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

Praharaj, Sambit; Scheffel, Maren; Drachsler, Hendrik; Specht, Marcus

DOI
10.1109/TLT.2021.3097766

Publication date
2021

Document Version
Final published version

Published in
IEEE Transactions on Learning Technologies

Citation (APA)
Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2021). Literature Review on Co-Located Collaboration Modeling Using Multimodal Learning Analytics - Can We Go the Whole Nine Yards? *IEEE Transactions on Learning Technologies*, *14*(3), 367-385. Article 9490371. https://doi.org/10.1109/TLT.2021.3097766

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

Sambit Praharaj, Maren Scheffel, Hendrik Drachsler, and Marcus Specht

***Abstract*—Collaboration is one of the important 21st-century skills. It can take place in remote or co-located settings. 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. 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 article, 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.**

***Index Terms*—Co-located collaboration (CC), CC analytics, collaboration analytics, collaborative learning tools, multimodal interactions, multimodal learning analytics (MMLA).**

I. Introduction

COLLABORATION is often mentioned as one of the important 21st-century skills [1] and a part of the 4Cs skill set [2] (along with critical thinking, communication, and creativity). When two or more persons work toward a common goal then collaboration occurs [3]. Most of the works in the field of learning analytics support for collaboration have focused on the analysis of distributed (or online) collaboration [4]. However, with the pervasive use of sensors [5], [6], multimodal learning analytics (MMLA) [7]–[9] 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 [5], [6], [10], [11]. Furthermore, sensor technology can be easily scaled up [12] and has become affordable and reliable in the past decade [13]. CC takes place in physical spaces, where all group members share each other's social and epistemic space [14]. "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." [15, pp.1–2, emphasis added]. According to Johnson and Johnson [16], 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.* [17] 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 [16], [17] 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;
5) motivation/individual accountability.

But, the work by Meier *et al.* [17] 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 [18]).

Quality of CC can be detected by different indicators of collaboration such as total speaking time [19] or eye gaze [20]. 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

Manuscript received July 13, 2019; revised May 28, 2020, January 19, 2021, June 20, 2021 and July 8, 2021; accepted July 12, 2021. Date of publication July 19, 2021; date of current version September 3, 2021. *(Corresponding author: Sambit Praharaj.)*

Sambit Praharaj is with the Open University of the Netherlands, 6419 AT Heerlen, Netherlands (e-mail: sambit.praharaj@ou.nl).

Maren Scheffel is with Ruhr-Universität Bochum, 44801 Bochum, Germany (e-mail: maren.scheffel@rub.de).

Hendrik Drachsler is with the DIPF Leibniz Institute for Research and Information in Education, 60323 Frankfurt am Main, Germany, also with the Goethe University, 60323 Frankfurt am Main, Germany, and also with the Open University of the Netherlands, 6419 AT Heerlen, Netherlands (e-mail: Hendrik.Drachsler@ou.nl).

Marcus Specht is with the Delft University of Technology, CD 2628 Delft, Netherlands (e-mail: m.m.specht@tudelft.nl).

Digital Object Identifier 10.1109/TLT.2021.3097766

members [19]. Moreover, different scenarios of CC such as collaborative programming [5], collaborative meetings [6], [21], or collaborative brainstorming [11] 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 [5]; whereas in collaborative meetings gaze direction, body posture, or speaking time of group members are more relevant indicators for collaboration quality [6], [21], [22]. This difference can be attributed to the goals of the tasks performed during CC and the structuring of the task [23], [24]. 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 [6]. 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.* [25], [26], 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 rest of this article is organized as follows. In Section II, we describe the approach taken for this review; it is followed by an explanation of the results obtained from the review in Section III; this is followed by a discussion of the results in Section IV. Finally, Section V concludes this article and we throw some light on limitations, future work, and open questions to be answered.

II. 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 [27]. 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, ScienceDirect, IEEE Xplore, International Society for 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 are 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 are 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.

III. 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

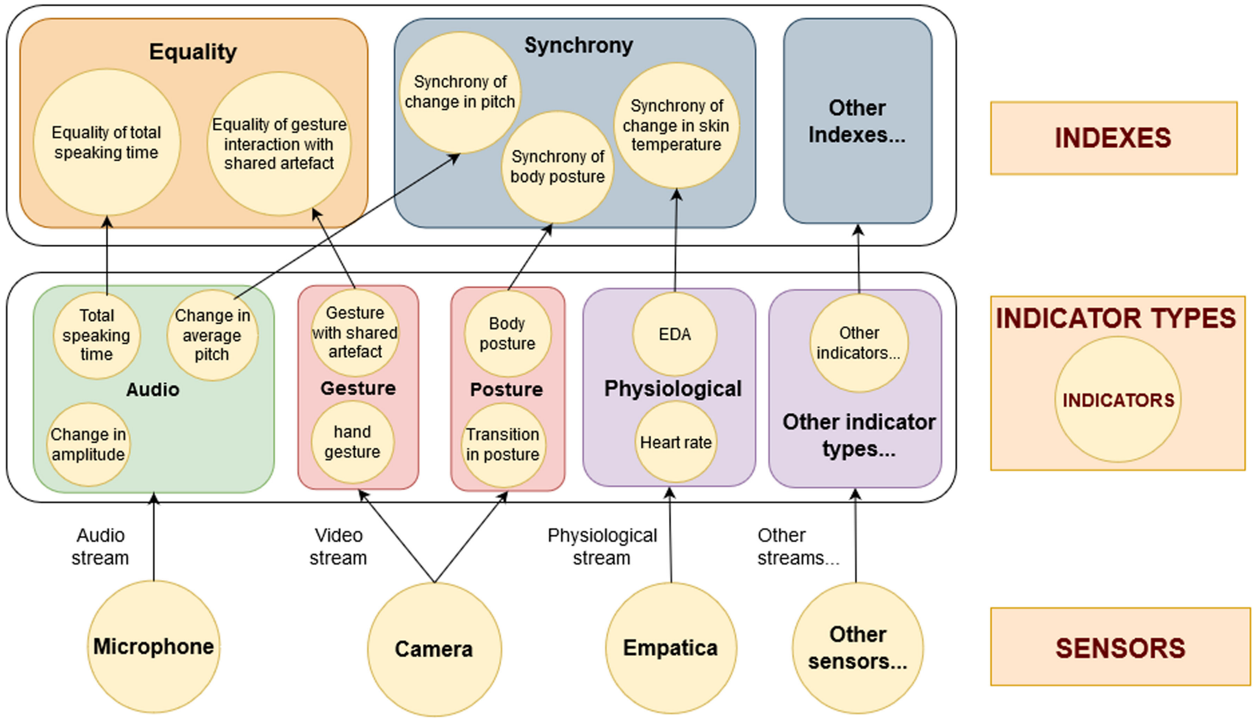

Fig. 1. Outline for the grouping of the articles along with the terminology used in the review (i.e., sensors, indicators, indicator types, and indexes).

studies. Each indicator type cluster is composed of multiple indicators of CC detected by the sensors. For example, audio data are 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 [28]–[30]. 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 [19].

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 [31] for each group member and in some tasks, roles emerge during collaboration [32]. 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., [11], [33]–[35]), sensor-based approach (e.g., [6], [36]–[38]), standard measurement scales like that of Meier's (e.g., [12]), self-reporting mechanism (e.g., [6], [39]), and indirect learning outcome performance measures (e.g., [12]). The types of study found are correlational study (e.g., [36], [37], [40]) and interventionist study (e.g., [6], [11], [35]).

