthcare simulation). In addition to it, some indicator types like eye gaze, gesture, and audio are dependent on context while some others like physiological ones are not. We find that higher occurrence of joint visual attention (JVA) (Schneider et al., 2015) measured from the eye gaze indicates better CC quality while the same is not true when individual eye gaze of speaker and listener is considered (Terken and Sturm, 2010). This indicates that CC is *scenario-dependent* and the collaboration indicators can vary depending on the scenario, its goal, and context. But, when we consider physiological indicator types then we find that instances of aroused and relaxed states are context-independent and can be misleading unless contextualized with other modalities like audio (Malmberg et al., 2019). Besides, the varying scenario-based goals, groups also vary in fundamental parameters. To understand this, we created a mapping of the parameters to the indicators and indexes to model a conceptual framework with the aim to detect CC quality.

Mapping of the parameters on indicators and indexes can define the measurable conceptual framework for CC quality detection

For this mapping, we chose one of the most occurring CC scenarios (i.e., collaborative problem solving) in the review which had well-defined task objectives too. Here, we mapped the CC parameters (such as behavior, composition, interaction, etc., of group members) onto the indicator types and the indexes. We found that *mapping* the parameters helped in furthering the *semantic enrichment* of the parameters, highlighting the relevance of the indicators, and thereby defines a complete measurable setup. For instance, dominance (a parameter of CC) can be mapped onto audio (an indicator type) taking into account the total speaking time (as an indicator) to measure the equality (an index) in the group; whereas the same parameter can be mapped onto synchrony (as an index) when posture (indicator type) is considered. Therefore, the same fundamental parameter (i.e., dominance in this case) can be measured differently depending on the indicator type (consisting of a group of similar indicators) and the indexes considered for measuring the quality of collaboration. If a group has higher dominance then specific members are more dominant than others. This is measured by synchrony or equality. Higher dominance means lesser synchrony or equality and worse quality of collaboration. Therefore, this measurable conceptual framework acts as a road map for future research and evaluation on CC quality.

b) Operationalization of CC quality

Operationalization of CC quality suffers from coding complexity and use of opaque inexplainable algorithms

The operationalization of the indexes from the indicators suffers from multiple limitations. Coding the indicators to compute the indexes is challenging at times (Cukurova et al., 2018, 2017a), as in the case of individual accountability; thus failing to detect CC quality. Another limitation is the use of machine learning approaches (Grover et al., 2016; Luz, 2013; Stewart et al., 2018; Viswanathan and VanLehn, 2017) which use one or more indicators to detect CC quality but fail to address the qualitative aspect of these indicators. For instance, silence and pause are good indicators of CC quality combined with other indicators (Luz, 2013) but it is unclear if more or less occurrence of silence in itself indicates anything about the quality of CC. This tension between the transparency of the learning analytics models and the accuracy highlighted by Cukurova et al. (2020) is still an open question. Some machine learning models which are like a black box (very opaque) have higher accuracy even though they are not transparent in terms of the role of each of the indicators of CC.

Furthermore, few indexes have been operationalized from a chosen set of indicator type while ignoring other types. For instance, synchrony has not been operationalized using the content indicator type. This may be attributed to the difficulty involved in detecting and analyzing the similarity of content of a discussion (or comparing the semantic nature of the discussion itself) during collaboration. This also brings into picture the importance of choosing the right sensing mechanisms and sensors in the respective CC scenario. However, equality index has been easily detected using the content indicator type (number of ideas as an indicator) as it is easier to measure a quantitative value (i.e., the number of ideas generated by each member during collaboration).

The gap between the theoretical measurable markers (i.e., the indexes) of CC quality and the practically detected measurable markers needs to be bridged

The mapping while defining the measurable conceptual framework for CC quality detection earlier is incomplete because of a lack of the operationalization of the indexes and a dearth of well-defined task goals. This restricted our conceptual framework design to only one of the most occurring scenarios. To overcome this scarcity, we substituted expected indexes based on our understanding from both the theory and practice. Thus, there is an urgent need for practitioners (or teachers) to act upon the other theoretical indexes (as outlined by Meier et al. (2007)) when measuring collaboration quality in CC. This can make more indexes from the theory visible in practice and bridge the gap to define a measurable setup for each scenario. Nonetheless, the framework is a starting point for making design-based decisions of a particular scenario of CC so that more measurable markers (i.e., indexes) can be added up to make it a complete set up for CC quality detection. Other collaboration frameworks (OECD, 2017; Ofstedal and Dahlberg, 2009; Koh et al., 2016) can be looked into to help in bridging this gap by comparing these frameworks.

c) Understand CC analytics and feedback design

All indicators of CC quality can be grouped into two categories (i.e., social and epistemic)

When reviewing the past literature on indicators of CC quality, we found broadly two groups (Praharaj et al., 2018b). Firstly, the *social* group consists of all the non-verbal indicators (such as different gestures and postures, eye gaze, non-verbal audio indicators like the total speaking time and turn taking) and secondly, the *epistemic* group consists of the indicators relevant to understand the content of the conversation (such as the actual content of discussion, ideas presented, any content related data obtained from data logs). The idea of grouping the indicators comes from the social and epistemic collaboration scripts grouping (Weinberger et al., 2005).

A group of epistemic indicators cannot be detected and analyzed by sensors alone

For detecting the social group of indicators, sensors have been mostly used (Kim et al., 2008; Schneider et al., 2015). But, for detecting the epistemic group of indicators human help was required (Praharaj et al., 2018b) as it is difficult for sensors to automatically detect when ideas are generated from speech by understanding the semantics. It is also difficult to understand the actual content and context of the discussion. This was evident from different studies needing human observers to overcome this semantic difficulty (Tausch et al., 2014; Wise et al., 2017; Harrer, 2013). Lately, with the maturity of automatic speech recognition (ASR) techniques and natural language processing (NLP), the automation of speech to text transcription and analyzing the content of the conversations is showing promise. But, it is far from completely removing humans in the loop yet. For example, humans are needed to change certain names in speech to text transcription, to understand the boundary of a long conversation and separate ideas contributed in a particular context from non-ideas (Tausch et al., 2014).

CC analytics and feedback design is dependent on the stakeholders

While addressing the review on the feedback mechanisms during CC, we found feedback is either real-time (for acting as reflection or guiding the group members) or post-hoc (for the purpose of reflection). We found two main stakeholders for whom this feedback was designed in different studies: The teachers (or facilitators) and the group members. This distinction was essential to understand the feedback design requirements. Some works used smart devices (such as tangible user interfaces, interactive smart white boards and tablets) during collaboration which require a lot of preparation to set up before a collaborative task. Therefore, it is difficult to use these in real-world dynamic settings. There were also two categories (i.e., private and public) of feedback display. So, there is a trade-off between personalization and privacy for the group. More personalized feedback for the whole group meant that it is also less privacy preserving. Thus, a design decision needs to be made on the level (i.e., group, individual or both) of feedback to be shown for different stakeholder groups.

Humans in the prototyping and feedback loop help to speed up the design

To give real-time feedback using a hybrid set up of humans and sensors, we took a quick preliminary step in building an initial prototype design. The main goal was to facilitate collaboration in real-time during group PhD meetings and to get a feel of how this set up works in authentic in-the-wild setting. We were successful in building a click-based interface for the human annotator who clicked the interface based on "who spoke when". This was made possible with microphones capturing the audio of the group. Then the total speaking time and turn taking patterns for the group were generated temporally with a click and shown as a real-time reflective feedback on a public display. This helped us to create a hybrid setup very quickly without building an actual automated sensor-based system to experiment with different types of real-time feedback mechanisms during CC and enabled us to gather some insights for future feedback designs. Therefore, humans in the prototyping and feedback loop help to speed up the design by bootstrapping. Using these insights, other modular components can be built to track other indicators of collaboration quality; and integrate them on a single dashboard.

2. Prototype an automatic CC analytics set up

Using the definitions of CC quality, analytics in part I, an automatic CC analytics set up can be prototyped.

a) Content of the conversations can provide rich information about CC quality

Surface level methods (like the participation measured by total speaking time, turn takings) provide rough analysis of collaboration in the group. Content analysis helps us to know why one group member contributes more and appears to be more influential in the group (Strijbos et al., 2006). This content analysis on a more deeper level has been analyzed manually using knowledge convergence (i.e., increase in shared knowledge in the group) and cognitive convergence studies by (Jeong and Chi, 2007; Teasley et al., 2008). High knowledge convergence implies better quality of collaboration. Similarly, other studies have done content analysis with human observers by measuring the number of ideas contributed by the group members (Tausch et al., 2014). They found that if every member makes an equal contribution to the number of ideas then that group has better quality of collaboration. Other studies were on content analysis with the help of external artifacts like Tangible User Interface and Smart Tabletops where a comparison of a stored solution with the solution of the group members was made to detect the collaboration quality based on how much the solution matched to the expected answer (Echeverria et al., 2017; Wong-Villacrés et al., 2016; Granda et al., 2015).

Recently with the maturity of the automatic speech recognition (ASR) techniques, doing content analysis from audio conversations has become easier because of the accuracy of the automatic speech to text transcription. Few studies (e.g., Chandrasegaran et al. (2019); Huber et al. (2019)) have done topical keyword cluster abstract overviews by utilizing ASR and NLP techniques. Then they compared these clusters to the meeting agenda to detect the quality of collaboration based on how much the discussions match with the agenda. Therefore, content of the conversations can provide rich information about CC quality. We also did content analysis of the conversations in an university setting which we will discuss in the following subsection after discussing the challenges.

b) Design and prototype of a CC analytics set up is difficult to fully automate

There are a lot of challenges in the technical implementation and prototyping of an automatic CC analytics set up. First, architectural challenges are full automation, accuracy of speaker diarization, and accuracy of speech to text transcription. During speaker diarization, sometimes labels of roles are misplaced which needs to be manually corrected. Next, there are challenges in processing and analysing the data which is primarily dependent on the transcription. The unstructured text data obtained from audio in a co-located setting is quite different and noisy compared to data obtained from any online forums. It needs sentence segmentation to split it into meaningful sentences. But, the sentence segmentation working on only spoken text without punctuation marks or delimiters can cause sentence boundary detection problems.

Another challenge in text processing is to correct the names which are most of the time wrongly transcribed. For example, “moodle” (the name of a learning management system) was wrongly transcribed to “moeder” (the Dutch word for mother) and we had to manually fix this in the text corpus. So, when studies are in-the-wild without a controlled lab environment then there are more chances of natural unstructured conversations requiring cleaning and structuring before analysis can yield meaningful results. Sometimes automatic stop word removal by the algorithms is not sufficient. We also needed to manually remove some contextual stop words like some action verbs depending on their usage in our context. When we lemmatize and stem the words then the lemmatizer for Dutch text is often not accurate enough because of less usage and popularity as compared to English lemmatizers resulting in same meaning for two different root words. Therefore, sometimes we need to manually correct some root words.

However, lately with the maturity of the ASR systems and the NLP text processing techniques, some of the challenges like punctuated speech to text transcriptions, contextual removal of stop words, lemmatization of text in other languages than English is being improved further. But, one needs to keep this in mind during the design and prototyping of a CC analytics setup.

c) Interactive network graphs can provide rich insights into the interconnected conversation patterns between group members

While doing content analysis of the text corpus obtained from the conversations, first, we analyzed different bigrams (or two consecutive word phrases) and made a distinction between some of the frequently occurring bigrams and rarely occurring bigrams. We did this in one session and one phase of the CC task which is more inclined to technology-based discussion for two roles i.e., "Technology Enhanced Learning and Learning Analytics Advisor" and "Teacher" because they had the highest turn-takings during that phase. The primary objective was to make sense of the dominant role-role interactions in a phase with the help of the bigrams. To understand the conversations further in-depth, we built the co-occurrence matrix which shows the strength of the co-occurence of the words and how often they occur together in the same utterance. We visualized this using an interactive (i.e., highlight specific portions to reduce distraction and information overload) social network graph (as done earlier in online settings by Xie et al. (2018)). Here the strength of the word-to-word linkage was shown by varying the thickness of the edge and the word frequency was shown by varying the size of the node. Then we highlighted the conversations of different roles in the whole conversation corpus using different shapes for the nodes.

Therefore, the interactive network graph made it easier to understand contribution of different roles and role-role exchanges during the group conversations. To understand the influential and controlling words and phrases during the conversation, we computed different centrality measures (as done in the past by Das et al. (2018)) from the network graph. The influential words (measured by eigenvector centrality) and the controlling words (measured by betweenness centrality) bring out the important discussion points which were not visible plainly with the word frequency. For instance, in the technology phase, the word "technology" was not frequently used during discussion but it was one of the words with high betweenness centrality. It is because the word "technology" has more influence over the network of words (or more often mentioned with other words in a phrase) during that phase of CC.

3. Visualize the CC analytics to move towards quantifying the quality of CC

Finally CC analytics needs to be visualized to be able to get a holistic understanding and make an attempt to understand the quality of CC.

a) There needs to be a focus shift from the social space to the epistemic space of CC analytics

Using the set up developed in Part II, we conducted field trials across 14 different collaboration sessions and built a dashboard to visualize the group speech data in Part III. Expanding the short literature review on indicators of CC quality from group speech data in Part II, we defined the contextual operationalization of the indicators from the perspectives of different spaces, i.e., physical, social and epistemic (based on Praharaj (2019); Praharaj et al. (2018b)). We found that most studies (Bachour et al., 2010; Kim et al., 2008) in the past focused on the analysis of the social space of collaboration and that the very few studies (Chandrasegaran et al., 2019; Huber et al., 2019) that did focus on the epistemic space were abstract in nature or followed a manual approach (Jeong and Chi, 2007; Teasley et al., 2008). Epistemic space analysis helps us to know why one group member contributes more and appears to be more influential in the group (Strijbos et al., 2006) instead of a rough surface level analysis on collaboration (e.g., who spoke more or which group members exchanged most turns). The indicators from the social space are less obvious to understand collaboration as compared to indicators from the epistemic space (or the actual content of the conversation). Depending on the cultural background of a group member, tone of the voice can vary; thereby, acting as a good or bad indicator of collaboration. On the other hand, the "what" of the conversations is more obvious in meaning in most of the circumstances irrespective of the background. One reason for the heavy inclination to the social space of CC could be attributed to the immaturity of automatic speech recognition (ASR) systems which gives rise to the difficulty of transcribing the conversations. However, recently, with the ubiquitous use of different ASR systems like Google speech-to-text, it has become much easier to convert speech to text with higher accuracy and speed. Therefore, there needs to be a focus shift from the social to the epistemic space of CC analytics.

b) Collaborative convergence (i.e., an increase in shared knowledge among group members) measured from the epistemic space helps to quantify the quality of CC

We investigated how co-located collaboration indicators from group speech data can be analyzed automatically. To this end, we conducted field trials in 14 different collaboration sessions where we collected the audio recordings in an university setting. The CC task was to design a learning activity. Here, each group member was pre-assigned a role (such as teacher, all advisors, study coach, learner and game master) before collaboration. Each session had three different phases (i.e., blue, red and yellow), each phase with different objectives. For the CC analytics we used the definition of quality of CC from the literature review (Praharaj et al., 2021a). We analyzed the collaboration convergence (i.e., increase in shared knowledge 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 Jeong and Chi (2007); Teasley et al. (2008). This helped us to understand the change in role-role conversation patterns with time. Furthermore, this helped us to infer if more phase-related keywords were discussed or not which in turn defines their contextual importance. Along-with with the social space analysis (shown as role-role interaction network graph with speaking time and turn-taking between roles), the holistic overview of how the conversations evolved could be obtained. This helped us to quantify the collaboration quality.

Our primary objective of doing this was to show how to measure collaboration quality using convergence instead of categorizing whether higher or lower convergence is good or bad. It turned out that this categorization fully depends on the collaboration context and how on-topic or off-topic the word or phrase being used is. As we did not define fixed objectives before the collaboration task and did not conduct a controlled lab-based study, the detection of CC quality is quite open to interpretation.

c) Both the social and the epistemic space can complement each other when visualized on a dashboard (using interactive network graphs and bar graphs) to understand CC analytics holistically

Finally, we investigated how to visualize the quality indicators of collaboration from group speech data. We built a dashboard including 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 get an in-depth understanding of the epistemic space, we built an extended interactive network graph (which can be highlighted with a search bar, selection or hovering) to do role-based profiling of the dominant conversations (measured from the social space using indicators like total speaking time and turn takings) during collaboration. Using this graph, we understood the frequently used words, strength between certain words, importance of the words and the importance of the uttered phrases. These visualizations on a dashboard helped us to analyze the evolving CC conversation patterns holistically, hence understand the CC analytics holistically. Therefore, from the social space one can understand "who" dominated the conversation and then from the epistemic space it will be clear as to "why" that member dominated and in "what" way.

Besides, the dashboards will need customization in the future depending on the purpose of use (i.e., post hoc or real-time) and the stakeholders (teachers or group members). The dashboard is now a generic version which is information rich instead of being tailor-made for a specific stakeholder group. Even though it provides a holistic view of the evolving collaboration patterns, it is not targeted to a specific stakeholder group (i.e., teacher or group members) which can be modified further depending on their needs to make it suitable for them. Furthermore, the network graph in the dashboard 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 in future iterations.

