

Review

How Can High-Frequency Sensors Capture Collaboration? A Review of the Empirical Links between Multimodal Metrics and Collaborative Constructs

Bertrand Schneider *, Gahyun Sung, Edwin Chng  and Stephanie Yang

Harvard Graduate School of Education, Harvard University, Cambridge, MA 02138, USA;
gsung@g.harvard.edu (G.S.); chng_weimingedwin@g.harvard.edu (E.C.); szhang@g.harvard.edu (S.Y.)

* Correspondence: bertrand_schneider@g.harvard.edu

Abstract: This paper reviews 74 empirical publications that used high-frequency data collection tools to capture facets of small collaborative groups—i.e., papers that conduct Multimodal Collaboration Analytics (MMCA) research. We selected papers published from 2010 to 2020 and extracted their key contributions. For the scope of this paper, we focus on: (1) the sensor-based metrics computed from multimodal data sources (e.g., speech, gaze, face, body, physiological, log data); (2) outcome measures, or operationalizations of collaborative constructs (e.g., group performance, conditions for effective collaboration); (3) the connections found by researchers between sensor-based metrics and outcomes; and (4) how theory was used to inform these connections. An added contribution is an interactive online visualization where researchers can explore collaborative sensor-based metrics, collaborative constructs, and how the two are connected. Based on our review, we highlight gaps in the literature and discuss opportunities for the field of MMCA, concluding with future work for this project.

Keywords: collaboration; multimodal; review



Citation: Schneider, B.; Sung, G.; Chng, E.; Yang, S. How Can High-Frequency Sensors Capture Collaboration? A Review of the Empirical Links between Multimodal Metrics and Collaborative Constructs. *Sensors* **2021**, *21*, 8185. <https://doi.org/10.3390/s21248185>

Academic Editors: Mimi Recker, Colin Hennessy Elliott and Quentin Biddy

Received: 30 October 2021
Accepted: 3 December 2021
Published: 8 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the last decades, there has been a growing recognition of the importance of skills beyond content knowledge that are transferable to the unknown careers of the future. With the increasing automation for routine and manual tasks, skills such as expert thinking and complex communication also rise in value [1]. In particular, the ability to collaborate within diverse teams is a foundational skill for tackling global problems that traverse several domains. Various organizations highlighted the need to foster collaborative skills in the future generation. The American Association of College and Universities identified teamwork skills as an essential intellectual and practical skill that students should develop when working towards a post-secondary degree [2]. In the K-12 domain, the Program for International Student Assessment (PISA) added a measure of Collaborative Problem Solving to its battery in 2015, motivated by an increase in the incorporation of learning teamwork tasks [3]. In summary, collaboration is increasingly recognized as a crucial skill for work and success in modern society.

Traditionally, collaborative processes were studied through qualitative observation and manual coding, which has the benefits of leveraging human intuition and heuristics, but is limited in scale. Recently, sophisticated data collection tools such as eye-trackers, motion sensors, and wearables have become affordable and reliable, which opens new doors for capturing large, multidimensional, and fine-grained information during collaboration. A sensor can collect data at a rate of 30 to 120 times per second (Hz) on dimensions including group physiological state, gestures, body postures, and speech, among others. The maturation of sensors and data mining techniques led researchers to envision new instruments and methods to measure and understand collaborative skills. As the field is

nascent and occasionally referred to through different terms, we use the term Multimodal Collaboration Analytics (MMCA) hereafter. The term draws from multimodal learning analytics [4], an overlapping field where interest extends to studying general learning processes with high-frequency data, and from the adjacent field of collaboration analytics [5], which includes research endeavors employing purely computer-generated data (e.g., log files).

Despite the increasing popularity, challenges remain in using sensor data in collaboration research. To move away from singular approaches and allow for collective knowledge building and the cross-pollination of ideas, we take note of synthesis efforts in more established fields: the taxonomy of socio-emotional learning (SEL) constructs by [6] in developmental psychology, the human genome project in genomics [7], or the human connectome project in neuroscience [8]. With the belief that MMCA research could benefit from similar organizing attempts, the current paper summarizes prior research into a map of how data, metrics and collaborative constructs are connected, and visualizes this mapping to make the state of the field accessible to other researchers.