A. Indicators to Assess Collaboration Quality

As a first step, all articles obtained were grouped according to the sensors, indicators, and indicator types. Fig. 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

TABLE I
OVERVIEW OF SENSORS[1] AND INDICATOR TYPES[2]

Indicator types	Sensors	References
Audio	Kinect, microphones, sociometric badge	[41], [42], [43], [5], [44], [45], [46], [47], [48], [38], [49], [50], [51], [33], [21], [6], [52], [53], [54], [28], [55], [56], [19], [38], [35], [57], [34], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [29], [30], [70], [71], [72], [73], [74], [75], [76], [77], [78]
Posture	Kinect, camera, ceiling mounted time-of-flight sensors, sociometric badge	[5], [50], [79], [33], [6], [37], [36], [40], [28], [80], [34], [22], [81], [75], [12]
Gesture	Kinect, camera	[5], [44], [82], [49], [50], [83], [79], [33], [6], [39], [37], [36], [40], [28], [56], [57], [34], [81], [58], [64], [66], [29], [30], [71], [73]
Eye gaze	camera, eye tracker, eye tracking glass	[84], [42], [85], [5], [48], [82], [20], [33], [21], [80], [86], [34], [22], [64], [87], [66], [68], [30]
Spatial	Kinect, camera	[45], [88], [79], [6], [89], [38], [90], [81], [87], [12]
Content (i.e., ideas, knowledge, or task related log data)	tangible-user-interface (TUI), human observer, tablets	[11], [6], [70], [91], [92], [93], [49], [50], [83], [79], [94], [33], [39], [95], [54], [28], [56], [96], [60], [61], [97], [62], [63], [64], [87], [98], [65], [67], [29], [99], [71], [72], [77]
Writing	digital pen	[42], [100], [47], [50], [94], [95]
Physiological	empatica	[44], [101], [80], [96], [57], [102], [103], [78], [104], [13], [105]
Self-reports	online forms, questionnaires	[94], [6], [39], [101], [95], [84]

[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.

proxy for measuring collaboration quality. The indexes (i.e., *synchrony*, *equality*, *individual accountability*, *intraindividual 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.* [17], 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.

1) 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 I gives an overview of the different sensors and their indicator types.

a) 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 [53]; nonverbal features like the total speaking time of group members [19], [52], the number of interruptions [46], and overlap or no overlap duration of speech [53]; the total speaking time of a member together with the attention of other group members measured by their gaze [21]; linguistic features such as frequency of pronouns used, length of the used sentences, and number of prepositions used [48], [51]. 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 [53]. The audio was captured in different settings (e.g., working around a tangible user interface (TUI) [56], working with a sociometric badge worn around the neck [6], and working under camera observation in videotaped *post hoc* studies [33], [34]).

To report further, Terken and Strum [21] 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 [19], [52]. The group was having a conversation around a smart table. The total speaking time of each member was reflected back to them by an LED light display [19] and concentric circle visualization [52] 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 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 does not drive them to actively collaborate.

To analyze other audio indicators in depth, Bassiou *et al.* [53] used *nonlexical* 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 three 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 [41]. 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 nonlexical features like pitch but missed the fine-grained semantic meaning of different nonlexical 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 [55] indicate a good quality of collaboration. Previous research also indicated that overlap in speech is associated with positive group performance [106], [107].

In addition, other works on CC quality focused on expertise detection and productive problem-solving [43], [46], [50], estimation of success [38], collaboration detection [28], and differentiating student learning strategies [44] during CC using the audio indicator type. Zhou *et al.* [46] 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 and Saturnino [43] used the nonverbal audio indicators like presence or absence of speech, silence, pause, and 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.* [38] 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 did not 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 changes 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.

b) Indicator type—posture: This indicator type comprises *body posture* [5], [6], [79], *head movements* [36], [37], [40], and *transitions* between these postures [79]. Schneider and Blikstein [79] 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.

c) Indicator type—gesture: Other works used gestures of group members in open-ended CC scenarios such as building prototypes [36], [37], [82]. Gesture is comprised of *hand movements* [36], [37], *hand gestures* like pointing [5], *hand interactions with an object* [37], [82], *hand interactions around touch screens* [49], [83], or special interaction devices like a TUI. To elaborate further, Spikol *et al.* [82] and Cukurova *et al.* [36], [37] 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.* [5] 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 types, 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.

d) 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.* [20] 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.* [84] used an augmented reality (AR) setup 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 [21]; determining the social context from gaze [85] during group work; observing gaze patterns in post hoc studies [33] 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 [36], [37]. Some studies [21] 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.

e) Indicator type—spatial: This indicator type is a mix of the proximity indicator (i.e., the distance between the group members) [38], [79], [82], [89] and the positioning of the members (i.e., their mobility) [88], [89]. 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 [38], [82]. Other studies, however, did not find any correlation between the distance of group members and the quality of collaboration [79].

f) Indicator type—content: Apart from the indicator types discussed above, this indicator type is a combination of ideas [6], [11] and knowledge (i.e., the content-related knowledge obtained from the interactive devices or the task itself) [39], [83]. Tausch *et al.* [11] 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 [6]. Content of interaction during an activity (around a TUI) was tracked and communicated back to the group members using textual and haptic feedback [39] 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.* [83] 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 multitabletop touch interfaces to track the content of the collaborative task and activity [49], [94], [95]. Echeverria *et al.* [49] 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.* [95] used a multitabletop 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 *et al.* [94] tracked content-related activities by comparing a TUI-based setup and a paper-based setup. They found that the group in the TUI setup had more respect for their peers, better communication, and in turn better collaboration as compared to the group in the TUI setup. 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.

g) Indicator types—writing, physiological, and self-reports: Writing includes different indicators derived from the interactions using a digital pen like the pen stroke analysis [100]. Physiological indicator type has skin temperature [101] and heart rate [78] as indicators. Pijeira *et al.* [101] 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 [39], [84] or some extra information related to the collaboration task (like information about the self-perceived levels of rapport [41]). 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 that 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 that 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.

h) Combined indicator types: Some works (e.g., [6], [44], [56]) used a combination of multiple indicator types. For instance, Martinez-Maldonado *et al.* [56] 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. In addition, the more collaborative group also had more verbal responses after someone spoke.

Worsley and Blikstein [44] 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

TABLE II
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	[41]
		rise and fall of average amplitude	[42], [38]
	Posture	body position, leaning forward	[5]
		relaxed or active body posture	[79]
		head direction	[36], [37]
	Gesture	pointing	[108]
		hand movement	[38]
		using both hands	[79]
		hand position	[36], [37]
	Eye gaze	gaze at speaker, non-speaker/note	[42]
		joint visual attention (JVA)	[84], [85], [48], [51], [86]
	Writing	presence or absence of writing	[42]
	Physiological	electrodermal activation (EDA)	[80], [103], [105], [13]
		heart rate	[78], [104]
Equality ⇑	Audio	jitter	[41]
		total speaking time	[6], [52], [19], [35]
	Posture	body movement, sitting, walking	[6]
		head direction	[37], [40]
	Gesture	all types (i.e., pointing, nodding)	[6]
		hand interactions	[37], [40]
	Content	identifying patterns between data	[91]
		number of ideas and questions	[6]
	Writing	proportion of participation in database modelling	[95]
Individual Accountability ∗	Posture	head direction	[37], [36]
	Gesture	hand position	[37], [36]
Intra-individual Variability ⇓	Posture	head direction	[37], [40]
	Gesture	hand position	[37], [40]
Information Pooling ⇑	Content	web search	[99]
Mutual Understanding ⇑	Audio	dialogues, verbal discourse, statements, questions	[59], [63], [93], [60], [62], [73]
	Content	task related content, knowledge construction, quantitative and conceptual discourse, idea flow	[63], [93], [60], [62], [87], [61], [33], [97], [70]
	Gesture	touch actions on tabletop, hand movement	[83], [73]
	Posture	head orientation	[22]
	Eye gaze	eye gaze on peers, shared devices	[87]
	Spatial	position in the room	[87]
Reciprocal Interaction ⇑	Gesture	hand movement on tabletop	[81]
	Content	explanation, initiation and arguments	[81]

[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.

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.* [6] 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 [96].