Limitations

Limitation of the literature review done in **Part I** is having to exclude different types of study (i.e., correlational vs interventionist) due to the scope of the review. This can open doors for another direction of future work to look into different feedback mechanisms used and how this model can be helpful with this regard.

The final limitation of the review is that we did not consider the number of groups used by different past studies. We think that this will be a good direction of future research even though it will be difficult to determine a threshold as to how many groups considered in a study will make it worthy of inclusion in the review. As per the title of Chapter 2, we do not think we are there yet (i.e., the whole nine yards) because CC modeling is dependent on many factors like how it is operationalized, in what context, and the impact of culture. Thus, we have a made a starting step to model CC in one of the scenarios taking into account the indicators, indexes, and parameters but not considering the number of groups, or types of algorithms (such as pattern mining, random forest classifiers) used.

The architecture and prototype related limitations in **Part II** have been addressed in the recommendations. Similarly, in **Part III**, the dashboard has some limitations which have also been addressed in the recommendations.

Moreover, in this current study in the thesis, we have used only one group setting. If we use different group settings then we can have different outcomes. We also did not test the impact of the visualizations on the stakeholders which will be an interesting direction to go later.

Implications and Final thoughts

The thesis has major implications in the field of co-located (or face-to-face) collaboration or team(work) research. It gives recommendations for the CSCL, collaboration and team(work) research community based on different phases (i.e., early, mid and later) of the research in 3 different clusters.

In Part I, CC quality is defined along with its operationalization and then CC analytics, feedback design is understood. We provide a theoretical definition of the CC quality and contextualized it in different scenarios. For instance, in collaborative brainstorming (i.e., the CC scenario), equality (i.e., index) of the number of ideas contributed (i.e., the indicator) by each member in the group helps to determine the quality of CC (Tausch et al., 2014). So, the higher the equality, the lower is the dominance (i.e., the parameters) in the group, and thus the better is the CC quality. Therefore, this example shows that in the case of a collaborative brainstorming context here, the presence of less dominant members is what makes the qualities of effective team(work).

Moreover, we model a conceptual framework for CC quality detection based on a scenario-driven prioritization. This can act as a starting step for researchers to operationalize other CC indexes from theory to practice. They can borrow from research on collaboration indexes in an online setting. For instance, previous works have detected different indexes during remote or online collaboration (from the eye gaze as an indicator) like reaching consensus, information pooling, and time management (Schneider and Pea, 2014b) (as outlined by Meier et al. (2007)) with the help of network analysis and graph theory. There needs to be a clear definition of the goals of the collaboration task. The final confluence of both approaches of modeling collaboration quality (i.e., sensor-based and scenario-based) gives a holistic picture of CC quality detection in a particular scenario.

CC analytics is needed to understand the collaboration processes, make the group aware of how they collaborate and support collaborators to do better team work.

The indicators of CC quality can be broadly grouped into social and epistemic categories. Researchers using the epistemic group of indicators need to pay careful attention to choosing the right sensors and sensing mechanisms. The reason being the semantic limitations of sensors (e.g., to exactly find out when a new idea starts in the conversation) in understanding the content and context of the conversations. Thus, depending on the type of collaboration task, humans combined with sensors can help to speed up the preliminary prototyping and feedback design.

In Part II, aim was to prototype an automatic CC analytics set up using the CC quality definition in Part I. Literature review revealed that content of the conversations can provide rich information about CC quality. This can be a good direction for researchers to focus on. However, designing and prototyping a CC analytics setup to analyze the content of the conversations is difficult to automate fully. So, the limitations in automation need to be kept in mind when thinking about scalability and time necessary to build such a set up. Using interactive network graphs can be a way forward to understand the huge text corpus of conversation among group members.

In Part III, CC analytics is visualized with an aim to move towards quantifying the CC quality. The main recommendation for researchers is to shift their focus from social and epistemic space. Automated analysis of the epistemic space is still underutilized which can be tapped into now with the maturity of the automatic speech recognition techniques and natural language processing. Different epistemic CC quality measures (like convergence in our case) can be implemented efficiently now similar to their manual implementation in the past to understand in depth the content of the conversations. To this end, we have built a generic dashboard to visualize collaboration analytics 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 the role-role exchanges and "what" they interacted. Now, the next important question is: "Who would use it and why?".

This question will be answered by understanding the needs of the dashboard design. This will be an interesting direction for researchers to look into. The dashboard design will be primarily 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) who will be using it. To cater to the temporal needs, a differentiation needs to be made between what can be shown as an immediate formative feedback and what can be shown as a summative feedback at the end of collaboration. To uncover this, a qualitative study needs to be done by interviewing different stakeholders to identify the user requirements. For example, a 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. It is also necessary to distinguish the purpose of the feedback (i.e., whether feedback is used for mirroring or guiding) that needs to be built (Soller et al., 2005).

Therefore, the major outlook for the future will be to measure the quality of collaboration by adding additional modules to this set up and give feedback (an example of feedback in Praharaj et al. (2019)) to facilitate collaboration. Our primary goal will be to use the model of CC in the designed scenarios and then look into different feedback mechanisms that have been built using these indicators to facilitate collaboration. This combined with the indicators of collaboration quality can help to derive the conceptual and implementation model to discover other indexes of collaboration for the community. As a result of which, it will pave the way to form the feedback mechanism to facilitate collaboration for a particular collaboration task.

Besides, due to the emergence of COVID-19, a massive shift of education is happening from the physical space to the online space. Therefore, we are looking into adapting our approach to an online setting from a face-to-face setting. Because of our modular approach, it will be easier to adapt the technical setup. Moreover, we will not need speaker diarization in an online setting, and it will be much easier to get different clean audio streams from each group member in an online setting.

Collaboration analytics and support plays an important role to facilitate collaboration (i.e., a 21st century skill). Collaboration enhances interaction, problem solving and productivity (Kivunja, 2015). Collaboration improves efficiency not only in teaching and learning but also in all walks of life after school (Johnson and Johnson, 1991). Collaboration can be useful for different purposes (such as brainstorming, problem-solving, programming, learning, engineering design) as found in the literature review earlier. Engaging in a collaborative task does not necessarily build collaborative skills (Dillenbourg, 1999); rather on-time feedback and support encourages self-reflection (O'Donnell, 2006). Therefore, effective collaboration analytics and support needs to be implemented.

References

Abdu, R. (Jun. 7–11, 2015). The development of situational-misconceptions in math problem solving. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 78–85, Gothenburg, Sweden.

Ahonen, L., Cowley, B. U., Hellas, A., and Puolamäki, K. (Feb. 2018). Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment. *Scientific Rep.*, 8(1):3138.

Anastasiou, D. and Ras, E. (Mar. 13, 2017). A questionnaire-based case study on feedback by a tangible interface. In *Proc. ACM Workshop Intelligent Interfaces for Ubiquitous & Smart Learning*, pages 39–42, Limassol, Cyprus.

Andrade, A. (Jun. 7–11, 2015). Using situated-action networks to visualize complex learning. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 372–379, Gothenburg, Sweden.

Andrade-Lotero, A., Danish, J., Moreno, J., and Perez, L. (Jun. 15–19, 2013). Measuring "framing" differences of single-mouse and tangible inputs on patterns of collaborative learning. In *Proc. 10th Int. Conf. Computer Supported Collaborative Learning (CSCL'13)*, volume 1, pages 34–41, Madison, WI, USA.

Andrist, S., Ruis, A., and Shaffer, D. W. (Dec. 2018). A network analytic approach to gaze coordination during a collaborative task. *Comput. Human Behav.*, 89:339–348.

Bachour, K., Kaplan, F., and Dillenbourg, P. (2010). An interactive table for supporting participation balance in face-to-face collaborative learning. *IEEE Trans. Learn. Technol.*, 3(3):203–213.

Balaam, M., Fitzpatrick, G., Good, J., and Harris, E. (2011). Enhancing interactional synchrony with an ambient display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 867–876. ACM.

Barkley, E. F., Cross, K. P., and Major, C. H. (2014). *Collaborative learning techniques: A handbook for college faculty*. John Wiley & Sons.

Barthelmess, P., Kaiser, E., Huang, X., and Demirdjian, D. (Oct. 4–6, 2005). Distributed pointing for multimodal collaboration over sketched diagrams. In *Proc. 7th Int. Conf. Multimodal Interfaces (ICMI'05)*, pages 10–17, Torento, Italy.

Bassiou, N. et al. (Sep. 8–12, 2016). Privacy-preserving speech analytics for automatic assessment of student collaboration. In *INTERSPEECH*, pages 888–892, San Francisco, CA, USA.

Bergstrom, T. and Karahalios, K. (Jan. 3–6, 2007). Conversation clock: Visualizing audio patterns in co-located groups. In *Proc. 40th Annu. Hawaii Int. Conf. System Sciences (HICSS'07)*, pages 78–78, Big Island, HI, USA.

Bhattacharya, I. et al. (Oct. 16–20, 2018). A multimodal-sensor-enabled room for unobtrusive group meeting analysis. In *Proc. 20th Int. Conf. Multimodal Interaction (ICMI'18)*, pages 347–355, Boulder, Colorado, USA.

Blikstein, P. (Apr. 8–13, 2013). Multimodal learning analytics. In *Proc. 3rd Int. Conf. Learning Analytics & Knowledge (LAK‚Äô13)*, pages 102–106, Leuven, Belgium.

Çetin, Ö. and Shriberg, E. (May 1–4, 2006). Overlap in meetings: ASR effects and analysis by dialog factors, speakers, and collection site. In *Proc. 3rd Int. Workshop Machine Learning Multimodal Interaction (MLMI'06)*, pages 212–224, Bethesda, MD, USA.

Chandrasegaran, S., Bryan, C., Shidara, H., Chuang, T.-Y., and Ma, K.-L. (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*, pages 1–14.

Chapman, C., Ramondt, L., and Smiley, G. (2005). Strong community, deep learning: Exploring the link. *Innov. Educ. & Teaching Int.*, 42(3):217–230.

Chikersal, P., Tomprou, M., Kim, Y. J., Woolley, A. W., and Dabbish, L. (2017). Deep structures of collaboration: Physiological correlates of collective intelligence and group satisfaction. In *CSCW*, pages 873–888.

Child, S. and Shaw, S. (2015). Collaboration in the 21st century: Implications for assessment. *Econ.*, 21:2008.

Collazos, C. A., Guerrero, L. A., Pino, J. A., and Ochoa, S. F. (Sep. 28–Oct. 2, 2003). Collaborative scenarios to promote positive interdependence among group members. In *Proc. Int. Conf. Collaboration & Technology (CRIWG'03)*, pages 356–370, Autrans, France.

Collazos, C. A., Guerrero, L. A., Pino, J. A., Ochoa, S. F., and Stahl, G. (2007). Designing collaborative learning environments using digital games. *J. Univ. Comput. Sci.*, 13(7):1022–1032.

Coopey, E., Shapiro, R. B., and Danahy, E. (Mar. 24–28, 2014). Collaborative spatial classification. In *Proc. 4th Int. Conf. Learning Analytics & Knowledge (LAK'14)*, pages 138–142, Indianapolis, IN, USA.

Cukurova, M., Luckin, R., Mavrikis, M., and Millán, E. (2017a). Machine and human observable differences in groups' collaborative problem-solving behaviours. In *Proc. 12th European Conf. Technology Enhanced Learning (EC-TEL'17)*, pages 17–29, Tallinn, Estonia.

Cukurova, M., Luckin, R., Millán, E., and Mavrikis, M. (Jan. 2018). The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Comput. & Educ.*, 116:93–109.

Cukurova, M., Luckin, R., Millán, E., Mavrikis, M., and Spikol, D. (Sep. 12–15, 2017b). Diagnosing collaboration in practice-based learning: Equality and intra-individual variability of physical interactivity. In *Proc. 12th European Conf. Technology Enhanced Learning (EC-TEL'17)*, pages 30–42, Tallinn, Estonia.

Cukurova, M., Zhou, Q., Spikol, D., and Landolfi, L. (Mar. 23–27, 2020). Modelling collaborative problem-solving competence with transparent learning analytics: Is video data enough? In *Proc. 10th Int. Conf. Learning Analytics & Knowledge (LAK'20)*, pages 270–275, Frankfurt, Germany.

Das, K., Samanta, S., and Pal, M. (2018). Study on centrality measures in social networks: a survey. *Social network analysis and mining*, 8(1):1–11.

Davidsen, J. and Ryberg, T. (Feb. 2017). 'This is the size of one meter': Children's bodily-material collaboration. *Int. J. Comput.-Supported Collaborative Learn.*, 12(1):65–90.

Dede, C. (2010). Comparing frameworks for 21st century skills. *21st Century Skills: Rethinking how Students Learn*, 20(2010):51–76.

Di Mitri, D., Schneider, J., Drachsler, H., and Specht, M. (2018a). From signals to knowledge. a conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*.

Di Mitri, D., Schneider, J., Specht, M., and Drachsler, H. (Jul. 2018b). From signals to knowledge: A conceptual model for multimodal learning analytics. *J. Comput. Assisted Learn.*, 34(4):338–349.

Dich, Y., Reilly, J., and Schneider, B. (Jun. 27–30, 2018). Using physiological synchrony as an indicator of collaboration quality, task performance and learning. In *Proc. 19th Int. Conf. Artificial Intelligence Education (AIED'18)*, pages 98–110, London, UK.

Dierker, A., Mertes, C., Hermann, T., Hanheide, M., and Sagerer, G. (Nov. 2–4, 2009). Mediated attention with multimodal augmented reality. In *Proc. 2009 Int. Conf. Multimodal Interfaces (ICMI-MLMI'09)*, pages 245–252, Cambridge, MA, USA.

Dillenbourg, P. (1999). What do you mean by collaborative learning? In *Collaborative-learning: Cognitive & Computational Approaches*, pages 1–19. Oxford: Elsevier.

Dillenbourg, P., Baker, M. J., Blaye, A., and O'Malley, C. (1996). The evolution of research on collaborative learning. In *Learning in Humans & Machine: Towards an Interdisciplinary Learning Science*, pages 189–211. Oxford: Elsevier.

Dillenbourg, P. and Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *J. Learn. Sci.*, 15(1):121–151.

DiMicco, J. M., Pandolfo, A., and Bender, W. (2004). Influencing group participation with a shared display. In *Proceedings of the 2004 ACM Conference on CSCW*, pages 614–623. ACM.

Dong, W., Kim, T., and Pentland, A. (Oct. 26–30, 2009). A quantitative analysis of the collective creativity in playing 20-questions games. In *Proc. 7th ACM Conf. Creativity & Cognition (C&C'09)*, pages 365–366, Berkeley, CA, USA.

Dornfeld, C. and Puntambekar, S. (Jun. 7–11, 2015). Emergent roles and collaborative discourse over time. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 380–387, Gothenburg, Sweden.

Dornfeld, C., Zhao, N., and Puntambekar, S. (Jun. 18–22, 2017). A mixed-methods approach for studying collaborative learning processes at individual and group levels. In *Proc. 12th Int. Conf. Computer Supported Collaborative Learning (CSCL'17)*, volume 1, pages 191–198, Philadelphia, PA, USA.

Dyckhoff, A. L. (2014). Action research and learning analytics in higher education. *eleed*, 10(1).

Echeverria, V., Martinez-Maldonado, R., Chiluiza, K., and Shum, S. B. (Dec. 4–8, 2017). DBCollab: Automated feedback for face-to-face group database design. In *Proc. 25th Int. Conf. Computers Education (ICCE'17)*, Christchurch, New Zealand.

Elkins, A. N., Muth, E. R., Hoover, A. W., Walker, A. D., Carpenter, T. L., and Switzer, F. S. (Nov. 2009). Physiological compliance and team performance. *Appl. Ergonom.*, 40(6):997–1003.

Emara, M., Tscholl, M., Dong, Y., and Biswas, G. (Jun. 18–22, 2017). Analyzing students' collaborative regulation behaviors in a classroom-integrated open ended learning environment. In *Proc. 12th Int. Conf. Computer Supported Collaborative Learning (CSCL'17)*, volume 1, pages 319–326, Philadelphia, PA, USA.

Evans, A. C., Wobbrock, J. O., and Davis, K. (Feb. 27–Mar. 2, 2016). Modeling collaboration patterns on an interactive tabletop in a classroom setting. In *Proc. 19th ACM Conf. Computer-Supported Cooperative Work & Social Computing (CSCW'16)*, pages 860–871, San Francisco, CA, USA.

Fake, H., Whiteman, J., and Bannan, B. (Jun. 18–22, 2017). Collaborative game design: A bounded case study of undergraduate students in a capstone course. In *Proc. 12th Int. Conf. Computer Supported Collaborative Learning (CSCL'17)*, volume 1, pages 119–126, Philadelphia, PA, USA.

Fischer, F., Bruhn, J., Gräsel, C., and Mandl, H. (Apr. 2002). Fostering collaborative knowledge construction with visualization tools. *Learn. & Instruction*, 12(2):213–232.