2) 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 II shows the overview of the indexes that have been detected practically from different indicators of CC.

a) 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 [41]. Synchrony has been detected using audio indicator type [41], [42] and writing indicator type [42]. Lubold and Pon-Barry [41] 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 [109]. Nakano *et al.* [42] 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 cooccurrence 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 [110]. 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 [37]). 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* [36], [37], [40]. 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 quality of collaboration. Other indicator types like eye gaze [111] (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 [79]. Higher level of physiological synchrony of the skin temperature as seen by [80] and [103] can also indicate good quality of collaboration.

b) Equality: In the work by Lubold and Pon-Barry [41], 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 [37]. Some of these works [37], [40] 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 [19], [52]. 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 toward 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.* [6] 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.* [11] 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.

c) 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 [36], [37]. 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 [36], [37], 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.

d) 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 [37]. 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 [112].

e) Information pooling: It is the accumulation of information measured from the content of the conversation in a particular CC task [99]. In this study, the group members try to gather as much information as possible regarding the shared web search to move toward 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.

f) 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 [59], [63]. 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 [83], [87]. Higher level of mutual understanding indicates higher quality of collaboration.

g) 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 that 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 is 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 II. 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.

B. 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 III gives an 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 two to four 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 subscenario 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.* [6] had problem solving as the scenario that 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.

1) 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

TABLE III
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 $\Uparrow$	[42], [41], [47]
		verbal discourse such as statement, questions, numbers, context of problem, dialogues $*$, misconception in problem solving $*$	[50], [54], [59], [76], [63]
		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 $\Uparrow$	[43], [46], [28], [55], [19], [6]
		linguistic features: pronouns, prepositions $*$, number of anaphoras (anybody, anyone, all, another) used $\Uparrow$	[48], [51]
	Writing	handwriting signal $*$, area used to sketch geometric representation $\Uparrow$, digital pen stroke, writing speed and pressure $*$	[100], [47], [50]
	Spatial	mobility in the room $*$, distance between individuals $*$	[88], [79]
	Eye gaze	JVA (*synchrony*) $\Uparrow$	[48], [51]
	Gesture	using both hands $\Uparrow$, touch actions on tabletop $*$	[79], [113], [50]
	Posture	*synchrony* in active or relaxed body posture $*$, specific body movements $*$	[79], [6], [50]
	Content	task related $\Uparrow$, tabletop content logs matching near to the solution $\Uparrow$, type of contribution (i.e., symmetric$\Uparrow$, asymmetric$\Downarrow$ or individual$\Downarrow$)	[63], [77], [54], [39], [28], [113]
	Self-reports	closeness in group $\Uparrow$, tools used for CC $*$	[113], [39]
Design	Content	*equality* between patterns of data $\Uparrow$, *mutual understanding* from content logs $\Uparrow$, topic of discussion $\Uparrow$, idea flow $*$, task related content $\Uparrow$	[91], [93], [60], [62], [87], [61], [64]
	Eye gaze	JVA (*synchrony*) $\Uparrow$, eye gaze on peers $*$, eye gaze on shared devices $*$	[20], [87], [64]
	Audio	small group talk $*$, dialogues $*$, verbal discourse $*$	[66], [60], [58], [62], [61], [64]
	Gesture	task related touch actions on the TUI $\Uparrow$, facial expression $\Uparrow$, prompts, patterns of gesture $*$	[114], [102], [58]
	Physiological	simultaneous arousal by electrodermal activation (EDA) $\Downarrow$	[102]
	Posture	body motion $*$	[64]
	Spatial	space usage $*$	[87]
Learning	Audio	speech intonation $*$, initiator or listener, verbal discourse $*$	[33], [34], [86], [72], [70], [97], [29], [68], [30], [67]
	Eye gaze	gaze at peers, JVA (*synchrony*) $\Uparrow$	[33], [86], [34]
	Gesture	average hand movement, number of pointing $\Uparrow$	[33], [34], [70]
	Physiological	*synchrony* in arousal by electrodermal activation $\Uparrow$	[103]
	Posture	standing or sitting, bending body $*$	[34], [33]
	Content	topic of discussion related to the task $\Uparrow$, *mutual understanding* from knowledge construction $\Uparrow$	[33], [34], [72], [98], [70], [97], [29], [68], [30], [67]
Engineering design	Spatial	distance between group members $*$	[79]
		distance between group members $\Downarrow$	[40], [82], [36], [37], [38], [90]
	Physiological	near or below average arousal by EDA $\Uparrow$	[44], [57]
	Gesture	*synchrony* in hand movement, wrist movement $\Uparrow$, *synchrony* in pointing $\Uparrow$	[79], [40], [82], [36], [37], [38], [44], [90], [57]
	Posture	*synchrony* in posture $*$, *equality* of posture $\Uparrow$, synchrony in posture $\Uparrow$, *Intra-individual variability* of active or passive posture $\Downarrow$, *Individual accountability* of active or passive posture $*$	[79], [40], [82], [36], [37], [38], [90]
	Audio	amplitude $*$	[38], [90]
Meetings	Audio	*equality* of total speaking time $\Uparrow$, average speech segment length $*$, presence of speech $*$	[21], [6], [52], [35], [75], [78]
		prosody such as average amplitude and average energy variation in speech $*$	[6], [52]
	Content	*equality* of number of ideas $\Uparrow$, content logs matching solution $\Uparrow$	[6], [83]
	Eye gaze	speaker and listener eye gaze $*$	[21], [22]
	Gesture	*equality* of average gesture variation $\Uparrow$	[6]
	Posture	*mutual understanding* from head orientation $\Uparrow$	[22], [75]
	Self-reports	self-report on dominance $*$	[6]
	Physiological	*synchrony* in heart rate $\Uparrow$	[78]
Brainstorming	Content	*equality* of number of ideas $\Uparrow$, *equality* of total speaking time $\Uparrow$, TUI content logs matching the solution $\Uparrow$	[11], [6], [56]
	Gesture	*mutual understanding* from touch actions with the TUI $\Uparrow$	[83], [56]
	Posture	*equality* of body movement like sitting and walking $\Uparrow$	[6]
	Audio	nonverbal audio features like presence or absence, length, verbal interactions $*$	[83], [6], [56]
Healthcare simulation	Spatial	correct mobility and positioning in the room around the patient manikin $\Uparrow$	[45], [89]
Browsing	Content	content of search (*information pooling*) $\Uparrow$	[99]
Gaming	Eye gaze	JVA (*synchrony*) $\Uparrow$	[84], [85]
	Gesture	*reciprocal interaction* by hand movement on tabletop $\Uparrow$	[81]
	Posture	body movement $*$	[81]
	Spatial	space usage around the tabletop $*$	[81]
	Audio	verbal discourse $*$	[73]
Planning	Physiological	*synchrony* in direction of arousal $\Uparrow$	[101]
	Gesture	*synchrony* in pointing $\Uparrow$	[108]
	Content	content of the task like planning, editing, modifying $*$	[92]
Programming	Physiological	*synchrony* in arousal from EDA $\Uparrow$	[96], [80], [13]
	Spatial	proximity $*$	[12]
	Gesture	grabbing mouse from the partner $*$	[5]
	Posture	body position, leaning forward $*$, spending more time in iterating (actively involved in programming a solution) pose as compared to planning, tinkering $\Uparrow$	[5], [12]
Database design	Gesture	number of touch actions on the interactive surface $*$	[94], [49], [95]
	Content	content logs from TUI matching the correctness of solution $\Uparrow$	[94], [49], [95]
	Self-reports	evaluation of own and peers' group work skills $*$	[94], [49], [95]
Concept mapping	Content	content logs from TUI matching the correctness of solution $\Uparrow$	[83]
	Audio	verbal interactions, verbal response to peers $\Uparrow$	[56], [71]
	Gesture	concurrent and parallel touch actions with the TUI $\Downarrow$	[56], [71]

[1]Marked in italics when practically detected.
[2]positive $\Uparrow$, negative $\Downarrow$, unclear or no effect $*$.

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 categories 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 [21], higher instances of synchrony in eye gaze (or JVA) is indicative of the better quality of CC [20], [84]. So, eye gaze is dependent on the context. However, synchrony in posture may not be a good predictor of collaboration quality [79].

Besides, less distance between the collaborating group members means that they have a higher level of comfort and the quality of collaboration is better [38], [82] for that group. However, this was that not consistent with all the previous works [79]. Schneider and Blikstein [79] 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 [6], [19], interruptions [46], and overlap in the speech [100], [106], [107] 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 [41]. Silence has been used as a feature to train a machine learning model to detect the CC quality [43] 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 [6], [21] 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 [114]. It meant that they are working collaboratively toward the task objective instead of working individually. In general, groups where members used both their hands [79] and had near average hand movement [44] showed better quality of collaboration. So, the gesture is dependent on the context. For example, a planning scenario [108] 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 [5] 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 [101]. However, it was not true all the time [103]. During collaborative learning, using electrodermal activation (EDA) in a group of the 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 [103] 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 [102]. 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 [102], this arousal can be because of rising stress levels or confusion levels leading to unproductive collaboration.

As mentioned earlier in Section III-A, 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 setup 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 setup only or a hybrid setup 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 I, it is clear that these indicators define the type of setup needed for collaboration detection. In addition, these indicators vary a lot in different scenarios because of the differing goals and parameters.

2) 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 coconstruction, reflection, coherence, misconception, or uncertainty). To elaborate the parameters, *dominance* [6], [79] includes the dominance and leadership parameter. The *coupling* [41], [77] includes the comfort level, coupling, coordination, and rapport between the group members. *Coherence* [48], [51] includes verbal coherence where group members build upon each other's ideas and verbal-discourse coherence. *Engagement* [37], [38],

TABLE IV
MODELING A CC SCENARIO—COLLABORATIVE PROBLEM-SOLVING

Parameters[a]	**Indicator types**	**Indexes**[b]	**References**
Expertise ⤊	Audio	Dialogue management∗	[43], [46], [50]
	Posture	Synchrony∗, IA∗	[50]
	Gesture	Synchrony∗, IA∗	[50]
	Writing	Synchrony∗, IA∗	[100], [50]
Dominance ⤋	Audio	Equality	[6], [19]
	Posture	Synchrony, Equality	[79], [6]
	Gesture	Synchrony, Equality	[79], [6]
	Writing	Synchrony∗, Equality∗	[100]
	Self-reports	Equality	[6]
Coupling ⤊	Audio	Synchrony, Equality	[41], [77]
	Posture	Equality	[79]
	Gesture	Mutual understanding∗, Synchrony	[113], [79]
	Content	Mutual understanding∗, Synchrony	[113], [79]
	Self-reports	Mutual understanding∗	[113]
Reflection ⤊	Self-reports	Information pooling∗	[39]
	Audio	Information pooling∗	[54]
	Content	Information pooling∗	[54]
Roles ⤊	Audio	Mutual understanding∗, Task division∗	[55]
Coherence ⤊	Audio	Reaching consensus∗	[48], [51], [77]
	Eye gaze	Reaching consensus∗	[48], [51]
	Content	Reaching consensus∗	[77]
Uncertainty ⤋	Audio	Mutual understanding	[59]
Misconception ⤋	Audio	Mutual understanding	[63]
	Content	Mutual understanding	[63]
Engagement ⤊	Audio	Equality∗, Synchrony	[53], [28], [42]
	Posture	Mutual understanding∗	[28]
	Gesture	Mutual understanding∗	[28]
	Content	Mutual understanding∗	[28]
	Eye gaze	Synchrony	[42]
	Writing	Synchrony	[42]
	Spatial	IVA∗	[88]

[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.* [17] and the practically detected ones.

[40] includes engagement, participation, and interaction. *Learning strategy* [44], [57] reports the strategy adopted by the group during CC. *Heterogeneity* [64], [72], [98] refers to the difference in previous knowledge or the difference in capabilities among the group members. Fischer *et al.* [115], [116] proposed that heterogeneous collaborating teams possess a symmetry of ignorance [117], which makes them more interesting, wherein neither team possesses the full breadth of knowledge to solve the problem independently, thus collaborating with each other can help to resolve the problem. *Roles* (another parameter of CC) [65], [67], when self-assigned, evokes a sense of responsibility within the group that are designed to facilitate group progress toward a goal [31]. Some roles are preassigned and some emerge during collaboration [32]. Moving on to another parameter, *misconception*, it arises when group members share a common understanding among themselves without any refutation and reflection [63]. *Uncertainty* is more common among group members with less mutual understanding [59]. During design scenarios of CC, *critique* [62] 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 coconstruction* [33], [70], [97] is sharing each other's ideas and using them to build the shared understanding of the situation or the task at hand.

C. *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 IV gives an overview of this mapping.

Now, we drill down further into the CPS scenario in Table IV. 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

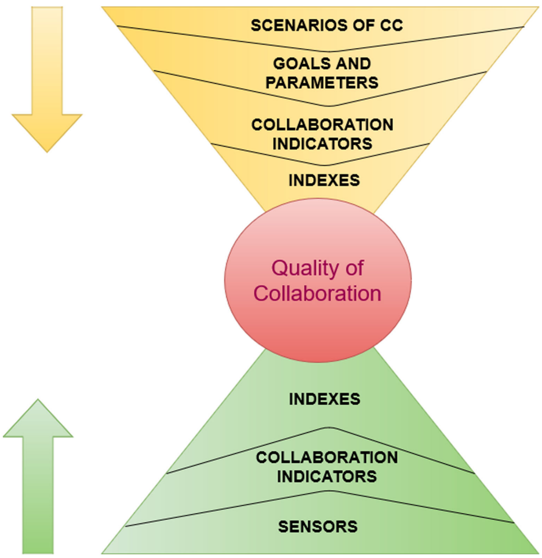

Fig. 2. Confluence of both approaches of CC quality detection.

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 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 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 [6], [19].

IV. 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 [6], [19], and number and duration of overlap of audio [100] 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 [38], [82] or sometimes there is no relation [79] with CC quality. The comprehensive overview of the indicators will help practitioners to choose the sensors and indicators according to their setup. 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 [36], [37], as in the case of individual accountability, thus failing to detect CC quality. Another limitation is the use of machine learning approaches [5], [28], [43], [76], 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 [43] 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.* [118] and is still an open question. Some machine learning models that 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 that use these external objects tend to be inclined toward 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 [20] 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 [21]. 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 [102]. 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 setup. 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 setup 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.

V. Conclusion 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 toward 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 Table IV 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 [119] (as outlined in [17]) 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 versus 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 Section I, 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.

References

[1] C. Dede, "Comparing frameworks for 21st century skills," in *21st Century Skills: Rethinking how Students Learn*. Bloomington, IN, USA: Solution Tree Press, vol. 20, 2010, pp. 51–76.

[2] C. Kivunja, "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.*, vol. 6, no. 2, p. 224, 2015

[3] P. Dillenbourg, "What do you mean by collaborative learning?" in *Collaborative-Learning: Cognitive and Computational Approaches*. Oxford, U.K.: Elsevier, 1999.

[4] H. Jeong and C. E. Hmelo-Silver, "An overview of CSCL methodologies," in *Proc. 9th Int. Conf. Learn. Sci.*, Chicago, IL, USA, vol. 1, 2010, pp. 921–928.

[5] S. Grover, M. Bienkowski, A. Tamrakar, B. Siddiquie, D. Salter, and A. Divakaran, "Multimodal analytics to study collaborative problem solving in pair programming," in *Proc. 6th Int. Conf. Learn. Analytics Knowl.*, New York, NY, USA, 2016, pp. 516–517.

[6] T. Kim, A. Chang, L. Holland, and A. S. Pentland, "Meeting mediator: Enhancing group collaboration using sociometric feedback," in *Proc. ACM Conf. Comput. Supported Cooperative Work*, San Diego, CA, USA, 2008, pp. 457–466.

[7] P. Blikstein, "Multimodal learning analytics," in *Proc. 3 rd Int. Conf. Learn. Analytics Knowl.*, Leuven, Belgium, 2013, pp. 102–106.

[8] S. Praharaj, M. Scheffel, H. Drachsler, and M. Specht, "MULTIFOCUS: Multimodal learning analytics for co-located collaboration understanding and support," in *Proc. 13th Eur. Conf. Technol. Enhanced Learn. Doctoral Consortium*, Leeds, U.K., 2018, pp. 1–6.

[9] D. Di Mitri, J. Schneider, M. Specht, and H. Drachsler, "From signals to knowledge: A conceptual model for multimodal learning analytics," *J. Comput. Assisted Learn.*, vol. 34, no. 4, pp. 338–349, Jul. 2018.