Fischer, G. (Dec. 2000). Symmetry of ignorance, social creativity, and meta-design. *Knowl.-Based Syst.*, 13(7–8):527–537.

Fischer, G. and Ostwald, J. (2005). Knowledge communication in design communities. In Bromme, R., Hesse, F. W., and Spada, H., editors, *Barriers and Biases in Computer-Mediated Knowledge Communication*, pages 213–242. Springer US, Boston, MA.

Flood, V. J., Neff, M., and Abrahamson, D. (Jun. 7–11, 2015). Boundary interactions: Resolving interdisciplinary collaboration challenges using digitized embodied performances. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 94–101, Gothenburg, Sweden.

Granda, R. X., Echeverría, V., Chiluiza, K., and Wong-Villacrés, M. (Jul. 14–16, 2015). Supporting the assessment of collaborative design activities in multi-tabletop classrooms. In *2015 Asia-Pacific Conf. Computer Aided System Engineering (APCASE'15)*, pages 270–275, Quito, Ecuador.

Greller, W. and Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15(3):42–57.

Grover, S., Bienkowski, M., Tamrakar, A., Siddiquie, B., Salter, D., and Divakaran, A. (Apr. 25–29, 2016). Multimodal analytics to study collaborative problem solving in pair programming. In *Proc. 6th Int. Conf. Learning Analytics & Knowledge (LAK'16)*, pages 516–517, New York, NY, USA.

Hardy, L. and White, T. (Jun. 7–11, 2015). Meaning-making in collaborative activity: Effort toward coherent, but not shared, interpretations of the problem. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 435–442, Gothenburg, Sweden.

Hare, A. P. (1994). Types of roles in small groups: A bit of history and a current perspective. *Small Group Res.*, 25(3):433–448.

Harrer, A. (Apr. 8–13, 2013). Analytics of collaborative planning in metafora: Architecture, data, and analytic methods. In *Proc. 3rd Int. Conf. Learning Analytics & Knowledge (LAK'13)*, pages 255–259, Leuven, Belgium.

Hashida, T., Nomura, K., Iida, M., and Naemura, T. (Jun. 15–19, 2013). Interpersonal browsing: Supporting cooperative web searching by face-to-face sharing of browser pages. In *Proc. 10th Int. Conf. Computer Supported Collaborative Learning (CSCL'13)*, volume 1, pages 224–231, Madison, WI, USA.

Healion, D., Russell, S., Cukurova, M., and Spikol, D. (Mar. 13–17, 2017). Tracing physical movement during practice-based learning through multimodal learning analytics. In *Proc. 7th Int. Conf. Learning Analytics & Knowledge (LAK'17)*, pages 588–589, Vancouver, Canada.

Henning, R., Armstead, A., and Ferris, J. (Nov. 2009). Social psychophysiological compliance in a four-person research team. *Appl. Ergonom.*, 40(6):1004–1010.

Henning, R. A., Boucsein, W., and Gil, M. C. (Apr. 2001). Social–physiological compliance as a determinant of team performance. *Int. J. Psychophysiology*, 40(3):221–232.

Huber, B., Shieber, S., and Gajos, K. Z. (2019). Automatically analyzing brainstorming language behavior with meeter. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–17.

Isenberg, P., Fisher, D., Morris, M. R., Inkpen, K., and Czerwinski, M. (Oct. 24–29, 2010). An exploratory study of co-located collaborative visual analytics around a tabletop display. In *Proc. 2010 IEEE Symp. Visual Analytics Science & Technology (VAST'10)*, pages 179–186, Salt Lake City, UT, USA.

Jeong, H. and Chi, M. T. (2007). Knowledge convergence and collaborative learning. *Instructional Science*, 35(4):287–315.

Jeong, H. and Hmelo-Silver, C. E. (Jun. 29–Jul. 2, 2010). An overview of CSCL methodologies. In *Proc. 9th Int. Conf. Learning Sciences (ICLS'10)*, volume 1, pages 921–928, Chicago, IL, USA.

Jermann, P., Mullins, D., Nüssli, M.-A., and Dillenbourg, P. (2011). Collaborative gaze footprints: correlates of interaction quality. In *CSCL 2011 Conference Proceedings*, volume 1, pages 184–191. International Society of the Learning Sciences.

Johansson, C., Ahmet, Z., Tholander, J., Aleo, F., Jonsson, M., and Sumon, S. (July 4–8, 2011). Weather Gods and Fruit Kids—Embodying abstract concepts using tactile feedback and whole body interaction. In *Proc. 9th Int. Conf. Computer Supported Collaborative Learning (CSCL'11)*, pages 160–167, Hong Kong, China.

Johnson, D. W. and Johnson, F. P. (1991). *Joining together: Group theory and group skills*. Prentice-Hall, Inc.

Johnson, D. W. and Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, 38(5):365–379.

Kim, J. et al. (Sep. 6–10, 2015). Vocal turn-taking patterns in groups of children performing collaborative tasks: An exploratory study. In *INTERSPEECH*, Dresden, Germany.

Kim, T., Chang, A., Holland, L., and Pentland, A. S. (Nov. 8–12, 2008). Meeting mediator: Enhancing group collaboration using sociometric feedback. In *Proc. 2008 ACM Conf. Computer Supported Cooperative Work (CSCW'08)*, pages 457–466, San Diego, CA, USA.

Kivunja, C. (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 Educ.*, 6(02):224.

Koh, E., Shibani, A., Tan, J. P.-L., and Hong, H. (2016). A pedagogical framework for learning analytics in collaborative inquiry tasks: An example from a teamwork competency awareness program. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, pages 74–83.

Kulyk, O., Wang, J., and Terken, J. (2005). Real-time feedback on nonverbal behaviour to enhance social dynamics in small group meetings. In *International WS on Machine Learning for Multimodal Interaction*, pages 150–161. Springer.

Li, W., Nüssli, M.-A., and Jermann, P. (Nov. 8–10, 2010). Gaze quality assisted automatic recognition of social contexts in collaborative Tetris. In *Proc. 2010 Int. Conf. Multimodal Interfaces (ICMI-MLMI'10)*, pages 1–8, Beijing, China.

Liberati, A. et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, 6(7):e1–e34.

Lubold, N. and Pon-Barry, H. (Nov. 12, 2014). Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In *Proc. 2014 ACM Workshop Multimodal Learning Analytics Workshop & Grand Challenge (MLA'14)*, pages 5–12, Istanbul, Turkey.

Luz, S. (Dec. 9–13, 2013). Automatic identification of experts and performance prediction in the multimodal math data corpus through analysis of speech interaction. In *Proc. 15th ACM Int. Conf. Multimodal Interaction (ICMI'13)*, pages 575–582, Sydney, Australia.

Madan, A., Caneel, R., and Pentland, A. S. (2004). Groupmedia: distributed multimodal interfaces. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, pages 309–316. ACM.

Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X., and Siipo, A. (Jul. 2019). Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Comput. Human Behav.*, 96:235–245.

Manske, S., Hecking, T., Chounta, I.-A., Werneburg, S., and Hoppe, H. U. (Jun. 7–11, 2015). Using differences to make a difference: A study on heterogeneity of learning groups. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 182–189, Gothenburg, Sweden.

Martin, N. D., Gnesdilow, D., and Puntambekar, S. (Jun. 7–11, 2015). Peer scaffolding to learn science in symmetrical groups collaborating over time. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 340–347, Gothenburg, Sweden.

Martinez, R., Kay, J., and Yacef, K. (July 4–8, 2011). Visualisations for longitudinal participation, contribution and progress of a collaborative task at the tabletop. In *Proc. 9th Int. Conf. Computer Supported Collaborative Learning (CSCL'11)*, pages 25–32, Hong Kong, China.

Martinez-Maldonado, R., Clayphan, A., Yacef, K., and Kay, J. (Apr.-Jun. 2015a). MTFeedback: Providing notifications to enhance teacher awareness of small group work in the classroom. *IEEE Trans. Learn. Technol.*, 8(2):187–200.

Martinez-Maldonado, R., Dimitriadis, Y., Martinez-Monés, A., Kay, J., and Yacef, K. (Nov. 2013). Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. *Int. J. Comput.-Supported Collaborative Learn.*, 8(4):455–485.

Martinez-Maldonado, R. et al. (Jun. 7–11, 2015b). Learning about collaborative design for learning in a multi-surface design studio. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 174–181, Gothenburg, Sweden.

Martinez-Maldonado, R. et al. (Mar. 13–17, 2017a). Analytics meet patient manikins: Challenges in an authentic small-group healthcare simulation classroom. In *Proc. 7th Int. Conf. Learning Analytics & Knowledge (LAK'17)*, pages 90–94, Vancouver, Canada.

Martinez-Maldonado, R., Pechenizkiy, M., Buckingham-Shum, S., Power, T., Hayes, C., and Axisa, C. (Jul. 9–12, 2017b). Modelling embodied mobility teamwork strategies in a simulation-based healthcare classroom. In *Proc. 25th Conf. User Modeling, Adaptation & Personalization (UMAP'17)*, pages 308–312, Bratislava, Slovakia.

McBride, E., Vitale, J., Applebaum, L., and Linn, M. (Jun. 18–22, 2017). Examining the flow of ideas during critique activities in a design project. In *Proc. 12th Int. Conf. Computer Supported Collaborative Learning (CSCL'17)*, volume 1, pages 41–48, Philadelphia, PA, USA.

Meier, A., Spada, H., and Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *Int. J. Comput.-Supported Collaborative Learn.*, 2(1):63–86.

Nakano, Y. I., Nihonyanagi, S., Takase, Y., Hayashi, Y., and Okada, S. (Nov. 9–13, 2015). Predicting participation styles using co-occurrence patterns of nonverbal behaviors in collaborative learning. In *Proc. 17th ACM Int. Conf. Multimodal Interaction (ICMI'15)*, pages 91–98, Seattle, WA, USA.

Noel, R. et al. (Oct. 2018). Exploring collaborative writing of user stories with multimodal learning analytics: A case study on a software engineering course. *IEEE Access*, 6:67783–67798.

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., and Castells, J. (Dec. 9–13, 2013). Expertise estimation based on simple multimodal features. In *Proc. 15th ACM Int. Conf. Multimodal Interaction (ICMI'13)*, pages 583–590, Sydney, Australia.

O'Donnell, A. M. (2006). The role of peers and group learning.

OECD (2017). *PISA 2015 collaborative problem-solving framework*.

Ofstedal, K. and Dahlberg, K. (2009). Collaboration in student teaching: Introducing the collaboration self-assessment tool. *Journal of Early Childhood Teacher Education*, 30(1):37–48.

Olsen, J. K. and Finkelstein, S. (Jun. 18–22, 2017). Through the (thin-slice) looking glass: An initial look at rapport and co-construction within peer collaboration. In *Proc. 12th Int. Conf. Computer Supported Collaborative Learning (CSCL'17)*, volume 1, pages 511–518, Philadelphia, PA, USA.

Oshima, J., Ohsaki, A., Yamada, Y., and Oshima, R. (Jun. 18–22, 2017). Collective knowledge advancement and conceptual understanding of complex scientific concepts in the jigsaw instruction. In *Proc. 12th Int. Conf. Computer Supported Collaborative Learning (CSCL'17)*, volume 1, pages 57–64, Philadelphia, PA, USA.

Oviatt, S., Hang, K., Zhou, J., and Chen, F. (Nov. 9–13, 2015). Spoken interruptions signal productive problem solving and domain expertise in mathematics. In *Proc. 17th ACM Int. Conf. Multimodal Interaction (ICMI'15)*, pages 311–318, Seattle, WA, USA.

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., and Kirschner, P. A. (Apr. 25–29, 2016). Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. In *Proc. 6th Int. Conf. Learning Analytics & Knowledge (LAK'16)*, pages 64–73, Edinburgh, UK.

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., and Kirschner, P. A. (Mar. 2019). Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members? *Comput. Human Behav.*, 92:188–197.

Pijeira-Díaz, H. J., Drachsler, H., Kirschner, P. A., and Järvelä, S. (2018). Profiling sympathetic arousal in a physics course: How active are students? *Journal of Computer Assisted Learning*.

Praharaj, S. (Oct. 14–18, 2019). Co-located collaboration analytics. In *Proc. 21st Int. Conf. Multimodal Interaction (ICMI'19)*, pages 473–476, Suzhou, China.

Praharaj, S., Scheffel, M., Drachsler, H., and Specht, M. (2019). Group coach for co-located collaboration. In *European Conference on Technology Enhanced Learning*, pages 732–736. Springer.

Praharaj, S., Scheffel, M., Drachsler, H., and Specht, M. (2021a). 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):367–385.

Praharaj, S., Scheffel, M., Drachsler, H., and Specht, M. (Sep. 3–5, 2018a). MULTIFOCUS: Multimodal learning analytics for co-located collaboration understanding and support. In *Proc. 13th European Conf. Technology Enhanced Learning (EC-TEL'18) Doctoral Consortium*, Leeds, UK.

Praharaj, S., Scheffel, M., Drachsler, H., and Specht, M. (Sep. 3–6, 2018b). Multimodal analytics for real-time feedback in co-located collaboration. In *Proc. 13th European Conf. Technology Enhanced Learning (EC-TEL'18)*, pages 187–201, Leeds, UK.

Praharaj, S., Scheffel, M., Schmitz, M., Specht, M., and Drachsler, H. (May 2021b). Towards automatic collaboration analytics for group speech data using learning analytics. *Sensors*, 21(9):3156.

Reilly, J. M., Ravenell, M., and Schneider, B. (Jul. 16–20, 2018). Exploring collaboration using motion sensors and multi-modal learning analytics. In *Proc. Int. Conf. Educational Data Mining (EDM'18)*, Raleigh, NC, USA.

Richardson, D. C. and Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29(6):1045–1060.

Rittel, H. (1984). Second-generation design methods. *Developments in design methodology*, pages 317–327.

Rodríguez, F. J., Price, K. M., and Boyer, K. E. (Jun. 18–22, 2017). Expressing and addressing uncertainty: A study of collaborative problem-solving dialogues. In *Proc. 12th Int. Conf. Computer Supported Collaborative Learning (CSCL'17)*, volume 1, pages 207–214, Philadelphia, PA, USA.

Scherer, S., Weibel, N., Morency, L.-P., and Oviatt, S. (Oct. 26, 2012). Multimodal prediction of expertise and leadership in learning groups. In *Proc. 1st Int. Workshop Multimodal Learning Analytics (MLA'12)*, pages 1–8, Santa Monica, CA, USA.

Scherr, R. E. and Hammer, D. (Mar. 2009). Student behavior and epistemological framing: Examples from collaborative active-learning activities in physics. *Cognition & Instruction*, 27(2):147–174.

Schmitz, M., Scheffel, M., Bemelmans, R., and Drachsler, H. (2019). Fellowship of the learning activity - learning analytics 4 learning design game.

Schneider, B. and Blikstein, P. (Oct. 2015). Unraveling students' interaction around a tangible interface using multimodal learning analytics. *J. Educ. Data Mining*, 7(3):89–116.

Schneider, B. and Pea, R. (2015). Does seeing one another's gaze affect group dialogue? A computational approach. *J. Learn. Analytics*, 2(2):107–133.

Schneider, B. and Pea, R. (July 4–7, 2014a). The effect of mutual gaze perception on students' verbal coordination. In *Proc. 7th Int. Conf. Educational Data Mining (EDM'14)*, pages 138–144, London, UK.

Schneider, B. and Pea, R. (Oct. 2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *Int. J. Comput.-Supported Collaborative Learn.*, 8(4):375–397.

Schneider, B. and Pea, R. (Oct. 2014b). Toward collaboration sensing. *Int. J. Comput.-Supported Collaborative Learn.*, 9(4):371–395.

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., and Pea, R. D. (Jun. 7–11, 2015). 3D tangibles facilitate joint visual attention in dyads. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 156–165, Gothenburg, Sweden.

Schneider, J., Di Mitri, D., Limbu, B., and Drachsler, H. (2018). Multimodal learning hub: a tool for capturing customizable multimodal learning experiences. In *Proceedings of the 13th EC-TEL*. Springer.

Shih, P. C., Nguyen, D. H., Hirano, S. H., Redmiles, D. F., and Hayes, G. R. (2009). Groupmind: supporting idea generation through a collaborative mind-mapping tool. In *Proc. of the ACM 2009 Int. Conf. on Supporting Group Work*, pages 139–148. ACM.

Siemens, G. (2011). Learning analytics & knowledge: February 27-march 1, 2011. Banff.

Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *Int. J. Artif. Intell. Educ.*, 12:40–62.

Soller, A., Martinez, A., Jermann, P., and 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.

Spikol, D., Ruffaldi, E., and Cukurova, M. (Jun. 18–22, 2017a). Using multimodal learning analytics to identify aspects of collaboration in project-based learning. In *Proc. 12th Int. Conf. Computer Supported Collaborative Learning (CSCL'17)*, volume 1, pages 263–270, Philadelphia, PA, USA.

Spikol, D., Ruffaldi, E., Dabisias, G., and Cukurova, M. (2018a). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning*.

Spikol, D., Ruffaldi, E., Dabisias, G., and Cukurova, M. (May 2018b). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *J. Comput. Assisted Learn.*, 34(4):366–377.

Spikol, D., Ruffaldi, E., Landolfi, L., and Cukurova, M. (Jul. 3–7, 2017b). Estimation of success in collaborative learning based on multimodal learning analytics features. In *Proc. 17th IEEE Int. Conf. Advanced Learning Technologies (ICALT'17)*, pages 269–273, Timisoara, Romania.

Stahl, G., Law, N., and Hesse, F. (Oct. 2013). Reigniting CSCL flash themes. *Int. J. Comput.-Supported Collaborative Learn.*, 8(4):369–374.

Starr, E. L., Reilly, J. M., and Schneider, B. (Jun. 23–27, 2018). Toward using multi-modal learning analytics to support and measure collaboration in co-located dyads. In *Proc. 13th Int. Conf. Learning Sciences (ICLS'18)*, London, UK.

Stewart, A. E., Keirn, Z. A., and D'Mello, S. K. (Oct. 16–20, 2018). Multimodal modeling of coordination and coregulation patterns in speech rate during triadic collaborative problem solving. In *Proc. 20th Int. Conf. Multimodal Interaction (ICMI'18)*, pages 21–30, Boulder, Colorado, USA.

Stiefelhagen, R. and Zhu, J. (April 20–25, 2002). Head orientation and gaze direction in meetings. In *Proc. Int. Conf. Extended Abstracts Human Factors Computing Systems (CHI EA'02)*, pages 858–859, Minneapolis, MN, USA.

Strijbos, J.-W., Martens, R. L., Prins, F. J., and Jochems, W. M. (2006). Content analysis: What are they talking about? *Computers & education*, 46(1):29–48.

Strijbos, J.-W. and Weinberger, A. (Jul. 2010). Emerging and scripted roles in computer-supported collaborative learning. *Comput. Human Behav.*, 26(4):491–494.

Tausch, S., Hausen, D., Kosan, I., Raltchev, A., and Hussmann, H. (Oct. 26–30, 2014). Groupgarden: Supporting brainstorming through a metaphorical group mirror on table or wall. In *Proc. 8th Nordic Conf. Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI'14)*, pages 541–550, Helsinki, Finland.

Teasley, S., Fischer, F., Dillenbourg, P., Kapur, M., Chi, M., Weinberger, A., and Stegmann, K. (2008). Cognitive convergence in collaborative learning.

Terken, J. and Sturm, J. (March, 2010). Multimodal support for social dynamics in co-located meetings. *Pers. & Ubiquitous Comput.*, 14(8):703–714.

Thompson, K., Ashe, D., Yeoman, P., and Parisio, M. (Jun. 15–19, 2013). Phases of design: Following idea development and patterns of collaborative discussion in a learning by design project. In *Proc. 10th Int. Conf. Computer Supported Collaborative Learning (CSCL'13)*, volume 1, pages 494–501, Madison, WI, USA.

Thompson, K., Kennedy-Clark, S., Markauskaite, L., and Southavilay, V. (2014). Discovering processes and patterns of learning in collaborative learning environments using multi-modal discourse analysis. *Res. & Pract. Technol. Enhanced Learn.*, 9(2).

Triglianos, V., Praharaj, S., Pautasso, C., Bozzon, A., and Hauff, C. (2017). Measuring student behaviour dynamics in a large interactive classroom setting. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, pages 212–220. ACM.

Viswanathan, S. A. and VanLehn, K. (Jun. 18–22, 2017). High accuracy detection of collaboration from log data and superficial speech features. In *Proc. 12th Int. Conf. Computer Supported Collaborative Learning (CSCL'17)*, volume 1, pages 335–342, Philadelphia, PA, USA.

Wake, J. D., Guribye, F., and Wasson, B. (Jun. 7–11, 2015). Students' design decisions in collaborative design of location-based games for learning. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 324–331, Gothenburg, Sweden.

Weinberger, A., Ertl, B., Fischer, F., and Mandl, H. (2005). Epistemic and social scripts in computer–supported collaborative learning. *Instructional Science*, 33(1):1–30.

Wise, A. F., Antle, A. N., and Warren, J. (Jun. 18–22, 2017). Explanation-giving in a collaborative tangible tabletop game: Initiation, positionality, valence and action-orientation. In *Proc. 12th Int. Conf. Computer Supported Collaborative Learning (CSCL'17)*, volume 1, pages 471–478, Philadelphia, PA, USA.

Wise, A. F., Antle, A. N., Warren, J., May, A., Fan, M., and Macaranas, A. (Jun. 7–11, 2015). What kind of world do you want to live in? Positive interdependence and collaborative processes in the tangible tabletop land-use planning game youtopia. In *Proc. 11th Int. Conf. Computer Supported Collaborative Learning (CSCL'15)*, volume 1, pages 236–243, Gothenburg, Sweden.

Wong, L. H., Chen, W., Zhan, Y., and Chin, C. K. (July 4–8, 2011). Analysis of small group interactions in a seamless language learning environment: An artifact-oriented approach. In *Proc. 9th Int. Conf. Computer Supported Collaborative Learning (CSCL'11)*, pages 144–151, Hong Kong, China.

Wong-Villacrés, M., Granda, R., Ortiz, M., and Chiluiza, K. (Apr. 25–29, 2016). Exploring the impact of a tabletop-generated group work feedback on students' collaborative skills. In *Int. Workshop Learning Analytics Across Physical & Digital Spaces (CrossLAK'16)*, pages 58–64, Edinburgh, UK.

Worsley, M. and Blikstein, P. (2018). A multimodal analysis of making. *Int. J. Artif. Intell. Educ.*, 28(3):385–419.

Worsley, M. and Blikstein, P. (Mar. 16–20, 2015). Leveraging multimodal learning analytics to differentiate student learning strategies. In *Proc. 5th Int. Conf. Learning Analytics & Knowledge (LAK'15)*, pages 360–367, Poughkeepsie, NY, USA.

Xie, K., Di Tosto, G., Lu, L., and Cho, Y. S. (2018). Detecting leadership in peer-moderated online collaborative learning through text mining and social network analysis. *The Internet and Higher Education*, 38:9–17.

Zhou, J., Hang, K., Oviatt, S., Yu, K., and Chen, F. (Nov. 12, 2014). Combining empirical and machine learning techniques to predict math expertise using pen signal features. In *Proc. 2014 ACM Workshop on Multimodal Learning Analytics Workshop & Grand Challenge (MLA'14)*, pages 29–36, Istanbul, Turkey.

Appendices

Appendix A

A sample consent form in Dutch (used in Chapter 3 and 4)

Toestemmingsformulier

Tijdens dit experiment worden er op verschillende manieren gegevens verzameld:

- Er is een vragenlijst over hoe vaak u reeds met FoLA2 gewerkt hebt. In deze vragenlijst wordt gevraagd hoeveel jaren ervaring u heeft betreffende de rol die u gaat spelen.
- Er is een vragenlijst, gebaseerd op de UTUAT2 vragenlijst die inzicht verschaft in de kennis, ervaring en de verwachting rondom learning analytics.
- Er worden videobeelden gemaakt van het spelbord en de omgeving. Deze worden alleen voor interne analyse gebruikt over het spelverloop.
- Gedurende het spel wordt de stand van het spelbord opgeslagen.
- Er worden per persoon geluidsopnames per deelnemer gemaakt. Die worden gebruikt voor twee doeleinden:
    - 1) Analyse doeleinden in het onderzoek van Marcel Schmitz (Zuyd en OU) om het spelverloop in kaart te brengen
    - 2) Analyse doeleinden voor het onderzoek van Sambit Praharaj (OU), dat als doel heeft om samenwerking binnen de groep inzichtelijk te maken middels tekstanalyse gebaseerd op verschillende rollen.

De persoonsgegevens die worden opgeslagen zijn de ervaringsgegevens betreffende de rol die gespeeld wordt tijdens het spel. Deze worden niet gekoppeld aan andere persoonsgegevens.

Alle gegevens worden geanonimiseerd en daarna opgeslagen. De audiogegevens zullen via NLP omgezet worden in tekst en geanalyseerd worden. Na afloop van het PhD onderzoek zullen de videogegevens verwijderd worden. De audiogegevens zullen geanonimiseerd worden opgeslagen en na maximaal vijf jaar verwijderd worden. De videogegevens zullen 1 jaar na afloop van het onderzoek verwijderd worden.

Mocht u na het geven van de toestemming toch willen dat de geluidsopnames en ervaringsgegevens verwijderd worden dan kunt u een mail sturen naar marcel.schmitz@zuyd.nl. Deze gegevens worden dan verwijderd.

O Ik geef geen toestemming voor het gebruik van de informatie die over mij is verzameld.

O Ik geef toestemming voor het gebruik van de over mij verzamelde informatie voor het onderzoeksproject van Marcel, maar niet van Sambit.

O Ik geef toestemming voor het gebruik van de informatie die over mij is verzameld voor de onderzoeksprojecten van Marcel en van Sambit.

Naam, `datum, locatie, handtekening

____

Appendix B

Overview of social space (i.e., dominant role-role exchanges in terms of speaking time and turn taking) and epistemic space (i.e., relevant bigrams) for individual roles in red phase across all sessions (Chapter 4)

Sessions	All Advisors (A)	Study Coach (S)	Teacher (T)	Learner (L)
1 (A-T, A-S)	Smart screen, make photo, mobile phone, technology need, technology effort, online collaboration, screen wiki, mobile phone, make post, post its, teacher prepared	Concept mapping, map tool, question test, then Belbin, finally visualizer, necessarily mindmap, interesting wearable, action finally	Team composition, fourth block, trigger interaction, blog vlog, various roles, vlog wiki, mobile telephone, make mindmap, group outcome, ideal group, make post, post its	Via moodle, moodle work, new team, form film, good post, post its, its use
2 (A-T, A-S, A-L, T-S)	Interaction booster, technology need, smart screen, teacher student, normal interaction	Team attribute, discussion attribute, different context, stay screen, wordcloud or so	New group, old group, question pair, screen on top, question DUO, physically screen, target interaction	Demand interaction, team communication, more interaction, year shakespeak, teacher communication
3 (A-T, T-S)	No screen, smart screen, interaction booster, technology need	Same shakespeak, smart screen, last evaluation goal, interaction booster	Interaction booster, online collaboration, bring lesson material, collaboration environment, environment moodle, smartscreen note, make note	Make problem, student interaction, search quiz
5 (A-T, A-L, T-S, T-L)	Quiz moodle, apply knowledge test, make telephone, study place, paper digital, conflict handling	Shakespeak mean, write paper, red card, basic conflict, learning technology card, make app, learning technology reach	Oral class, individual input, make quiz, criticize DUO, teacher analysis, laptop oral, late student	Video clip commercial, mobile person, easy laptop, laptop pack, classroom mobile, click send
6 (A-T)	Leave smartscreen, smartscreen need, handout paper	Moodle mail, question answer, personal reflection, receive smartscreen, learn material, leave tool	Learning analytics, whiteboard like that, yes upload, clear communication, student week, finally paper, possibly board, student trust	Moodle body, pairing rules
7 (A-T, A-S)	Online collaboration, shakespeak after all, videoclip target, make online	Monty python, extroverted person, extroverted student, want shakespeak, person exclusive, random tool	Interaction booster, state feedback, student class, live question, ask teacher, expected feedback	Monty python, smart board, smart screen, send shakespeak, personal grade, write question
8 (A-T, A-L, T-L)	Shakespeak form, smart screen, excites student, discussion manner, data use, student app	Lubach film, good film, discussion trigger, at student	Save data, manner data, good question	Interaction booster, teacher learner, real questionnaire, minutes time, via shakespeak, say quiz, smartscreen stake
9 (A-T)	Clear padlet, smart screen, describe assignment, padlet lecture, booster shakespeak, screen use, moodle question mark, whiteboard kind of	Present powerpoint, map tool, manner data, concept map, data visualizer	Smart board, see histogram, see graphics, last whiteboard, whiteboard number, padlet open	Little moodle, search internet, smart screen, assignment document, collect data, verbally explain, smartphone screen, email teacher, inform data
10 (A-T, A-L)	Smart board, padlet wall, mindmap tool, group mindmap, interaction booster, put padlet, aspect survey, survey think, yesterday statistic	[No contribution]	End evaluation, student smart, smart board, padlet wall, student discount, student access	Map tool, listen online, bad assignment, concept map
11 (T-L)	The sendstep, difficult discussion, work online, go online, find present	Shakespeak pair, via shakespeak, ask evaluation question, qualitative survey	Reflective feedback, feedback present, technology need, supply shakespeak, question answer	Smart screen, throw inbox, make shakespeak, online things, need dropbox, quarter click, online difficult, difficult discussion
12 (T-S, L-S)	More padlet, go reflect, digital white, white board	Put smartscreen, interaction booster, student paper, booster use, find feedback, think DUO	Moodle quiz, leave creativity, yes digital, survey form, target pair	Naturally laptop, make powerpoint, use teacher, target question
14 (A-T, A-S)	Put moodle, interaction booster, enclose mindmap, say videoclip, learning analytics, yes results	Complex theory, class result, unclear moodle, Socratic lesson, self study form, quiz result, moodle result, extended abstract	Student theory, activate inside knowledge, watched quiz, quiz question	Mind maps, maps recap, make mindmap, finally observe

Appendix C

Code Repository

This website contains all the code used in different chapters and how to replicate the data processing and analysis for the different studies in Chapter 1, Chapter 2, Chapter 3 and Chapter 4.

https://bit.ly/SambitPhDThesis

List of Tables

List of Figures

Summary

Collaboration is one of the 4 important 21st century skills (Kivunja, 2015). When two or more people work towards a common goal then collaboration is said to occur (Dillenbourg, 1999). Collaboration can take place either in an online setting or in a co-located (or face-to-face) setting. The measurement of online collaboration processes is possible due to the measurement, collection and analysis of the learner data using learning analytics (Siemens, 2011; Greller and Drachsler, 2012). With the ubiquitous usage of sensors lately, a new branch of learning analytics otherwise known as multimodal learning analytics (MMLA) has risen to prominence (Di Mitri et al., 2018a; Martinez-Maldonado et al., 2017a). Moreover, sensor technology has become more scalable (Reilly et al., 2018), affordable and reliable in the past decade (Starr et al., 2018). There is a focus shift towards the analysis of CC due to the rise of MMLA.

In this thesis we focused on co-located (or face-to-face) collaboration. Co-located collaboration takes place in the intersection of physical, social and epistemic space of the group members. Collaboration quality can be detected using different indicators of collaboration in different spaces (i.e., epistemic and social spaces) (Praharaj et al., 2018b). The social space comprises the non-verbal indicators (posture, gesture, eye gaze and non-verbal audio indicators like speaking time, pitch, turn-taking). The epistemic space comprises the verbal audio indicators (such as the content of the conversation) and content log data.

The main objectives of this thesis are: 1) To define and understand the co-located collaboration (CC) quality and the analytics; 2) To develop a technical prototype and test it with field trials to move towards automated collaboration analytics using the aforementioned definitions; 3) To visualize the CC analytics and move towards quantifying CC quality.

To fulfil these objectives, we have subdivided the thesis into three parts and four chapters. In **Part I** consisting of 2 chapters (i.e., Chapter 1 and 2), we describe the definition for CC quality, and how quality is contextualised in certain scenarios of collaboration. We also connect quality to suitable analytics and feedback mechanisms to support learners in collaboration. In Chapter 1, we do an exploratory state-of-the-art review to understand how indicators help to detect the quality of collaboration. Then we looked into the studies on CC feedback and analytics to understand how these indicators helped to facilitate collaboration. For example, during collaborative meeting, total speaking time was used as an indicator of collaboration quality to give real-time feedback to support collaboration. The real-time reflective feedback was shown by glowing the required number of LED lights (i.e., proportional to the

total speaking time) in front of that group member on a smart table. This helped to create a balance between the participants who expressed more verbally (i.e., who spoke more) and those who expressed less verbally (i.e., who spoke less); thereby improving the quality of collaboration. Therefore, based on the review of these indicator, feedback examples from several studies, we designed a hybrid set up (with the combination of human observers and sensors like microphone) to test real-time feedback with the help of a small field study. We tested this set up during PhD meetings and tracked two indicators of collaboration quality (i.e., total speaking time and turn taking). Then we showed these indicators as a reflective feedback during the meeting on a large public shared display as temporal graphs. The aim of this feedback set up was to get a feel of it instead of testing the efficacy of the feedback on collaboration quality.

Our findings indicate that the indicators of CC quality can be grouped into two categories (i.e., social and epistemic). The group of epistemic indicators cannot be detected and analyzed alone by senors and need the help of humans because of the semantic nature of understanding them. Furthermore, CC analytics and feedback design is dependent on the stakeholders and having human in the loop helps to speed up the design.