[10] S. Praharaj, M. Scheffel, M. Schmitz, M. Specht, and H. Drachsler, "Towards automatic collaboration analytics for group speech data using learning analytics," *Sensors*, vol. 21, no. 9, May 2021, Art. no. 3156.

[11] S. Tausch, D. Hausen, I. Kosan, A. Raltchev, and H. Hussmann, "Groupgarden: Supporting brainstorming through a metaphorical group mirror on table or wall," in *Proc. 8th Nordic Conf. Human-Comput. Interact.: Fun, Fast, Foundational*, Helsinki, Finland, Oct. 26-30, 2014, pp. 541–550.

[12] J. M. Reilly, M. Ravenell, and B. Schneider, "Exploring collaboration using motion sensors and multi-modal learning analytics," in *Proc. Int. Conf. Educ. Data Mining*, Raleigh, NC, USA, 2018.

[13] E. L. Starr, J. M. Reilly, and B. Schneider, "Toward using multimodal learning analytics to support and measure collaboration in co-located dyads," in *Proc. 13th Int. Conf. Learn. Sci.*, London, U.K., 2018, pp. 448–455.

[14] S. Praharaj, "Co-located collaboration analytics," in *Proc. 21st Int. Conf. Multimodal Interaction*, Suzhou, China, 2019, pp. 473–476.

[15] G. Stahl, N. Law, and F. Hesse, "Reigniting CSCL flash themes," *Int. J. Comput.-Supported Collaborative Learn.*, vol. 8, no. 4, pp. 369–374, Oct. 2013.

[16] D. W. Johnson and R. T. Johnson, "An educational psychology success story: Social interdependence theory and cooperative learning," *Educational Researcher*, vol. 38, no. 5, pp. 365–379, 2009.

[17] A. Meier, H. Spada, and N. Rummel, "A rating scheme for assessing the quality of computer-supported collaboration processes," *Int. J. Comput.-Supported Collaborative Learn.*, vol. 2, no. 1, pp. 63–86, 2007.

[18] S. Child and S. Shaw, "Collaboration in the 21st century: Implications for assessment," *Econ.*, vol. 21, 2015, Art. no. 2008.

[19] K. Bachour, F. Kaplan, and P. Dillenbourg, "An interactive table for supporting participation balance in face-to-face collaborative learning," *IEEE Trans. Learn. Technol.*, vol. 3, no. 3, pp. 203–213, Jul./Sep. 2010.

[20] B. Schneider, K. Sharma, S. Cuendet, G. Zufferey, P. Dillenbourg, and R. D. Pea, "3D tangibles facilitate joint visual attention in dyads," in *Proc. 11th Int. Conf. Comput. Supported Collaborative Learn.*, vol. 1, Gothenburg, Sweden, 2015, pp. 156–165.

[21] J. Terken and J. Sturm, "Multimodal support for social dynamics in co-located meetings," *Pers. Ubiquitous Comput.*, vol. 14, no. 8, pp. 703–714, Mar. 2010.

[22] R. Stiefelhagen and J. Zhu, "Head orientation and gaze direction in meetings," in *Proc. Int. Conf. Extended Abstr. Hum. Factors Comput. Syst.*, Minneapolis, MN, USA, 2002, pp. 858–859.

[23] C. A. Collazos, L. A. Guerrero, J. A. Pino, S. F. Ochoa, and G. Stahl, "Designing collaborative learning environments using digital games," *J. Universal Comput. Sci.*, vol. 13, no. 7, pp. 1022–1032, 2007.

[24] C. A. Collazos, L. A. Guerrero, J. A. Pino, and S. F. Ochoa, "Collaborative scenarios to promote positive interdependence among group members," in *Proc. Int. Conf. Collaboration Technol.*, Autrans, France,, 2003, pp. 356–370.

[25] P. Dillenbourg, M. J. Baker, A. Blaye, and C. O'Malley, "The evolution of research on collaborative learning," in *Learning in Humans and Machine: Towards an Interdisciplinary Learning Science*. Oxford: Elsevier, 1996, pp. 189–211.

[26] P. Dillenbourg and D. Traum, "Sharing solutions: Persistence and grounding in multimodal collaborative problem solving," *J. Learn. Sciences*, vol. 15, no. 1, pp. 121–151, 2006.

[27] A. Liberati *et al.*, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration," *PLoS Med.*, vol. 6, no. 7, pp. e1–e34, 2009.

[28] S. A. Viswanathan and K. VanLehn, "High accuracy detection of collaboration from log data and superficial speech features," in *Proc. 12th Int. Conf. Comput. Supported Collaborative Learn.*, Philadelphia, PA, USA, 2017, vol. 1, pp. 335–342.

[29] L. Hardy and T. White, "Meaning-making in collaborative activity: Effort toward coherent, but not shared, interpretations of the problem," in *Proc. 11th Int. Conf. Comput. Supported Collaborative Learn.*, vol. 1, Gothenburg, Sweden, 2015, pp. 435–442.

[30] A. Andrade-Lotero, J. Danish, J. Moreno, and L. Perez, "Measuring 'framing' differences of single-mouse and tangible inputs on patterns of collaborative learning," in *Proc. 10th Int. Conf. Comput. Supported Collaborative Learn.*, vol. 1, Madison, WI, USA, 2013, pp. 34–41.

[31] A. P. Hare, "Types of roles in small groups: A bit of history and a current perspective," *Small Group Res.*, vol. 25, no. 3, pp. 433–448, 1994.

[32] J.-W. Strijbos and A. Weinberger, "Emerging and scripted roles in computer-supported collaborative learning," *Comput. Hum. Behav.*, vol. 26, no. 4, pp. 491–494, Jul. 2010.

[33] R. E. Scherr and D. Hammer, "Student behavior and epistemological framing: Examples from collaborative active-learning activities in physics," *Cogn. Instruct.*, vol. 27, no. 2, pp. 147–174, Mar. 2009.

[34] J. Davidsen and T. Ryberg, "'This is the size of one meter': Children's bodily-material collaboration," *Int. J. Comput.-Supported Collaborative Learn.*, vol. 12, no. 1, pp. 65–90, Feb. 2017.

[35] S. Praharaj, M. Scheffel, H. Drachsler, and M. Specht, "Multimodal analytics for real-time feedback in co-located collaboration," in *Proc. 13th Eur. Conf. Technol. Enhanced Learn.*, Leeds, U.K., 2018, pp. 187–201.

[36] M. Cukurova, R. Luckin, M. Mavrikis, and E. Millán, "Machine and human observable differences in groups' collaborative problem-solving behaviours," in *Proc. 12th Eur. Conf. Technol. Enhanced Learn.*, Tallinn, Estonia, 2017, pp. 17–29.

[37] M. Cukurova, R. Luckin, E. Millán, and M. Mavrikis, "The NISPI framework: Analysing collaborative problem-solving from students' physical interactions," *Comput. Educ.*, vol. 116, pp. 93–109, Jan. 2018.

[38] D. Spikol, E. Ruffaldi, L. Landolfi, and M. Cukurova, "Estimation of success in collaborative learning based on multimodal learning analytics features," in *Proc. 17th IEEE Int. Conf. Adv. Learn. Technol.*, Timisoara, Romania, 2017, pp. 269–273.

[39] D. Anastasiou and E. Ras, "A questionnaire-based case study on feedback by a tangible interface," in *Proc. ACM Workshop Intell. Interfaces Ubiquitous Smart Learn.*, Limassol, Cyprus, 2017, pp. 39–42.

[40] M. Cukurova, R. Luckin, E. Millán, M. Mavrikis, and D. Spikol, "Diagnosing collaboration in practice-based learning: Equality and intra-individual variability of physical interactivity," in *Proc. 12th Eur. Conf. Technol. Enhanced Learn.*, Tallinn, Estonia, 2017, pp. 30–42.

[41] N. Lubold and H. Pon-Barry, "Acoustic-prosodic entrainment and rapport in collaborative learning dialogues," in *Proc. ACM Workshop Multimodal Learn. Analytics Workshop Grand Challenge*, Istanbul, Turkey, 2014, pp. 5–12.

[42] Y. I. Nakano, S. Nihonyanagi, Y. Takase, Y. Hayashi, and S. Okada, "Predicting participation styles using co-occurrence patterns of nonverbal behaviors in collaborative learning," in *Proc. 17th ACM Int. Conf. Multimodal Interact.*, Seattle, WA, USA, 2015, pp. 91–98.

[43] S. Luz, "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 Interact.*, Sydney, Australia, 2013, pp. 575–582.

[44] M. Worsley and P. Blikstein, "Leveraging multimodal learning analytics to differentiate student learning strategies," in *Proc. 5th Int. Conf. Learn. Analytics Knowl.*, Poughkeepsie, NY, USA, 2015, pp. 360–367.

[45] R. Martinez-Maldonado *et al.* "Analytics meet patient manikins: Challenges in an authentic small-group healthcare simulation classroom," in *Proc. 7th Int. Conf. Learn. Analytics Knowl.*, 2017, pp. 90–94.

[46] S. Oviatt, K. Hang, J. Zhou, and F. Chen, "Spoken interruptions signal productive problem solving and domain expertise in mathematics," in *Proc. 17th ACM Int. Conf. Multimodal Interact.*, Seattle, WA, USA, 2015, pp. 311–318.

[47] S. Scherer, N. Weibel, L.-P. Morency, and S. Oviatt, "Multimodal prediction of expertise and leadership in learning groups," in *Proc. 1st Int. Workshop Multimodal Learn. Analytics*, Santa Monica, CA, USA, 2012, pp. 1–8.

[48] B. Schneider and R. Pea, "The effect of mutual gaze perception on students' verbal coordination," in *Proc. 7th Int. Conf. Educ. Data Mining*, London, U.K., 2014, pp. 138–144.

[49] V. Echeverria, R. Martinez-Maldonado, K. Chiluiza, and S. B. Shum, "DBCollab: Automated feedback for face-to-face group database design," in *Proc. 25th Int. Conf. Comput. Educ.*, Christchurch, New Zealand, 2017, pp. 156–165.

[50] X. Ochoa, K. Chiluiza, G. Méndez, G. Luzardo, B. Guamán, and J. Castells, "Expertise estimation based on simple multimodal features," in *Proc. 15th ACM Int. Conf. Multimodal Interact.*, 2013, pp. 583–590.

[51] B. Schneider and R. Pea, "Does seeing one another's gaze affect group dialogue? A computational approach," *J. Learn. Analytics*, vol. 2, no. 2, pp. 107–133, 2015.

[52] T. Bergstrom and K. Karahalios, "Conversation clock: Visualizing audio patterns in co-located groups," in *Proc. 40th Annu. Hawaii Int. Conf. Syst. Sci.*, Big Island, HI, USA, 2007, p. 78.

[53] N. Bassiou *et al.*, "Privacy-preserving speech analytics for automatic assessment of student collaboration," in *Proc. 17th Annu. Conf. Int. Speech Commun. Assoc.*, San Francisco, CA, USA, 2016, pp. 888–892.

[54] K. Thompson, S. Kennedy-Clark, L. Markauskaite, and V. Southavilay, "Discovering processes and patterns of learning in collaborative learning environments using multi-modal discourse analysis," *Res. Pract. Technol. Enhanced Learn.*, vol. 9, no. 2, pp. 215–240, 2014.

[55] J. Kim *et al.*, "Vocal turn-taking patterns in groups of children performing collaborative tasks: An exploratory study," in *Proc. 16th Annu. Conf. Int. Speech Commun. Assoc.*, Dresden, Germany, 2015, pp. 1645–1649.

[56] R. Martinez-Maldonado, Y. Dimitriadis, A. Martinez-Monés, J. Kay, and K. Yacef, "Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop," *Int. J. Comput.-Supported Collaborative Learn.*, vol. 8, no. 4, pp. 455–485, Nov. 2013.

[57] M. Worsley and P. Blikstein, "A multimodal analysis of making," *Int. J. Artif. Intell. Educ.*, vol. 28, no. 3, pp. 385–419, 2018.

[58] M. Emara, M. Tscholl, Y. Dong, and G. Biswas, "Analyzing students' collaborative regulation behaviors in a classroom-integrated open ended learning environment," in *Proc. 12th Int. Conf. Comput. Supported Collaborative Learn.*, Philadelphia, PA, USA, 2017, vol. 1, pp. 319–326.

[59] F. J. Rodríguez, K. M. Price, and K. E. Boyer, "Expressing and addressing uncertainty: A study of collaborative problem-solving dialogues," in *Proc. 12th Int. Conf. Comput. Supported Collaborative Learn.*, Philadelphia, PA, USA, vol. 1, 2017, pp. 207–214.

[60] C. Dornfeld, N. Zhao, and S. Puntambekar, "A mixed-methods approach for studying collaborative learning processes at individual and group levels," in *Proc. 12th Int. Conf. Comput. Supported Collaborative Learn.*, Philadelphia, PA, USA, 2017, vol. 1, pp. 191–198.

[61] H. Fake, J. Whiteman, and B. Bannan, "Collaborative game design: A bounded case study of undergraduate students in a capstone course," in *Proc. 12th Int. Conf. Comput. Supported Collaborative Learn.*, Philadelphia, PA, USA, vol. 1, 2017, pp. 119–126.

[62] E. McBride, J. Vitale, L. Applebaum, and M. Linn, "Examining the flow of ideas during critique activities in a design project," in *Proc. 12th Int. Conf. Comput. Supported Collaborative Learn.*, Philadelphia, PA, USA, 2017, pp. 41–48.

[63] R. Abdu, "The development of situational-misconceptions in math problem solving," in *Proc. 11th Int. Conf. Comput. Supported Collaborative Learn.*, Gothenburg, Sweden, 2015, pp. 78–85.