To understand this further, we do an in-depth literature review of the indicators of collaboration quality in Chapter 2. It is because the indicators of collaboration vary depending on the scenarios and the context of collaboration. Here, we define the quality of collaboration with a event-process framework made up of the indicators and indexes. The indicators are low-level events obtained after processing and aggregation from the sensors. The indexes (i.e., high level processes) act as the measurable markers helping to detect the quality of collaboration. The are made up of one or more indicators. For instance, in collaborative meetings (i.e., the scenario of CC), the equality (i.e., the index) of the total speaking time (i.e., the indicator) measures the quality of collaboration. If all group members have similar total speaking time with no one dominating the conversation then there is higher equality of total speaking time for the group and better quality of collaboration. These indicators vary across different scenarios because of the differing goals and parameters (i.e., primary aspects such as team composition, behaviour of team members and behaviour during collaboration) of CC. For example, indicators of CC quality for collaborative programming can differ completely from indicators of collaborative brainstorming.

Thus, with the help of this literature review, we define a conceptual model that encompasses the indicators, indexes and parameters to detect the CC quality. In this model, we map the parameters in different scenarios onto the indicators and indexes to pave the way for designing of a CC quality detection and prediction system. Our findings also indicate that the operationalization of CC quality suffers from coding complexity and use of opaque machine learning algorithms. There is a huge gap between the theoretical indexes (i.e., measurable markers) of CC quality and the practically detected indexes. This needs to be bridged.

Using this definition of CC quality, we focus on audio-based indicators of collaboration in **Part II**, Chapter 3. Audio is the dominantly used modality as found in the literature review and very easy to capture with microphones. Most prior studies focused on "how" group members speak and not on "what" they speak. Thus, the focus was on the social space (comprising of non-verbal audio indicators such as total speaking time, change in pitch) and not on the epistemic space (comprising of the content of the conversations) of the audio modality. Very few works focused on the epistemic space semi-automatically in a lab setting using manual coding. These approaches were based on predefined conditions, gave abstract overview of the topics of discussion and laborious to implement. So, we developed a prototype to overcome this and analyze the richness of the epistemic space in an authentic real world setting with the help of field trials. For this, we recorded the audio conversations during a CC task where university staff played a board game with pre-assigned roles to create awareness of the connection between learning analytics and learning design. We transcribed these audio recordings (i.e., convert from speech to text), processed them and then visualized them (using network graphs to understand the interconnected nature of the spoken text). For this, we also did a role-based profiling to get a holistic overview of the conversations in an automatic manner. We tested this prototype for one CC session with an aim to make a step towards automatic collaboration analytics.

Our findings indicate that content of the conversations can provide rich information about CC quality. With the help of interactive network graphs, rich insights about the interconnected conversation patterns between group members can be obtained. However, designing and prototyping a CC analytics set up is difficult to have full automation. We need help of humans to clean the data corpus especially when name is uttered in the conversations.

Using the developed prototype, we moved towards quantifying the quality of collaboration in **Part III**, Chapter 4 with the help of field trials across 14 different CC sessions played in an university setting using board games. We did a holistic analysis of the social (i.e., total speaking time and turn taking) and epistemic (i.e., the content of the conversations) space also considering the role based contributions and interactions and then visualize it. We define quality of collaboration taking into account the convergence of the discussion (i.e., shared epistemic space knowledge as analyzed from the content of the conversations) among the group members with different roles. Finally, we visualized both the social and epistemic space using a dashboard; then discuss the stakeholders who can use such a dashboard and what future research can be done on that dashboard.

We find that most of the focus on CC analytics is now on the social space and it needs to shift to the epistemic space. CC quality measures such as convergence (i.e., an increase in shared knowledge among group members) measured from the epistemic space can be useful to quantify the CC quality. Both the social and epistemic space can be useful to give a holistic view of the CC quality when visualized on a single dashboard.

In the **Discussion**, based on the findings listed above, we derive and elaborate on a mix of observations and recommendations. We grouped these according to the phase of the research on CSCL and CC analytics: defining the CC quality and analytics in the early phase, prototyping the automatic CC analytics set up in the mid phase and visualizing the CC analytics to move towards the quantification of CC quality in the late phase.

Samenvatting

Samenwerken is een van de vier belangrijke 21este-eeuwse vaardigheden (Kivunja, 2015). Wanneer twee of meer mensen werken aan een gemeenschappelijk doel, dan is er sprake van samenwerking (Dillenbourg, 1999). Samenwerking kan zowel in een online setting als in een co-located (of face-to-face) setting plaatsvinden. Het meten van online samenwerkingsprocessen is mogelijk door het meten, verzamelen en analyseren van de gegevens van leerlingen met behulp van learning analytics (Siemens, 2011; Greller and Drachsler, 2012). Met het alomtegenwoordige gebruik van sensoren de laatste tijd, is een nieuwe tak van learning analytics, ook wel bekend als multimodale learning analytics (MMLA), op de voorgrond getreden (Di Mitri et al., 2018a; Martinez-Maldonado et al., 2017a). Bovendien is sensortechnologie in het afgelopen decennium schaalbaarder (Reilly et al., 2018), betaalbaar en betrouwbaar geworden (Starr et al., 2018). Door de opkomst van MMLA is de aandacht verschoven naar de analyse van CC.

In dit proefschrift hebben we ons gericht op co-located (of face-to-face) samenwerking. Co-located samenwerking vindt plaats op het snijvlak van fysieke, sociale en epistemische ruimte van de groepsleden. De kwaliteit van samenwerking kan worden gedetecteerd met behulp van verschillende indicatoren van samenwerking in verschillende ruimtes (d.w.z. epistemische en sociale ruimtes, (Praharaj et al., 2018b)). De sociale ruimte omvat de non-verbale indicatoren (houding, gebaar, oogopslag en non-verbale audio-indicatoren zoals spreektijd, toonhoogte, beurt nemen). De epistemische ruimte omvat de verbale audio-indicatoren (zoals de inhoud van het gesprek) en inhoudslogboekgegevens.

De belangrijkste doelstellingen van dit proefschrift zijn: 1) Het definieren en begrijpen van de co-located collaboration (CC) kwaliteit en de analyses; 2) Het ontwikkelen van een technisch prototype en het testen met veldproeven om te komen tot geautomatiseerde collaboration analytics met behulp van de eerder genoemde definities; 3) Het visualiseren van de CC analytics en te komen tot kwantificering van CC kwaliteit.

Om aan deze doelstellingen te voldoen, hebben we het proefschrift onderverdeeld in drie delen en vier hoofdstukken. In **Deel I**, bestaande uit 2 hoofdstukken (d.w.z. Hoofdstuk 1 en 2), beschrijven we de definitie voor CC-kwaliteit en hoe kwaliteit wordt gecontextualiseerd in bepaalde scenario's van samenwerking. We koppelen kwaliteit ook aan geschikte analyse- en feedbackmechanismen om leerlingen te ondersteunen bij het samenwerken. In hoofdstuk 1 doen we een verkennende state-of-the-art review om te begrijpen hoe indicatoren helpen om de kwaliteit van samenwerking te detecteren. Vervolgens hebben we de onderzoeken naar

CC-feedback en -analyses bekeken om te begrijpen hoe deze indicatoren de samenwerking hebben vergemakkelijkt. Tijdens een samenwerkingsvergadering werd bijvoorbeeld de totale spreektijd gebruikt als een indicator van de samenwerkingskwaliteit om realtime feedback te geven ter ondersteuning van de samenwerking. De realtime reflectieve feedback werd getoond door het vereiste aantal LED-lampjes (dat wil zeggen, evenredig aan de totale spreektijd) voor dat groepslid op een slimme tafel te laten branden. Dit hielp om een evenwicht te creëren tussen de deelnemers die meer verbale uiting gaven (d.w.z. wie meer sprak) en de deelnemers die minder verbale uiting gaven (d.w.z. wie minder sprak); waardoor de kwaliteit van de samenwerking verbeterde. Daarom hebben we op basis van de beoordeling van deze indicator, de feedbackvoorbeelden uit verschillende onderzoeken, en een ontworpen hybride opstelling (met de combinatie van menselijke waarnemers en sensoren zoals een microfoon) om realtime feedback te testen met behulp van een klein veldonderzoek. We testten deze opzet tijdens PhD-bijeenkomsten en volgden twee indicatoren van samenwerkingskwaliteit (d.w.z. totale spreektijd en beurt nemen). Vervolgens toonden we deze indicatoren als reflectieve feedback tijdens de vergadering op een groot openbaar gedeeld display als temporele grafieken. Het doel van deze feedback-opstelling was om er gevoel voor te krijgen in plaats van de effectiviteit van de feedback op de samenwerkingskwaliteit te testen.

Onze bevindingen geven aan dat de indicatoren van CC-kwaliteit kunnen worden gegroepeerd in twee categorieën (d.w.z. sociaal en epistemisch). De groep van epistemische indicatoren kan niet worden gedetecteerd en geanalyseerd door senoren alleen en heeft de hulp van mensen nodig vanwege de semantische aard van het begrijpen ervan. Bovendien is het ontwerp van CC-analyses en feedback afhankelijk van de stakeholders en helpt het om het ontwerp te versnellen door een mens in de lus te hebben.

Om dit verder te begrijpen, doen we een diepgaand literatuuronderzoek naar de indicatoren van samenwerkingskwaliteit in hoofdstuk 2. Dit komt omdat de indicatoren van samenwerking variëren afhankelijk van de scenario's en de context van samenwerking. Hier definiëren we de kwaliteit van samenwerking met een event-proces framework dat bestaat uit de indicatoren en indexen. De indicatoren zijn gebeurtenissen op laag niveau die zijn verkregen na verwerking en aggregatie van de sensoren. De indexen (d.w.z. processen op hoog niveau) fungeren als meetbare markers die helpen om de kwaliteit van samenwerking te detecteren. Ze bestaan uit een of meer indicatoren. In samenwerkingsvergaderingen (d.w.z. het scenario van CC) bijvoorbeeld, meet de gelijkheid (d.w.z. de index) van de totale spreektijd (d.w.z. de indicator) de kwaliteit van de samenwerking. Als alle groepsleden dezelfde totale spreektijd hebben en niemand het gesprek domineert, is er een grotere gelijkheid van de totale spreektijd voor de groep en een betere kwaliteit van samenwerking. Deze indicatoren variëren in verschillende scenario's vanwege de verschillende doelen en parameters (d.w.z. primaire aspecten zoals teamsamenstelling, gedrag van teamleden en gedrag tijdens samenwerking) van CC. Indicatoren van CC-kwaliteit voor collaboratief programmeren kunnen bijvoorbeeld volledig verschillen van indicatoren van collaboratief brainstormen.

Met behulp van dit literatuuronderzoek definiëren we dus een conceptueel model dat de indicatoren, indexen en parameters omvat om de CC-kwaliteit te detecteren. In dit model brengen we de parameters in verschillende scenario's in kaart op de indicatoren en indexen om de weg vrij te maken voor het ontwerpen van een CC-kwaliteitsdetectie- en voorspellingssysteem. Onze bevindingen geven ook aan dat de operationalisering van CC-kwaliteit lijdt onder de coderingscomplexiteit en het gebruik van ondoorzichtige algoritmen voor machine learning. Er is een enorme kloof tussen de theoretische indexen (d.w.z. meetbare markers) van CC-kwaliteit en de praktisch gedetecteerde indexen. Dit moet worden overbrugd.

Gebruikmakend van deze definitie van CC-kwaliteit richten we ons op audiogebaseerde indicatoren van samenwerking in **Deel II**, Hoofdstuk 3. Audio is de meest gebruikte modaliteit zoals gevonden in het literatuuronderzoek en zeer gemakkelijk vast te leggen met microfoons. De meeste eerdere studies richtten zich op "hoe" groepsleden spreken en niet op "wat" ze spreken. De focus lag dus op de sociale ruimte (bestaande uit non-verbale audio-indicatoren zoals totale spreektijd, verandering in toonhoogte) en niet op de epistemische ruimte (bestaande uit de inhoud van de gesprekken) van de audiomodaliteit. Zeer weinig werken richtten zich op de epistemische ruimte semi-automatisch in een lab setting met behulp van handmatige codering. Deze benaderingen waren gebaseerd op vooraf gedefinieerde voorwaarden, gaven een abstract overzicht van de discussieonderwerpen en waren arbeidsintensief om te implementeren. Daarom hebben we een prototype ontwikkeld om dit te ondervangen en de rijkdom van de epistemische ruimte te analyseren in een authentieke, reële wereldsetting met behulp van veldproeven. Hiervoor namen we de audiogesprekken op tijdens een CC-taak waarbij universiteitsmedewerkers een bordspel speelden met vooraf toegewezen rollen om bewustzijn te creëren over de verbinding tussen learning analytics en learning design. We transcribeerden deze audio-opnames (d.w.z. zetten ze om van spraak naar tekst), verwerkten ze en visualiseerden ze vervolgens (met behulp van netwerkgrafieken om de onderlinge verbondenheid van de gesproken tekst te begrijpen). Hiervoor hebben we ook een rol-gebaseerde profilering gedaan om een holistisch overzicht van de gesprekken te krijgen op een automatische manier. We hebben dit prototype getest gedurende één CC sessie met als doel een stap te zetten in de richting van automatische analyse van de samenwerking.

Onze bevindingen geven aan dat de inhoud van de gesprekken rijke informatie kan opleveren over de CC-kwaliteit. Met behulp van interactieve netwerkgrafieken kunnen rijke inzichten worden verkregen over de onderling verbonden gesprekspatronen tussen groepsleden. Het ontwerpen en prototypen van een CC-analyse-opstelling is echter moeilijk om volledig te automatiseren. We hebben hulp van mensen nodig om het gegevenscorpus op te schonen, vooral wanneer de naam in de gesprekken wordt uitgesproken.

Aan de hand van het ontwikkelde prototype hebben we de kwaliteit van de samenwerking gekwantificeerd in **Deel III**, Hoofdstuk 4, met behulp van veldproeven in 14 verschillende CC-sessies die in een universitaire setting met behulp van bordspellen werden gespeeld. We hebben een holistische analyse gemaakt van de sociale ruimte (d.w.z. de totale spreektijd en het nemen van beurten) en de epistemische ruimte (d.w.z. de inhoud van de gesprekken), waarbij we ook rekening hebben gehouden met de rolgebaseerde bijdragen en interacties, en hebben deze vervolgens gevisualiseerd. We definiëren de kwaliteit van de samenwerking rekening houdend met de convergentie van de discussie (i.e., gedeelde epistemische ruimte kennis zoals geanalyseerd uit de inhoud van de gesprekken) tussen de groepsleden met verschillende rollen. Ten slotte hebben we zowel de sociale als de epistemische ruimte gevisualiseerd met behulp van een dashboard; vervolgens bespreken we de stakeholders die zo’n dashboard kunnen gebruiken en welk toekomstig onderzoek kan worden gedaan op dat dashboard.

We vinden dat de meeste focus op CC-analyse nu op de sociale ruimte ligt en dat deze moet verschuiven naar de epistemische ruimte. Maatregelen voor CC-kwaliteit, zoals convergentie (d.w.z. een toename van gedeelde kennis tussen groepsleden) gemeten vanuit de epistemische ruimte, kunnen nuttig zijn om de CC-kwaliteit te kwantificeren. Zowel de sociale als de epistemische ruimte kan nuttig zijn om een holistisch beeld te geven van de CC-kwaliteit wanneer deze wordt gevisualiseerd op een enkel dashboard.

In de **Discussie** worden op basis van de bovenstaande bevindingen een aantal observaties en aanbevelingen geformuleerd en verder uitgewerkt. We hebben deze gegroepeerd naar de fase van het onderzoek naar CSCL en CC analytics: het definiëren van de CC kwaliteit en analytics in de vroege fase, het prototypen van de automatische CC analytics set up in de mid fase en het visualiseren van de CC analytics om naar de kwantificering van CC kwaliteit te gaan in de late fase.

Summary in Hindi

सहयोग 21वीं सदी के 4 महत्वपूर्ण कौशलों में से एक है (किवुंजा, 2015)। कब दो या दो से अधिक लोग एक समान लक्ष्य की दिशा में कार्य करते हैं तो सहयोग घटित होने को कहा जाता है (डिलनबर्ग, 1999)। सहयोग या तो ऑनलाइन सेटिंग में हो सकता है या एक सह-स्थित (या आमने-सामने) सेटिंग में। ऑनलाइन सहयोग का मापन शिक्षार्थी के माप, संग्रह और विश्लेषण के कारण प्रक्रियाएं संभव हैं लर्निंग एनालिटिक्स का उपयोग कर डेटा (सीमेंस, 2011; ग्रीलर और ड्रेक्स्लर, 2012)। साथ हाल ही में सेंसर का सर्वव्यापी उपयोग, अन्यथा सीखने की विश्लेषिकी की एक नई शाखा मल्टीमॉडल लर्निंग एनालिटिक्स (एमएमएलए) के रूप में जाना जाता है, प्रमुखता से बढ़ गया है (डी मित्रि) एट अल।, 2018ए; मार्टिनेज-मालडोनाडो एट अल।, 2017ए)। इसके अलावा, सेंसर तकनीक है पिछले एक दशक में अधिक स्केलेबल (रेली एट अल।, 2018), सस्ती और विश्वसनीय बनें (स्टार एट अल।, 2018)। वृद्धि के कारण सीसी के विश्लेषण पर ध्यान केंद्रित किया गया है एमएमएलए की। किया गया है।

इस थीसिस में हमने सह-स्थित (या आमने-सामने) सहयोग पर ध्यान केंद्रित किया। सह-स्थित सहयोग समूह के सदस्यों के भौतिक, सामाजिक और ज्ञान-मीमांसा स्थान के प्रतिच्छेदन में होता है। विभिन्न स्थानों (यानी, महामारी और सामाजिक स्थानों) में सहयोग के विभिन्न संकेतकों का उपयोग करके सहयोग गुणवत्ता का पता लगाया जा सकता है (प्रहराज एट अल।, 2018बी)। सामाजिक स्थान में गैर-मौखिक संकेतक (मुद्रा, हावभाव, नज़र और गैर-मौखिक ऑडियो संकेतक जैसे बोलने का समय, पिच, टर्न-टेकिंग) शामिल हैं। एपिस्टेमिक स्पेस में मौखिक ऑडियो संकेतक (जैसे बातचीत की सामग्री) और सामग्री लॉग डेटा शामिल हैं।

इस थीसिस के मुख्य उद्देश्य हैं: 1) सह-स्थित सहयोग (सीसी) गुणवत्ता और विश्लेषण को परिभाषित करना और समझना; 2) एक तकनीकी प्रोटोटाइप विकसित करना और उपरोक्त परिभाषाओं का उपयोग करके स्वचालित सहयोग विश्लेषण की ओर बढ़ने के लिए फील्ड परीक्षणों के साथ इसका परीक्षण करना; 3) सीसी एनालिटिक्स की कल्पना करना और सीसी गुणवत्ता को मापने की दिशा में आगे बढ़ना।

इन उद्देश्यों की पूर्ति के लिए हमने थीसिस को तीन भागों और चार अध्यायों में विभाजित किया है। 2 अध्यायों (अर्थात, अध्याय 1 और 2) से मिलकर, हम सीसी गुणवत्ता की परिभाषा का वर्णन करते हैं, और सहयोग के कुछ परिदृश्यों में गुणवत्ता को कैसे संदर्भित किया जाता है। हम सहयोग में शिक्षार्थियों का समर्थन करने के लिए गुणवत्ता को उपयुक्त विश्लेषण और प्रतिक्रिया तंत्र से भी जोड़ते हैं। अध्याय 1 में, हम यह समझने के लिए एक खोजपूर्ण अत्याधुनिक समीक्षा करते हैं कि कैसे संकेतक सहयोग की गुणवत्ता का पता लगाने में मदद करते हैं। फिर हमने सीसी फीडबैक और एनालिटिक्स पर अध्ययनों पर ध्यान दिया कि कैसे इन संकेतकों ने सहयोग को सुविधाजनक बनाने में मदद की। उदाहरण के लिए, सहयोगात्मक बैठक के दौरान, सहयोग का समर्थन करने के लिए रीयल-टाइम फ़ीडबैक देने के लिए सहयोग गुणवत्ता के संकेतक के रूप में कुल बोलने के समय का उपयोग किया गया था। एक स्मार्ट टेबल पर उस समूह के सदस्य के सामने आवश्यक संख्या में एलईडी रोशनी (यानी, कुल बोलने के समय के आनुपातिक) को चमकते हुए वास्तविक समय परावर्तक प्रतिक्रिया दिखाई गई थी। इससे उन प्रतिभागियों के बीच संतुलन बनाने में मदद मिली जिन्होंने अधिक मौखिक रूप से व्यक्त किया (यानी, जो अधिक बोलते थे) और जिन्होंने कम मौखिक रूप से व्यक्त किया (यानी, जो कम बोलते थे); जिससे सहयोग की गुणवत्ता में सुधार हो। इसलिए, इन संकेतकों की समीक्षा के आधार पर, कई अध्ययनों से फीडबैक उदाहरण, हमने एक छोटे से फील्ड अध्ययन की सहायता से रीयल-टाइम फीडबैक का परीक्षण करने के लिए एक हाइब्रिड सेट अप (मानव पर्यवेक्षकों और माइक्रोफ़ोन जैसे सेंसर के संयोजन के साथ) तैयार किया। हमने पीएचडी मीटिंग्स के दौरान इस सेट अप का परीक्षण किया और सहयोग गुणवत्ता के दो संकेतकों (यानी, कुल बोलने का समय और टर्न टेकिंग) को ट्रैक किया। फिर हमने इन संकेतकों को अस्थायी रेखांकन के रूप में एक बड़े सार्वजनिक साझा प्रदर्शन पर बैठक के दौरान

एक चिंतनशील प्रतिक्रिया के रूप में दिखाया। इस फीडबैक सेट अप का उद्देश्य सहयोग गुणवत्ता पर फीडबैक की प्रभावकारिता का परीक्षण करने के बजाय इसका अनुभव प्राप्त करना था।

हमारे निष्कर्ष बताते हैं कि सीसी गुणवत्ता के संकेतकों को दो श्रेणियों (अर्थात, सामाजिक और ज्ञान-मीमांसा) में बांटा जा सकता है। ज्ञान-मीमांसा संकेतकों के समूह का अकेले वरिष्ठों द्वारा पता नहीं लगाया जा सकता है और उनका विश्लेषण नहीं किया जा सकता है और उन्हें समझने की इन मानसिक प्रकृति के कारण मनुष्यों की सहायता की आवश्यकता होती है। इसके अलावा, सीसी एनालिटिक्स और फीडबैक डिजाइन हितधारकों पर निर्भर है और लूप में मानव होने से डिजाइन को गति देने में मदद मिलती है।

इसे और समझने के लिए, हम अध्याय 2 में सहयोग गुणवत्ता के संकेतकों की गहन साहित्य समीक्षा करते हैं। ऐसा इसलिए है क्योंकि सहयोग के संकेतक परिदृश्यों और सहयोग के संदर्भ के आधार पर भिन्न होते हैं। यहां, हम संकेतकों और अनुक्रमितों से बने एक घटना-प्रक्रिया ढांचे के साथ सहयोग की गुणवत्ता को परिभाषित करते हैं। संकेतक सेंसर से प्रसंस्करण और एकत्रीकरण के बाद प्राप्त निम्न-स्तरीय घटनाएं हैं। इंडेक्स (यानी, उच्च स्तरीय प्रक्रियाएं) सहयोग की गुणवत्ता का पता लगाने में मदद करने वाले विषय-मापने योग्य मार्कर के रूप में कार्य करते हैं। एक या एक से अधिक संकेतकों से बने होते हैं। उदाहरण के लिए, सहयोगी बैठकों (यानी, सीसी का परिदृश्य) में, कुल बोलने के समय (यानी, संकेतक) की समानता (यानी, सूचकांक) सहयोग की गुणवत्ता को मापती है। यदि समूह के सभी सदस्यों का कुल बोलने का समय समान है और कोई भी बातचीत पर हावी नहीं है, तो समूह के लिए बोलने के कुल समय की उच्च समानता और सहयोग की बेहतर गुणवत्ता है। ये संकेतक अलग-अलग परिदृश्यों में अलग-अलग लक्ष्यों और मापदंडों के कारण भिन्न होते हैं (यानी, प्राथमिक पहलू जैसे कि सीसी की टीम संरचना, टीम के सदस्यों का व्यवहार और सहयोग के दौरान व्यवहार)। उदाहरण के लिए, सहयोगी प्रोग्रामिंग के लिए सीसी गुणवत्ता के संकेतक सहयोगी विचार-मंथन के संकेतकों से पूरी तरह भिन्न हो सकते हैं।

इस प्रकार, इस साहित्य समीक्षा की सहायता से, हम एक वैचारिक मॉडल को परिभाषित करते हैं जिसमें सीसी गुणवत्ता का पता लगाने के लिए संकेतक, अनुक्रमणिका और पैरामीटर शामिल हैं। इस मॉडल में, हम संकेतकों पर विभिन्न परिदृश्यों में मापदंडों को मैप करते हैं और सीसी गुणवत्ता का पता लगाने और भविष्यवाणी प्रणाली के डिजाइन का मार्ग प्रशस्त करने के लिए अनुक्रमित करते हैं। हमारे निष्कर्ष यह भी संकेत करते हैं कि सीसी गुणवत्ता का संचालन कोडिंग जटिलता और अपारदर्शी मशीन लर्निंग एल्गोरिदम के उपयोग से ग्रस्त है। सीसी गुणवत्ता के सैद्धांतिक सूचकांक (अर्थात, मापन योग्य मार्कर) और व्यावहारिक रूप से खोजे गए सूचकांकों के बीच एक बड़ा अंतर है। इसे पाटने की जरूरत है।

सीसी गुणवत्ता की इस परिभाषा का उपयोग करते हुए, हम भाग 2, अध्याय 3 में सहयोग के ऑडियो-आधारित संकेतकों पर ध्यान केंद्रित करते हैं। ऑडियो प्रमुख रूप से उपयोग किया जाने वाला साधन है जैसा कि साहित्य समीक्षा में पाया गया है और माइक्रोफ़ोन के साथ कैप्चर करना बहुत आसान है। अधिकांश पूर्व अध्ययन "कैसे" समूह के सदस्य बोलते हैं, न कि "क्या" बोलते हैं। इस प्रकार, ध्यान सामाजिक स्थान (गैर-मौखिक ऑडियो संकेतक जैसे कुल बोलने का समय, पिच में परिवर्तन) पर केंद्रित था, न कि ऑडियो मोडेलिटी के महामारी स्थान (बातचीत की सामग्री को मिलाकर) पर। मैनुअल कोडिंग का उपयोग करते हुए प्रयोगशाला सेटिंग में अर्ध-स्वचालित रूप से महामारी अंतरिक्ष पर केंद्रित बहुत कम कार्य। ये दृष्टिकोण पूर्वनिर्धारित स्थितियों पर आधारित थे, चर्चा के विषयों का सार अवलोकन और कार्यान्वयन के लिए श्रमसाध्य थे। इसलिए, हमने इस पर काबू पाने के लिए एक प्रोटोटाइप विकसित किया और फील्ड परीक्षणों की मदद से एक प्रामाणिक वास्तविक दुनिया सेटिंग में महामारी अंतरिक्ष की समृद्धि का विश्लेषण किया। इसके लिए, हमने एक सीसी कार्य के दौरान ऑडियो वार्तालापों को रिकॉर्ड किया, जहां विश्वविद्यालय के कर्मचारियों ने सीखने के विश्लेषण और सीखने के डिजाइन के बीच संबंध के बारे में जागरूकता पैदा करने के लिए पूर्व-निर्धारित भूमिकाओं के साथ एक बोर्ड गेम खेला। हमने इन ऑडियो रिकॉर्डिंग्स को ट्रांसक्राइब किया (यानी, भाषण से टेक्स्ट में कनवर्ट करें), उन्हें प्रोसेस किया और फिर उन्हें विज़ुअलाइज़ किया (बोलने वाले टेक्स्ट की इंटरकनेक्टेड प्रकृति को समझने के लिए नेटवर्क ग्राफ़ का उपयोग करके)। इसके लिए, हमने स्वचालित तरीके से बातचीत का समग्र अवलोकन प्राप्त करने के लिए एक भूमिका-आधारित प्रोफाइलिंग भी की। हमने स्वचालित सहयोग विश्लेषण की दिशा में एक कदम बढ़ाने के उद्देश्य से एक सीसी सत्र के लिए इस प्रोटोटाइप का परीक्षण किया। किया।

हमारे निष्कर्ष बताते हैं कि बातचीत की सामग्री सीसी गुणवत्ता के बारे में समृद्ध जानकारी प्रदान कर सकती है। इंटरएक्टिव नेटवर्क ग्राफ़ की मदद से, समूह के सदस्यों के बीच परस्पर बातचीत पैटर्न के बारे में समृद्ध अंतर्दृष्टि प्राप्त की जा सकती है। हालांकि, सीसी एनालिटिक्स सेट अप को डिजाइन और प्रोटोटाइप करना पूर्ण स्वचालन के लिए मुश्किल है। हमें डेटा कॉर्पस को साफ करने के लिए मनुष्यों की मदद की ज़रूरत है, खासकर जब बातचीत में नाम का उच्चारण किया जाता है।

विकसित प्रोटोटाइप का उपयोग करते हुए, हम बोर्ड गेम का उपयोग करके विश्वविद्यालय की सेटिंग में खेले गए 14 अलग-अलग सीसी सत्रों में फील्ड ट्रायल की मदद से भाग 3, अध्याय 4 में सहयोग की गुणवत्ता को मापने की दिशा में आगे बढ़े। हमने भूमिका आधारित योगदानों और अंतःक्रियाओं पर विचार करते हुए सामाजिक (यानी, कुल बोलने का समय और बारी लेने) और एपिस्टेमिक (यानी, बातचीत की सामग्री) स्थान का समग्र विश्लेषण किया और फिर इसकी कल्पना की। हम अलग-अलग भूमिकाओं वाले समूह के सदस्यों के बीच चर्चा के अभिसरण (यानी, साझा महामारी अंतरिक्ष ज्ञान का विश्लेषण बातचीत की सामग्री से विश्लेषण) को ध्यान में रखते हुए सहयोग की गुणवत्ता को परिभाषित करते हैं। अंत में, हमने एडशबोर्ड का उपयोग करके सामाजिक और ज्ञान-मीमांसा दोनों स्थान की कल्पना की; फिर उन हितधारकों पर चर्चा करें जो इस तरह के डैशबोर्ड का उपयोग कर सकते हैं और उस डैशबोर्ड पर भविष्य में क्या शोध किया जा सकता है |

हम पाते हैं कि सीसी विश्लेषण पर अधिकांश ध्यान अब सामाजिक स्थान पर है और इसे ज्ञान-मीमांसा स्थान पर स्थानांतरित करने की आवश्यकता है। सीसी गुणवत्ता के उपाय जैसे कि अभिसरण (यानी, समूह के सदस्यों के बीच साझा ज्ञान में वृद्धि) को एपिस्टेमिक स्पेस से मापा जाता है, सीसी गुणवत्ता को मापने के लिए उपयोगी हो सकता है। सीसी गुणवत्ता का एक समग्र दृष्टिकोण देने के लिए सामाजिक और ज्ञान-मीमांसा दोनों स्थान उपयोगी हो सकते हैं जब एक सिंगल डैशबोर्ड पर कल्पना की जाती है।

चर्चा में, ऊपर सूचीबद्ध निष्कर्षों के आधार पर, हम टिप्पणियों और सिफारिशों के मिश्रण को प्राप्त करते हैं और विस्तृत करते हैं। हमने इन्हें सीएससीएल और सीसी एनालिटिक्स पर शोध के चरण के अनुसार समूहीकृत किया: प्रारंभिक चरण में सीसी गुणवत्ता और विश्लेषण को परिभाषित करना, मध्य चरण में स्थापित स्वचालित सीसी एनालिटिक्स का प्रोटोटाइप बनाना और सीसी एनालिटिक्स की कल्पना करना ताकि सीसी गुणवत्ता की मात्रा का ठहराव हो सके।

Summary in Odia

ଏକବିଂଶ ଶତାବ୍ଦୀର 4 ଟି ଗୁରୁତ୍ୱପୂର୍ଣ୍ଣ କୌଶଳ ମଧ୍ୟରୁ ସହଯୋଗ ହେଉଛି (Kivunja, 2015) ଗୋଟିଏ । ଏକ ଅନଲାଇନ୍ ସେଟିଂରେ କିମ୍ବା ଏକ ସହ-ଅବସ୍ଥାନ (କିମ୍ବା ମୁହାଁମୁହିଁ) ସେଟିଂରେ ସହଯୋଗ ହୋଇପାରେ । ଏହି ବିଷୟବସ୍ତୁରେ ଆମେ ସହ-ଅବସ୍ଥାନ (କିମ୍ବା ମୁହାଁମୁହିଁ) ସହଯୋଗ ଉପରେ ଧ୍ୟାନ ଦେଇଥିଲୁ । ସମବାୟ ସମିତି ଗୋଷ୍ଠୀ ସଦସ୍ୟଙ୍କ ଶାରୀରିକ, ସାମାଜିକ ଏବଂ ଜ୍ଞାନ ସ୍ଥାନର ଛକଠାରେ ହୋଇଥାଏ । ବିଭିନ୍ନ ସୂଚକ ବ୍ୟବହାର କରି ସହଯୋଗ ଗୁଣବତ୍ତା ଚିହ୍ନଟ କରାଯାଇପାରେ । ବିଭିନ୍ନ ସ୍ପେସରେ ସହଯୋଗ (ଯଥା, ଜ୍ଞାନ ଏବଂ ସାମାଜିକ ସ୍ପେସ୍) (Praharaj et al., 2018b) ହେଇପାରେ । ସାମାଜିକ ଜାଗାରେ ଅଣଭର୍ବାଲ୍ ସୂଚକ (ସ୍ଥିତି, ଅଙ୍ଗଭଙ୍ଗୀ, ଆଖି ନଜର ଏବଂ କଥା ନହେବା ସମୟ, ପିଚ୍, ଟର୍ନ ନେବା) ଭଳି ଅଣଭର୍ବାଲ୍ ଅଡିଓ ସୂଚକ ଥାଏ । ଜ୍ଞାନ ସ୍ପେସ୍ ଭର୍ବାଲ୍ ଅଡିଓ ସୂଚକ (ଯେପରିକି ବାର୍ତ୍ତାଳାପର ବିଷୟବସ୍ତୁ) ଏବଂ ବିଷୟବସ୍ତୁ ଲଗ୍ ତଥ୍ୟକୁ ନେଇ ଗଠିତ ।