[64] V. J. Flood, M. Neff, and D. Abrahamson, “Boundary interactions: Resolving interdisciplinary collaboration challenges using digitized embodied performances,” in *Proc. 11th Int. Conf. Comput. Supported Collaborative Learn.*, Gothenburg, Sweden, 2015, pp. 94–101.
[65] A. F. Wise, A. N. Antle, J. Warren, A. May, M. Fan, and A. Macaranas, “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. Comput. Supported Collaborative Learn.*, Gothenburg, Sweden, 2015, pp. 236–243.
[66] J. D. Wake, F. Guribye, and B. Wasson, “Students’ design decisions in collaborative design of location-based games for learning,” in *Proc. 11th Int. Conf. Comput. Supported Collaborative Learn.*, Gothenburg, Sweden, 2015, pp. 324–331.
[67] N. D. Martin, D. Gnesdilow, and S. Puntambekar, “Peer scaffolding to learn science in symmetrical groups collaborating over time,” in *Proc. 11th Int. Conf. Comput. Supported Collaborative Learn.*, Gothenburg, Sweden, 2015, pp. 340–347.
[68] A. Andrade, “Using situated-action networks to visualize complex learning,” in *Proc. 11th Int. Conf. Comput. Supported Collaborative Learn.*, Gothenburg, Sweden, 2015, pp. 372–379.
[69] C. Dornfeld and S. Puntambekar, “Emergent roles and collaborative discourse over time,” in *Proc. 11th Int. Conf. Comput. Supported Collaborative Learn.*, Gothenburg, Sweden, 2015, pp. 380–387.
[70] K. Thompson, D. Ashe, P. Yeoman, and M. Parisio, “Phases of design: Following idea development and patterns of collaborative discussion in a learning by design project,” in *Proc. 10th Int. Conf. Comput. Supported Collaborative Learn.*, Madison, WI, USA, 2013, pp. 494–501.
[71] R. Martinez, J. Kay, and K. Yacef, “Visualisations for longitudinal participation, contribution and progress of a collaborative task at the tabletop,” in *Proc. 9th Int. Conf. Comput. Supported Collaborative Learn.*, 2011, pp. 25–32.
[72] L. H. Wong, W. Chen, Y. Zhan, and C. K. Chin, “Analysis of small group interactions in a seamless language learning environment: An artifact-oriented approach,” in *Proc. 9th Int. Conf. Comput. Supported Collaborative Learn.*, Hong Kong, China, 2011, pp. 144–151.
[73] C. Johansson, Z. Ahmet, J. Tholander, F. Aleo, M. Jonsson, and S. Sumon, “Weather gods and fruit kids-embodying abstract concepts using tactile feedback and whole body interaction,” in *Proc. 9th Int. Conf. Comput. Supported Collaborative Learn.*, 2011, pp. 160–167.
[74] R. Noel *et al.*, “Exploring collaborative writing of user stories with multimodal learning analytics: A case study on a software engineering course,” *IEEE Access*, vol. 6, pp. 67 783–67 798, Oct. 2018.
[75] I. Bhattacharya *et al.*, “A multimodal-sensor-enabled room for unobtrusive group meeting analysis,” in *Proc. 20th Int. Conf. Multimodal Interaction*, Boulder, Colorado, USA, 2018, pp. 347–355.
[76] A. E. Stewart, Z. A. Keirn, and S. K. D’Mello, “Multimodal modeling of coordination and coregulation patterns in speech rate during triadic collaborative problem solving,” in *Proc. 20th Int. Conf. Multimodal Interaction*, Boulder, Colorado, USA, 2018, pp. 21–30.
[77] J. K. Olsen and S. Finkelstein, “Through the (thin-slice) looking glass: An initial look at rapport and co-construction within peer collaboration,” in *Proc. 12th Int. Conf. Comput. Supported Collaborative Learn.*, Philadelphia, PA, USA, Jun. 18-22, 2017, pp. 511–518.
[78] R. Henning, A. Armstead, and J. Ferris, “Social psychophysiological compliance in a four-person research team,” *Appl. Ergonom.*, vol. 40, no. 6, pp. 1004–1010, Nov. 2009.
[79] B. Schneider and P. Blikstein, “Unraveling students’ interaction around a tangible interface using multimodal learning analytics,” *J. Educ. Data Mining*, vol. 7, no. 3, pp. 89–116, Oct. 2015.
[80] Y. Dich, J. Reilly, and B. Schneider, “Using physiological synchrony as an indicator of collaboration quality, task performance and learning,” in *Proc. 19th Int. Conf. Artif. Intell. Educ.*, London, U.K., 2018, pp. 98–110.
[81] A. F. Wise, A. N. Antle, and J. Warren, “Explanation-giving in a collaborative tangible tabletop game: Initiation, positionality, valence and action-orientation,” in *Proc. 12th Int. Conf. Comput. Supported Collaborative Learn.*, Philadelphia, PA, USA, Jun. 2017, vol. 1, pp. 471–478.
[82] D. Spikol, E. Ruffaldi, and M. Cukurova, “Using multimodal learning analytics to identify aspects of collaboration in project-based learning,” in *Proc. 12th Int. Conf. Comput. Supported Collaborative Learn.*, Philadelphia, PA, USA, Jun. 2017, pp. 263–270.
[83] R. Martinez-Maldonado, A. Clayphan, K. Yacef, and J. Kay, “MTFeedback: Providing notifications to enhance teacher awareness of small group work in the classroom,” *IEEE Trans. Learn. Technol.*, vol. 8, no. 2, pp. 187–200, 2015.
[84] A. Dierker, C. Mertes, T. Hermann, M. Hanheide, and G. Sagerer, “Mediated attention with multimodal augmented reality,” in *Proc. Int. Conf. Multimodal Interfaces*, Cambridge, MA, USA, 2009, pp. 245–252.
[85] W. Li, M.-A. Nüssli, and P. Jermann, “Gaze quality assisted automatic recognition of social contexts in collaborative tetris,” in *Proc. Int. Conf. Multimodal Interfaces*, Beijing, China, 2010, pp. 1–8.
[86] S. Andrist, A. Ruis, and D. W. Shaffer, “A network analytic approach to gaze coordination during a collaborative task,” *Comput. Human Behav.*, vol. 89, pp. 339–348, Dec. 2018.
[87] R. Martinez-Maldonado *et al.*, “Learning about collaborative design for learning in a multi-surface design studio,” in *Proc. 11th Int. Conf. Comput. Supported Collaborative Learn.*, Gothenburg, Sweden, 2015, pp. 174–181.
[88] D. Healion, S. Russell, M. Cukurova, and D. Spikol, “Tracing physical movement during practice-based learning through multimodal learning analytics,” in *Proc. 7th Int. Conf. Learn. Analytics Knowl.*, 2017, pp. 588–589.
[89] R. Martinez-Maldonado, M. Pechenizkiy, S. Buckingham-Shum, T. Power, C. Hayes, and C. Axisa, “Modelling embodied mobility teamwork strategies in a simulation-based healthcare classroom,” in *Proc. 25th Conf. User Modeling, Adapt. Personalization*, Bratislava, Slovakia, 2017, pp. 308–312.
[90] D. Spikol, E. Ruffaldi, G. Dabisias, and M. Cukurova, “Supervised machine learning in multimodal learning analytics for estimating success in project-based learning,” *J. Comput. Assisted Learn.*, vol. 34, no. 4, pp. 366–377, May 2018.
[91] E. Coopey, R. B. Shapiro, and E. Danahy, “Collaborative spatial classification,” in *Proc. 4th Int. Conf. Learn. Analytics Knowl.*, Indianapolis, IN, USA, Mar. 24-28, 2014, pp. 138–142.
[92] A. Harrer, “Analytics of collaborative planning in metafora: Architecture, data, and analytic methods,” in *Proc. 3rd Int. Conf. Learn. Analytics Knowl.*, Leuven, Belgium, 2013, pp. 255–259.
[93] F. Fischer, J. Bruhn, C. Gräsel, and H. Mandl, “Fostering collaborative knowledge construction with visualization tools,” *Learn. Instruction*, vol. 12, no. 2, pp. 213–232, Apr. 2002.
[94] M. Wong-Villacrés, R. Granda, M. Ortiz, and K. Chiluiza, “Exploring the impact of a tabletop-generated group work feedback on students’ collaborative skills,” in *Int. Workshop Learn. Analytics Across Phys. Digit. Spaces*, Edinburgh, U.K., 2016, pp. 58–64.
[95] R. X. Granda, V. Echeverría, K. Chiluiza, and M. Wong-Villacrés, “Supporting the assessment of collaborative design activities in multi-tabletop classrooms,” in *Proc. Asia-Pacific Conf. Comput. Aided Syst. Eng.*, Quito, Ecuador, 2015, pp. 270–275.
[96] L. Ahonen, B. U. Cowley, A. Hellas, and K. Puolamäki, “Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment,”*Scientific Rep.*, vol. 8, no. 1, Feb. 2018, Art. no. 3138.
[97] J. Oshima, A. Ohsaki, Y. Yamada, and R. Oshima, “Collective knowledge advancement and conceptual understanding of complex scientific concepts in the jigsaw instruction,” in *Proc. 12th Int. Conf. Comput. Supported Collaborative Learn.*, Philadelphia, PA, USA, 2017, pp. 57–64.
[98] S. Manske, T. Hecking, I.-A. Chounta, S. Werneburg, and H. U. Hoppe, “Using differences to make a difference: A study on heterogeneity of learning groups,” in *Proc. 11th Int. Conf. Comput. Supported Collaborative Learn.*, Gothenburg, Sweden, 2015, vol. 1, pp. 182–189.
[99] T. Hashida, K. Nomura, M. Iida, and T. Naemura, “Inter-personal browsing: Supporting cooperative web searching by face-to-face sharing of browser pages,” in *Proc. 10th Int. Conf. Comput. Supported Collaborative Learn.*, Madison, WI, USA, 2013, pp. 224–231.
[100] J. Zhou, K. Hang, S. Oviatt, K. Yu, and F. Chen, “Combining empirical and machine learning techniques to predict math expertise using pen signal features,” in *Proc. ACM Workshop Multimodal Learn. Analytics Workshop Grand Challenge*, Istanbul, Turkey, 2014, pp. 29–36.
[101] H. J. Pijeira-Díaz, H. Drachsler, S. Järvelä, and P. A. Kirschner, “Investigating collaborative learning success with physiological coupling indices based on electrodermal activity,” in *Proc. 6th Int. Conf. Learn. Analytics Knowl.*, Edinburgh, U.K., 2016, pp. 64–73.
[102] J. Malmberg, S. Järvelä, J. Holappa, E. Haataja, X. Huang, and A. Siipo, “Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning?,” *Comput. Human Behav.*, vol. 96, pp. 235–245, Jul. 2019.
[103] H. J. Pijeira-Díaz, H. Drachsler, S. Järvelä, and P. A. Kirschner, “Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members?,” *Comput. Human Behav.*, vol. 92, pp. 188–197, Mar. 2019.