ଏହି ବିଷୟବସ୍ତୁର ମୂଳ ଉଦ୍ଦେଶ୍ୟଗୁଡ଼ିକ ହେଉଛି: 1) ସହ-ଅବସ୍ଥାନ ସହଯୋଗ (CC) ଗୁଣ ଏବଂ ଆନାଲିଟିକ୍ସକୁ ବ୍ୟାଖ୍ୟା ଏବଂ ବୁଝିବା; 2) ଏକ ଯାନ୍ତ୍ରିକ ପ୍ରୋଟୋଟାଇପ୍ ବିକଶିତ କରିବା ଏବଂ ଏହାକୁ ଉପରୋକ୍ତ ସଂଜ୍ଞା ବ୍ୟବହାର କରି ସ୍ୱୟଂଚାଳିତ ସହଯୋଗ ଆନାଲିଟିକ୍ସ ଆଡକୁ ଯିବା ପାଇଁ କ୍ଷେତ୍ର ପରୀକ୍ଷଣ ସହିତ ଏହାକୁ ପରୀକ୍ଷା କରିବା । 3) CC ଆନାଲିଟିକ୍ସକୁ ଭିଜୁଆଲ୍ କରିବା ଏବଂ CC ଗୁଣବତ୍ତା ମାପିବା ଆଡକୁ ଯିବା ।

ଆମେ ପାଇଲୁ ଯେ CC ଆନାଲିଟିକ୍ସ ଉପରେ ଅଧିକାଂଶ ଧ୍ୟାନ ବର୍ତ୍ତମାନ ସାମାଜିକ ସ୍ଥାନ ଉପରେ ଅଛି ଏବଂ ଏହା ଜ୍ଞାନ ସ୍ଥାନକୁ ସ୍ଥାନାନ୍ତରିତ ହେବା ଆବଶ୍ୟକ । CC ଗୁଣବତ୍ତା ପରିମାପ (ଯଥା, ଗୋଷ୍ଠୀ ସଦସ୍ୟଙ୍କ ମଧ୍ୟରେ ଅଂଶୀଦାରିତ ଜ୍ଞାନରେ ବୃଦ୍ଧି) CC ଗୁଣବତ୍ତା ପରିମାଣ ପାଇଁ ଉପଯୋଗୀ ହୋଇପାରେ । ଉଭୟ ସାମାଜିକ ଏବଂ ଜ୍ଞାନ ସ୍ପେସ କୁ ଉପଯୋଗ କରି CC ଗୁଣର ଏକ ସାମଗ୍ରିକ ଦୃଶ୍ୟ ଦେବା ପାଇଁ ଉପଯୋଗୀ |

ଆଲୋଚନାରେ, ଉପରୋକ୍ତ ତାଲିକାଭୁକ୍ତ ଫଳାଫଳ ଉପରେ ଆଧାର କରି, ଆମେ ପର୍ଯ୍ୟବେକ୍ଷଣ ଏବଂ ସୁପାରିଶର ମିଶ୍ରଣ ଉପରେ ବର୍ଣ୍ଣନା କରୁ | CSCL ଏବଂ CC ଆନାଲିଟିକ୍ସ ଉପରେ ଅନୁସନ୍ଧାନର ପର୍ଯ୍ୟାୟ ଅନୁଯାୟୀ ଆମେ ଏହାକୁ ଗୋଷ୍ଠୀ କରିଛୁ: ପ୍ରାରମ୍ଭିକ ପର୍ଯ୍ୟାୟରେ CC ଗୁଣବତ୍ତା ଏବଂ ଆନାଲିଟିକ୍ସକୁ ବ୍ୟାଖ୍ୟା କରିବା, ମଧ୍ୟଭାଗରେ ସ୍ଥାପିତ ସ୍ୱୟଂଚାଳିତ CC ଆନାଲିଟିକ୍ସକୁ ପ୍ରୋଟୋଟାଇପ୍ କରିବା ଏବଂ ବିଳମ୍ବିତ ପର୍ଯ୍ୟାୟରେ CC ଗୁଣବତ୍ତା ପରିମାଣକୁ ମାପିବା ପାଇଁ CC ଆନାଲିଟିକ୍ସକୁ ଭିଜୁଆଲ୍ କରିବା |

Acknowledgements

The path to starting this PhD was during my Master's in Computer Science in TU Delft where I was drawn to the Learning Analytics and Educational Technology community. There, I did experiments in a Bachelor's classroom to collect and analyze data logs from every student and help the teacher to get some insights about student engagement. At that time towards the end of my Master's, I came across Hendrik and Marcus' work which were very interesting for me. Later when I applied for this PhD, everything happened in a flash and finally after more than 4 years I am finishing writing this thesis. When I started my PhD then came Maren who helped me a lot to refine my writing skills, conceptualizing by giving regular feedback. I really loved to work with all three of my supervisors just like my second family away from home. All of my supervisors helped me a lot to make this PhD journey enjoyable, interesting and exciting. I think the biggest positive of being surrounded by these group of people is to never loose the excitement of doing a PhD through thick and thin.

The trips that I had with Hendrik to Leuven, Aachen and Frankfurt helped me to know his other side and talk more freely about my problems. I was lucky to have him and Marcus both on my side. I started my PhD with Marcus (as my main promotor) who was always very approachable, filled with ideas and a fun person to work with. Then in the second half of my PhD Hendrik took the reigns from Marcus who made me focus on the big picture instead of being distracted by the problems that arose. He is very caring and constantly motivated me. Maren was like my regular and more importantly emergency research contact. I asked her different types of questions and she was always there to help. Moreover, I would like to thank Sebastian, George, Roland and Daniel.

Furthermore, my paranymphs, Nardie and Marcel have been there throughout. Especially after Covid struck, my PhD was delayed, but Marcel was there to help me with the data collection in Zuyd University so that I can progress with my PhD. Nardie is like someone with whom I bonded very well from the beginning of joining the Open Universiteit. He always gives constant dose of positivity and his energy is infectious. I would like to thank many other colleagues (whom I can remember) who helped me a lot in the beginning and throughout the PhD. They are Ioana, Daniele, Bibeg, Liqin, Khaleel, Alessandra, Julia, Martine, Kevin, Angel, Jan, Katya, Esther, Howard, Slavi, Haoyu and Manuel. The weekly group meetings with Ioana, Daniele, Marcel and my supervisors were like fireball which propelled me through the week. The ping pong sessions after lunch with Liqin, Bibeg and Khaleel were the best that I can remember which sometimes became problem solving sessions too! I would be always grateful for the lunch talk and small talk with the past and current TELI colleagues and the current OLI colleagues.

One of my friend who helped me in designing the backbone of my technical set up was Manoj. Some of my other friends like Varun, Vivek, Abhijeet, Nirmal, Vibhor, Bhavya and Arnab gave immense support. One of the memorable trips with Vivek, Abhijeet and Arnab once helped me to write an article and also get it accepted in a journal. I would also like to thank all the anonymous reviewers who helped me to improve my articles. In addition to it, Roberto, Christian, Marco, Mikhail, Roland, Karel and Halszka helped a lot passively with ideas during my PhD. The JTELSS summer school was a rich session with a mix of fun interactions and idea exchange. This really helped me initially to meet fellow PhDs in my domain and talk with them. Besides, my YouTube family of subscribers helped a lot in encouraging me to continue my PhD with passion.

Last but not the least, I would like to thank my parents who always stood behind me and offered emotional and strategic help even from India. Almost everyday evening we talk with each other and I share all my problems if I have any to get better advice. It is because of them that I could come abroad and do whatever I can till today.

SIKS Dissertation Series

The complete list of dissertations carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems from 1998 on can be found at `http://www.siks.nl/dissertations.php`.

2016

2016-01 Syed Saiden Abbas (RUN)
Recognition of Shapes by Humans and Machines.

2016-015 Steffen Michels (RUN)
Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments.

2016-017 Berend Weel (VU)
Towards Embodied Evolution of Robot Organisms.

2016-019 Julia Efremova (TUE)
Mining Social Structures from Genealogical Data.

2016-02 Michiel Meulendijk (UU)
Optimizing medication reviews through decision support: prescribing a better pill to swallow.

2016-03 Maya Sappelli (RUN)
Knowledge Work in Context: User Centered Knowledge Worker Support.

2016-04 Laurens Rietveld (VU)
Publishing and Consuming Linked Data.

2016-05 Evgeny Sherkhonov (UVA)
Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers.

2016-06 Michel Wilson (TUD)
Robust scheduling in an uncertain environment.

2016-07 Jeroen de Man (VU)
Measuring and modeling negative emotions for virtual training.

2016-08 Matje van de Camp (TiU)
A Link to the Past: Constructing Historical Social Networks from Unstructured Data.

2016-09 Archana Nottamkandath (VU)
Trusting Crowdsourced Information on Cultural Artefacts.

2016-10 George Karafotias (VU)
Parameter Control for Evolutionary Algorithms.

2016-11 Anne Schuth (UVA)
Search Engines that Learn from Their Users.

2016-12 Max Knobbout (UU)
Logics for Modelling and Verifying Normative Multi-Agent Systems.

2016-13 Nana Baah Gyan (VU)
The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach.

2016-14 Ravi Khadka(UU)
Revisiting Legacy Software System Modernization.

2016-16 Guangliang Li (UVA)
Socially Intelligent Autonomous Agents that Learn from Human Reward.

2016-18 Albert Merono Penuela (VU)
Refining Statistical Data on the Web.

2016-20 Daan Odijk (VU)
Context & Semantics in News & Web Search.

2016-21 Alejandro Moreno Celleri (UT)
From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground.

2016-22 Grace Lewis (VU)
Software Architecture Strategies for Cyber-Foraging Systems.

2016-23 Fei Cai (UVA)
Query Auto Completion in Information Retrieval.

2016-24 Brend Wanders (UT)
Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach.

2016-25 Y. Kiseleva (TUE)
Using Contextual Information to Understand Searching and Browsing Behavior.

2016-26 Dilhan J. Thilakarathne (VU)
In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains.

2016-27 Wen Li (TUD)
Understanding Geo-spatial Information on Social Media.

2016-28 Mingxin Zhang (TUD)
Large-scale agent-based social simulation: A study on epidemic prediction and control.

2016-29 Nicolas Honing (TUD)
Understanding Geo-spatial Information on Social Media.

2016-30 Ruud Mattheij (UVT)
The Eyes Have IT.

2016-31 Mohammad Khelghati (UT)
Deep web content monitoring.

2016-32 Eelco Vriezekolk (UVT)
Assessing Telecommunication Service Availability Risks for Crisis Organisations.

2016-33 Peter Bloem (UVA)
Single Sample Statistics, exercises in learning from just one example.

2016-34 Dennis Schunselaar (TUE)
Configurable Process Trees: Elicitation, Analysis, and Enactment.

2016-35 Zhaochun Ren (UVA)
Monitoring Social Media: Summarization, Classification and Recommendation.

2016-36 Daphne Karreman (UT)
Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies.

2016-37 Giovanni Sileno (UVA)
Aligning Law and Action - a conceptual and computational inquiry.

2016-38 Andrea Minuto (UT)
Materials that matter - Smart Materials meet Art & Interaction Design.

2016-39 Merijn Bruijnes
Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect.

2016-40 Christian Detweiler (TUD)
Accounting for Values in Design.

2016-41 Thomas King (TUD)
Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance.

2016-42 Spyros Martzoukos (UVA)
Combinatorial and compositional aspects of bilingual aligned corpora.

2016-43 Saskia Koldijk (RUN)
Context-Aware Support for Stress Self-Management: From Theory to Practice.

2016-44 Thibault Sellam (UVA)
Automatic assistants for database exploration.

2016-45 Bram van Laar (UT)
Experiencing Brain-Computer Interface Control.

2016-46 Jorge Gallego Perez (UT)
Robots to Make you Happy.

2016-47 Christina Weber (UL)
Real-time foresight - Preparedness for dynamic innovation networks.

2016-48 Tanja Buttler (TUD)
Collecting Lessons Learned.

2016-49 Gleb Polevoy (TUD)
Participation and Interaction in Projects: A Game-Theoretic Analysis.

2016-50 Yan Wang (UVT)
The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains.

2017

2017-01 Jan-Jaap Oerlemans (UL)
Investigating Cybercrime.

2017-02 Sjoerd Timmer (UU)
Designing and Understanding Forensic Bayesian Networks using Argumentation.

2017-03 Daniel Harold Telgen (UU)
Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines.

2017-04 Mrunal Gawade (CWI)
Multi-core parallelism in a column-store.

2017-05 Mahdieh Shadi (UVA)
Collaboration Behavior; Enhancement in Co-development.

2017-06 Damir Vandic (EUR)
Intelligent Information Systems for Web Product Search.

2017-07 Roel Bertens (UU)
Insight in Information: from Abstract to Anomaly.

2017-08 Rob Konijn (VU)
Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery.

2017-09 Dong Nguyen (UT)
Text as Social and Cultural Data: A Computational Perspective on Variation in Text.

2017-10 Robby van Delden (UT)
(Steering) Interactive Play Behavior.

2017-11 Florian Kunneman (RUN)
Modelling patterns of time and emotion in Twitter #anticipointment.

2017-12 Sander Leemans (TUE)
Robust Process Mining with Guarantees.

2017-13 Gijs Huisman (UT)
Social Touch Technology - Extending the reach of social touch through haptic technology.

2017-14 Shoshannah Tekofsky (UVT)
You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior.

2017-15 Peter Berck (RUN)
Memory-Based Text Correction.

2017-16 Aleksandr Chuklin (UVA)
Understanding and Modeling Users of Modern Search Engines.

2017-17 Daniel Dimov (UL)
Crowdsourced Online Dispute Resolution.

2017-18 Ridho Reinanda (UVA)
Entity Associations for Search.

2017-19 Jeroen Vuurens (TUD)
Proximity of Terms, Texts and Semantic Vectors in Information Retrieval.

2017-20 Mohammadbashir Sedighi (TUD)
Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility.

2017-21 Jeroen Linssen (UT)
Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds).

2017-22 Sara Magliacane (VU)
Logics for causal inference under uncertainty.

2017-23 David Graus (UVA)
Entities of Interest — Discovery in Digital Traces.

2017-24 Chang Wang (TUD)
Use of Affordances for Efficient Robot Learning.

2017-25 Veruska Zamborlini (VUA)
Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search.

2017-26 Merel Jung (UT)
Socially intelligent robots that understand and respond to human touch.

2017-27 Michiel Joosse (UT)
Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors.

2017-28 John Klein (VU)
Architecture Practices for Complex Contexts.

2017-29 Adel Alhuraibi (UVT)
From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT.

2017-30 Wilma Latuny (UVT)
The Power of Facial Expressions.

2017-31 Ben Ruijl (UL)
Advances in computational methods for QFT calculations.

2017-32 Thaer Samar (RUN)
Access to and Retrievability of Content in Web Archives.

2017-33 Brigit van Loggem (OU)
Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity.

2017-34 Maren Scheffel (OUN)
The Evaluation Framework for Learning Analytics.

2017-35 Martine de Vos (VU)
Interpreting natural science spreadsheets.

2017-36 Yuanhao Guo (UL)
Shape Analysis for Phenotype Characterisation from High-throughput Imaging.

2017-37 Alejandro Montes Garcia (TUE)
WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy.

2017-38 Abdullah Kayal (TUD)
Normative Social Applications.

2017-39 Sara Ahmadi (RUN)
Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR.

2017-40 Altaf Hussain Abro (VUA)
Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems".

2017-41 Adnan Manzoor (VUA)
Minding a Healthy Lifestyle:An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle.

2017-42 Elena Sokolova (RUN)
Causal discovery from mixed and missing data with applications on ADHD datasets.

2017-43 Maaike de Boer (RUN)
Semantic Mapping in Video Retrieval.

2017-44 Garm Lucassen (UU)
Understanding User Stories - Computational Linguistics in Agile Requirements Engineering.

2017-45 Bas Testerink (UU)
Decentralized Runtime Norm Enforcement.

2017-46 Jan Schneider (OU)
Sensor-based Learning Support.

2017-47 Yie Yang (TUD)
Crowd Knowledge Creation Acceleration.

2017-48 Angel Suarez (OU)
Collaborative inquiry-based learning.

2018

2018-01 Han van der Aa (VU)
Comparing and Aligning Process Representations.

2018-02 Felix Mannhardt (TUE)
Multi-perspective Process Mining.

2018-03 Steven Bosems (UT)
Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction.

2018-04 Jordan Janeiro (TUD)
Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks.