[104] R. A. Henning, W. Boucsein, and M. C. Gil, "Social-physiological compliance as a determinant of team performance," *Int. J. Psychophysiol.*, vol. 40, no. 3, pp. 221–232, 2001.

[105] A. N. Elkins, E. R. Muth, A. W. Hoover, A. D. Walker, T. L. Carpenter, and F. S. Switzer, "Physiological compliance and team performance," *Appl. Ergonom.*, vol. 40, no. 6, pp. 997–1003, Nov. 2009.

[106] Ö. Çetin and E. Shriberg, "Overlap in meetings: ASR effects and analysis by dialog factors, speakers, and collection site," in *Proc. 3rd Int. Workshop Mach. Learn. Multimodal Interact.*, Bethesda, MD, USA, 2006, pp. 212–224.

[107] W. Dong, T. Kim, and A. Pentland, "A quantitative analysis of the collective creativity in playing 20-questions games," in *Proc. 7th ACM Conf. Creativity Cogn.*, Berkeley, CA, USA, 2009, pp. 365–366.

[108] P. Barthelmess, E. Kaiser, X. Huang, and D. Demirdjian, "Distributed pointing for multimodal collaboration over sketched diagrams," in *Proc. 7th Int. Conf. Multimodal Interfaces*, Torento, Italy, 2005, pp. 10–17.

[109] C. Chapman, L. Ramondt, and G. Smiley, "Strong community, deep learning: Exploring the link," *Innov. Educ. Teach. Int.*, vol. 42, no. 3, pp. 217–230, 2005.

[110] A. Soller, "Supporting social interaction in an intelligent collaborative learning system," *Int. J. Artif. Intell. Educ.*, vol. 12, pp. 40–62, 2001.

[111] B. Schneider and R. Pea, "Real-time mutual gaze perception enhances collaborative learning and collaboration quality," *Int. J. Comput.-Supported Collaborative Learn.*, vol. 8, no. 4, pp. 375–397, Oct. 2013.

[112] E. F. Barkley, K. P. Cross, and C. H. Major, *Collaborative Learning Techniques: A Handbook for College Faculty*. Hoboken, NJ, USA: Wiley, 2014.

[113] P. Isenberg, D. Fisher, M. R. Morris, K. Inkpen, and M. Czerwinski, "An exploratory study of co-located collaborative visual analytics around a tabletop display," in *Proc. IEEE Symp. Vis. Analytics Sci. & Technol.*, Salt Lake City, UT, USA, 2010, pp. 179–186.

[114] A. C. Evans, J. O. Wobbrock, and K. Davis, "Modeling collaboration patterns on an interactive tabletop in a classroom setting," in *Proc. 19th ACM Conf. Comput.-Supported Cooperative Work Social Comput.*, San Francisco, CA, USA, 2016, pp. 860–871.

[115] G. Fischer, "Symmetry of ignorance, social creativity, and meta-design" *Knowl. Based Syst.*, vol. 13, no. 7/8, pp. 527–537, Dec. 2000.

[116] G. Fischer and J. Ostwald, "Knowledge communication in design communities," in *Barriers and Biases in Computer-Mediated Knowledge Communication*. R. Bromme, F. W. Hesse, and H. Spada, Eds. Boston, MA, USA: Springer, 2005, pp. 213–242.

[117] H. Rittel, "Second-generation design methods," in *Developments in Design Methodology*, pp. 317–327, 1984.

[118] M. Cukurova, Q. Zhou, D. Spikol, and L. Landolfi, "Modelling collaborative problem-solving competence with transparent learning analytics: Is video data enough?" in *Proc. 10th Int. Conf. Learn. Analytics Knowl.*, Frankfurt, Germany, 2020, pp. 270–275.

[119] B. Schneider and R. Pea, "Toward collaboration sensing," *Int. J. Comput.-Supported Collaborative Learn.*, vol. 9, no. 4, pp. 371–395, Oct. 2014.

Sambit Praharaj received the bachelor's degree from VSSUT Burla, Odisha, India and the master's degree in computer science from TU Delft, Delft, The Netherlands. He is currently working toward the Ph.D. degree in multimodal learning analytics (EdTech) from the Open University of the Netherland, Heerlen, The Netherlands.

During his master thesis, he worked on using learning analytics on a large scale in the bachelors classroom at TU Delft to give insights about the engagement of the students for the teacher. He is looking into ways to detect the indicators of collaboration in a face-to-face setting using different sensors and then he wants to use these insights to give real-time feedback to facilitate collaboration. His research interests include co-located collaboration analytics using multimodal learning analytics.

Maren Scheffel received the Ph.D. degree in learning analytics from the Open University of the Netherlands, Heerlen, The Netherlands, in 2017.

She is currently a Professor of Educational Data Science with the Faculty of Educational Science, Ruhr University, Bochum, Germany. With a background in computational linguistics, she started to work in the field of technology-enhanced learning in 2009, and has since then been involved in the coordination and management (ROLE, LACE, CompetenSEA) as well as the research work (ROLE, LACE, SHEILA, CompetenSEA, SafePAT, DE-TEL) of several European projects. For her Ph.D., she developed the evaluation framework for learning analytics (EFLA). Her research interests include the connection of learning analytics and learning design as well as visualization techniques for dashboards.

Dr. Scheffel is the current President of the Society for Learning Analytics Research (SoLAR). She has served as the program Chair for European Conference on Technology Enhanced Learning (EC-TEL)'19, International Learning Analytics and Knowledge (LAK) Conference'20, EC-TEL'20, and LAK'21.

Hendrik Drachsler received the Ph.D. degree in computer science from the Open University of the Netherlands, Heerlen, The Netherlands, in 2009.

He is currently a Professor of Educational Technologies and Learning Analytics with the Leibniz Institute for Research and Information in Education (DIPF), Frankfurt, Germany; Goethe University, Germany; and the Open University of the Netherlands, Heerlen, The Netherlands. In the past, he has worked as a Principal Investigator and Scientific Coordinator with various national and EU projects (e.g., patient-project.eu, LinkedUp-project.eu, laceproject.eu, safepat.eu, sheilaproject.eu), and was responsible for a consortium budget of over 15 million euros in research grants and 3.2 million euros for his research group. His research interests include learning analytics, personalization technologies, recommender systems, educational data, mobile devices, and their applications in the fields of technology-enhanced learning and health 2.0.

Dr. Drachsler is an Elected Board Member of the Society of Learning Analytics Research (SoLAR), as well as a Member of the board of the European Association of Technology-Enhanced Learning (EATEL). He has regularly chaired international scientific conferences (e.g., International Learning Analytics and Knowledge (LAK)'16, European Conference on Technology Enhanced Learning (EC-TEL)'17, EC-TEL'18, and LAK'20).

Marcus Specht received the diploma in psychology and the Ph.D. degree in adaptive information technology from the University of Trier, Trier, Germany, in 1995 and 1998, respectively.

He is currently a Professor of Digital Education with the TU Delft, Delft, The Netherlands, and Director of the Leiden-Delft-Erasmus Center for Education and Learning. Since 2001 he has headed the Department of Mobile Knowledge with the Fraunhofer Institute for Applied Information Technology (FIT), Sankt Augustin, Germany. From 2005 to 2018, he was Professor of Learning Technologies at the Open University of the Netherlands, Heerlen, The Netherlands, and head of the Learning Innovation Lab. He is an Apple Distinguished Educator and was the President of the International Association of Mobile Learning, from 2013 to 2015. His research interests include mobile and contextualized learning technologies, and social and immersive media for learning.