2018-05 Hugo Huurdeman (UVA)
Supporting the Complex Dynamics of the Information Seeking Process.

2018-06 Dan Ionita (UT)
Model-Driven Information Security Risk Assessment of Socio-Technical Systems.

2018-07 Jieting Luo (UU)
A formal account of opportunism in multi-agent systems.

2018-08 Rick Smetsers (RUN)
Advances in Model Learning for Software Systems.

2018-09 Xu Xie (TUD)
Data Assimilation in Discrete Event Simulations.

2018-10 Julienka Mollee (VUA)
Moving forward: supporting physical activity behavior change through intelligent technology.

2018-11 Mahdi Sargolzaei (UVA)
Enabling Framework for Service-oriented Collaborative Networks.

2018-12 Xixi Lu (TUE)
Using behavioral context in process mining.

2018-13 Seyed Amin Tabatabaei (VUA)
Computing a Sustainable Future: Exploring the added value of computational models for increasing the use of renewable energy in the residential sector.

2018-14 Bart Joosten (UVT)
Detecting Social Signals with Spatiotemporal Gabor Filters.

2018-15 Naser Davarzani (UM)
Biomarker discovery in heart failure.

2018-16 Jaebok Kim (UT)
Automatic recognition of engagement and emotion in a group of children.

2018-17 Jianpeng Zhang (TUE)
On Graph Sample Clustering.

2018-18 Henriette Nakad (UL)
De Notaris en Private Rechtspraak.

2018-19 Minh Duc Pham (VUA)
Emergent relational schemas for RDF.

2018-20 Manxia Liu (RUN)
Time and Bayesian Networks.

2018-21 Aad Slootmaker (OU)
EMERGO: a generic platform for authoring and playing scenario-based serious games.

2018-22 Eric Fernandes de Mello Araujo (VUA)
Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks.

2018-23 Kim Schouten (EUR)
Semantics-driven Aspect-Based Sentiment Analysis.

2018-24 Jered Vroon (UT)
Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots.

2018-25 Riste Gligorov (VUA)
Serious Games in Audio-Visual Collections.

2018-26 Roelof de Vries (UT)
Theory-Based And Tailor-Made: Motivational Messages for Behavior Change Technology.

2018-27 Maikel Leemans (TUE)
Hierarchical Process Mining for Scalable Software Analysis.

2018-28 Christian Willemse (UT)
Social Touch Technologies: How they feel and how they make you feel.

2018-29 Yu Gu (UVT)
Emotion Recognition from Mandarin Speech.

2018-30 Wouter Beek (VU)
The K in semantic web stands for knowledge: scaling semantics to the web.

2019

2019-01 Rob van Eijk (UL)
Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification.

2019-02 Emmanuelle Beauxis- Aussalet (CWI, UU)
Statistics and Visualizations for Assessing Class Size Uncertainty.

2019-03 Eduardo Gonzalez Lopez de Murillas (TUE)
Process Mining on Databases: Extracting Event Data from Real Life Data Sources.

2019-04 Ridho Rahmadi (RUN)
Finding stable causal structures from clinical data.

2019-05 Sebastiaan van Zelst (TUE)
Process Mining with Streaming Data.

2019-06 Chris Dijkshoorn (VU)
Nichesourcing for Improving Access to Linked Cultural Heritage Datasets.

2019-07 Soude Fazeli (TUD)
Recommender Systems in Social Learning Platforms.

2019-08 Frits de Nijs (TUD)
Resource-constrained Multi-agent Markov Decision Processes.

2019-09 Fahimeh Alizadeh Moghaddam (UVA)
Self-adaptation for energy efficiency in software systems.

2019-10 Qing Chuan Ye (EUR)
Multi-objective Optimization Methods for Allocation and Prediction.

2019-11 Yue Zhao (TUD)
Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs.

2019-12 Jacqueline Heinerman (VU)
Better Together.

2019-13 Guanliang Chen (TUD)
MOOC Analytics: Learner Modeling and Content Generation.

2019-14 Daniel Davis (TUD)
Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses.

2019-15 Erwin Walraven (TUD)
Planning under Uncertainty in Constrained and Partially Observable Environments.

2019-16 Guangming Li (TUE)
Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models.

2019-17 Ali Hurriyetoglu (RUN)
Extracting actionable information from microtexts.

2019-18 Gerard Wagenaar (UU)
Artefacts in Agile Team Communication.

2019-19 Vincent Koeman (TUD)
Tools for Developing Cognitive Agents.

2019-20 Chide Groenouwe (UU)
Fostering technically augmented human collective intelligence.

2019-21 Cong Liu (TUE)
Software Data Analytics: Architectural Model Discovery and Design Pattern Detection.

2019-22 Martin van den Berg (VU)
Improving IT Decisions with Enterprise Architecture.

2019-23 Qin Lin (TUD)
Intelligent Control Systems: Learning, Interpreting, Verification.

2019-24 Anca Dumitrache (VU)
Truth in Disagreement- Crowdsourcing Labeled Data for Natural Language Processing.

2019-25 Emiel van Miltenburg (VU)
Pragmatic factors in (automatic) image description.

2019-26 Prince Singh (UT)
An Integration Platform for Synchromodal Transport.

2019-27 Alessandra Antonaci (OUN)
The Gamification Design Process applied to (Massive) Open Online Courses.

2019-28 Esther Kuindersma (UL)
Cleared for take-off:Game-based learning to prepare airline pilots for critical situations.

2019-29 Daniel Formolo (VU)
Using virtual agents for simulation and training of social skills in safety-critical circumstances.

2019-30 Vahid Yazdanpanah (UT)
Multiagent Industrial Symbiosis Systems.

2019-31 Milan Jelisavcic (VU)
Alive and Kicking: Baby Steps in Robotics.

2019-32 Chiara Sironi (UM)
Monte-Carlo Tree Search for Artificial General Intelligence in Games.

2019-33 Anil Yaman (TUE)
Evolution of Biologically Inspired Learning in Artificial Neural Networks.

2019-34 Negar Ahmadi (TUE)
EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES.

2019-35 Lisa Facey-Shaw (OUN)
Gamification with digital badges in learning programming.

2019-36 Kevin Ackermans (OUN)
Designing Video-Enhanced Rubrics to Master Complex Skills.

2019-37 Jian Fang (TUD)
Database Acceleration on FPGAs.

2019-38 Akos Kadar (OUN)
Learning visually grounded and multilingual representations.

2020

2020-01 Armon Toubman (UL)
Calculated Moves: Generating Air Combat Behaviour.

2020-02 Marcos de Paula Bueno (UL)
Unraveling Temporal Processes using Probabilistic Graphical Models.

2020-03 Mostafa Deghani (UvA)
Learning with Imperfect Supervision for Language Understanding.

2020-04 Maarten van Gompel (RUN)
Context as Linguistic Bridges.

2020-05 Yulong Pei (TUE)
On local and global structure mining.

2020-06 Preethu Rose Anish (UT)
Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support.

2020-07 Wim van der Vegt (OUN)
Towards a software architecture for reusable game components.

2020-08 Ali Mirsoleimani (UL)
Structured Parallel Programming for Monte Carlo Tree Search.

2020-09 Myriam Traub (UU)
Measuring Tool Bias & Improving Data Quality for Digital Humanities Research.

2020-10 Alifah Syamsiyah (TUE)
In-database Preprocessing for Process Mining.

2020-11 Sepideh Mesbah (TUD)
Semantic-Enhanced Training Data Augmentation-Methods for Long-Tail Entity Recognition Models.

2020-12 Ward van Breda (VU)
Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment.

2020-13 Marco Virgolin (CWI/ TUD)
Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming.

2020-14 Mark Raasveldt (CWI/UL)
Integrating Analytics with Relational Databases.

2020-15 Georgiadis Konstantinos (OU)
Smart CAT: Machine Learning for Configurable Assessments in Serious Games.

2020-16 Ilona Wilmont (RUN)
Cognitive Aspects of Conceptual Modelling.

2020-17 Daniele Di Mitri (OU)
The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences.

2020-18 Georgios Methenitis (TUD)
Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems.

2020-19 Guido van Capelleveen (UT)
Industrial Symbiosis Recommender Systems.

2020-20 Albert Hankel (VU)
Embedding Green ICT Maturity in Organisations.

2020-21 Karine da Silva Miras de Araujo (VU)
Where is the robot?: Life as it could be.

2020-22 Maryam Masoud Khamis (RUN)
Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar.

2020-23 Rianne Conijn (UT)
The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging.

2020-24 Lenin da Nobrega Medeiros (VUA/RUN)
How are you feeling, human? Towards emotionally supportive chatbots.

2020-25 Xin Du (TUE)
The Uncertainty in Exceptional Model Mining.

2020-26 Krzysztof Leszek Sadowski (UU)
GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization.

2020-27 Ekaterina Muravyeva (OUN)
Personal data and informed consent in an educational context.

2020-28 Bibeg Hang Limbu (OUN)
Multimodal interaction for deliberate practice: Training complex skills with augmented reality.

2020-29 Ioan Gabriel Bucur (RUN)
Being Bayesian about Causal Inference.

2020-30 Bob Zadok Blok (UL)
Creatief, Creatieve, Creatiefst.

2020-31 Gongjin Lan (VU)
Learning better – From Baby to Better.

2020-32 Jason Rhuggenaath (TUE)
Revenue management in online markets: pricing and online advertising.

2020-33 Rick Gilsing (TUE)
Supporting service-dominant business model evaluation in the context of business model innovation.

2020-34 Anna Bon (MU)
Intervention or Collaboration? Redesigning Information and Communication Technologies for Development.

2020-35 Siamak Farshidi (UU)
Multi-Criteria Decision-Making in Software Production.

2021

2021-01 Francisco Xavier Dos Santos Fonseca (TUD)
Location-based Games for Social Interaction in Public Space.

2021-02 Rijk Mercuur (TUD)
Simulating Human Routines:Integrating Social Practice Theory in Agent-Based Models.

2021-03 Seyyed Hadi Hashemi (UVA)
Modeling Users Interacting with Smart Devices.

2021-04 Ioana Jivet (OU)
The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning.

2021-05 Davide Dell'Anna (UU)
Data-Driven Supervision of Autonomous Systems.

2021-06 Daniel Davison (UT)
"Hey robot, what do you think?" How children learn with a social robot.

2021-07 Armel Lefebvre (UU)
Research data management for open science.

2021-08 Nardie Fanchamps (OU)
The Influence of Sense-Reason-Act Programming on Computational Thinking.

2021-09 Cristina Zaga (UT)
The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play.

2021-10 Quinten Meertens (UvA)
Misclassification Bias in Statistical Learning.

2021-11 Anne van Rossum (UL)
Nonparametric Bayesian Methods in Robotic Vision.

2021-12 Lei Pi (UL)
External Knowledge Absorption in Chinese SMEs.

2021-13 Bob R. Schadenberg (UT)
Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning.

2021-14 Negin Samaeemofrad (UL)
Business Incubators: The Impact of Their Support.

2021-15 Onat Ege Adali (TU/e)
Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm.

2021-16 Esam A. H. Ghaleb (MU)
BIMODAL EMOTION RECOGNITION FROM AUDIO-VISUAL CUES.

2021-17 Dario Dotti (UM)
Human Behavior Understanding from motion and bodily cues using deep neural networks.

2021-18 Remi Wieten (UU)
Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks.

2021-19 Roberto Verdecchia (VU)
Architectural Technical Debt: Identification and Management.

2021-20 Masoud Mansoury (TU/e)
Understanding and Mitigating Multi-Sided Exposure Bias in Recommender Systems.

2021-21 Pedro Thiago Timbó Holanda (CWI)
Progressive Indexes.

2021-22 Sihang Qiu (TUD)
Conversational Crowdsourcing.

2021-23 Hugo Manuel Proença (LIACS)
Robust rules for prediction and description.

2021-24 Kaijie Zhu (TUE)
On Efficient Temporal Subgraph Query Processing.

2021-25 Eoin Martino Grua (VUA)
The Future of E-Health is Mobile: Combining AI and Self-Adaptation to Create Adaptive E-Health Mobile Applications.

2021-26 Benno Kruit (CWI & VU)
Reading the Grid: Extending Knowledge Bases from Human-readable Tables.

2021-27 Jelte van Waterschoot (UT)
Personalized and Personal Conversations: Designing Agents Who Want to Connect With You.

2021-28 Christoph Selig (UL)
Understanding the Heterogeneity of Corporate Entrepreneurship Programs.

2022

2022-01 Judith van Stegeren (UT)
Flavor text generation for role-playing video games.

2022-02 Paulo da Costa (TU/e)
Data-driven Prognostics and Logistics Optimisation: A Deep Learning Journey.

2022-03 Ali el Hassouni (VUA)
A Model A Day Keeps The Doctor Away: Reinforcement Learning For Personalized Healthcare.

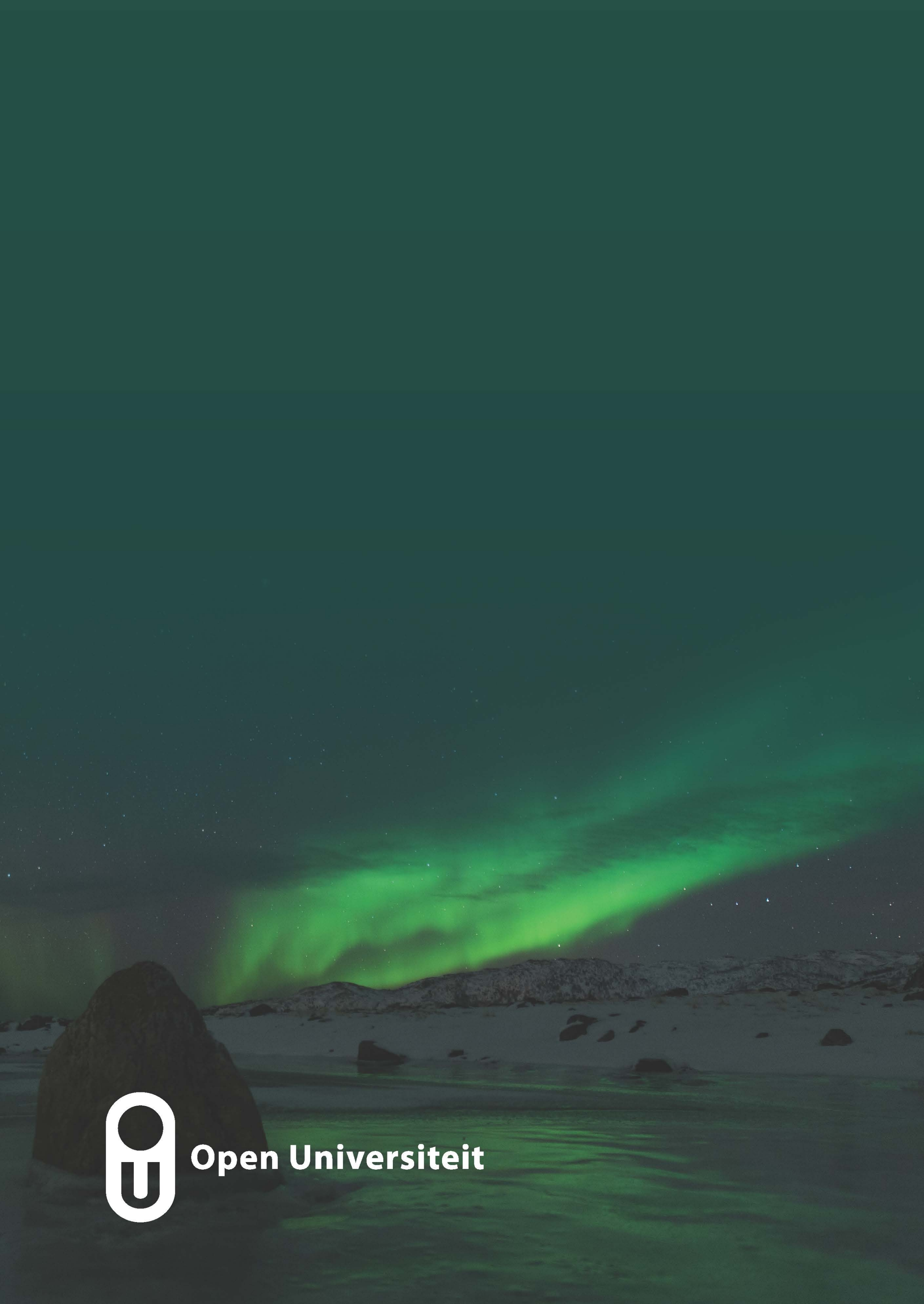
2022-04 Ünal Aksu (UU)
A Cross-Organizational Process Mining Framework.

2022-05 Shiwei Liu (TU/e)
Sparse Neural Network Training with In-Time Over-Parameterization.

2022-06 Reza Refaei Afshar (TU/e)
Machine Learning for Ad Publishers in Real Time Bidding.

2022-07 Sambit Praharaj (OU)
Measuring the Unmeasurable? Towards Automatic Co-located Collaboration Analytics.

2022-08 Maikel L. van Eck (TU/e)
Process Mining for Smart Product Design.


Open Universiteit