Our work differentiates itself from prior reviews by scope, looking at MMCA papers in both education and social computing. While there is work in both fields that shares the same goal of detecting and supporting collaboration [9], there is little communication between the two fields, and reviews either focus on multimodal learning analytics (MMLA; e.g., [4,10]), or social computing research on collaboration (e.g., [11,12]). Additionally, prior reviews in MMLA or social computing presented models of research (e.g., [13,14]) or surveyed the types of collaborative outcomes that research focused on [10,11], or organized the types of data sources used [9,15]. To our knowledge, there have been no attempts to empirically map out the entire space between data, metric, outcome, and constructs, which is difficult to achieve not only due to the sheer range in metrics and outcomes used, but also due to the diverse ways that metrics and outcomes are connected [11,16]. Our empirical survey of theory use in MMCA research is an additional novel contribution that responds to the call for more theory integration in social analytics [9,17].

This paper contributes the following: (1) a model of research that conceptualizes how constructs relate to sensor data, (2) a taxonomy to categorize collaborative outcomes and metrics, and (3) a map of metric–outcome connections in prior research, supported by an interactive online visualization. We also provide (4) an empirical analysis of how theory is used to connect data to constructs.

2. Methodology

2.1. Data Collection

The scope of this review includes MMCA papers that used multimodal sensing technology to study social interactions. We selected a date range of 2010–2020, given the fact that sensor-based data collection tools were increasingly deployed in the last decade. We survey the literature to select a subset of papers that aimed to connect collaborative constructs (operationalized as “outcomes”) with sensor-based proxies (“metrics”). This excludes several subtypes of wider MMCA research, such as those A/B testing two-sensor-based applications for collaboration, qualitatively describing events in high-frequency data (e.g., quantitative ethnography [18]), measuring the impact of an intervention with high-frequency data, or analyzing only non-sensor based process data, e.g., application log data or speech transcripts. While we acknowledge the contributions that these diverse approaches made to the field, we set a narrow criterion for the overall goal of this paper: to survey the assorted sensor data metric–outcome connections in MMCA research.

To source papers eligible for this review, we followed the principles of the PRISMA [19] framework, which provides a set of evidence-based guidelines for a systematic review and meta-analysis. The identification phase consisted of two parallel searches. First, we conducted a database search using Google Scholar and Web of Science, with a set of pre-determined key words. Examples of keywords include: “collaboration”, “collaboration analytics”, “multimodal”, “sensor” and “group work”. For the second prong of our search,

we gathered peer-reviewed articles from 2010–2020 conferences and journals that provided a publication for MMCA research. These conferences and journals were: the International Conference on Computer Supported Collaborative Learning (CSCL) and associated journal (ijCSCL), Conference on Computer–Human Interaction (CHI), ACM Transactions on Computer Human Interaction (TOCHI), International Conference on Learning Analytics and Knowledge (LAK), journal of Learning Analytics (jLA), International Conference on Educational Data Mining (EDM) and associated journal (jEDM), ACM Conference on Computer-Supported Cooperative Work (CSCW) and associated journal (jCSCW), IEEE Transactions on Learning Technologies (IEEE TLT), ACM International Conference on Multimodal Interaction (ICMI), and International Conference on Artificial Intelligence in Education (AIED) and associated journal (jAIED). We filtered the aggregated articles to select papers that were (1) full-length; (2) assessed collaboration instances between human participants; (3) collected at least one type of sensor data; (4) transformed at least one data source into metrics using computational methods (c.f., papers that solely used manually coded metrics from video recordings); and (5) connected metrics to at least one observed outcome. A visual representation of the study selection process under the PRISMA framework [19] can be found in Figure 1.

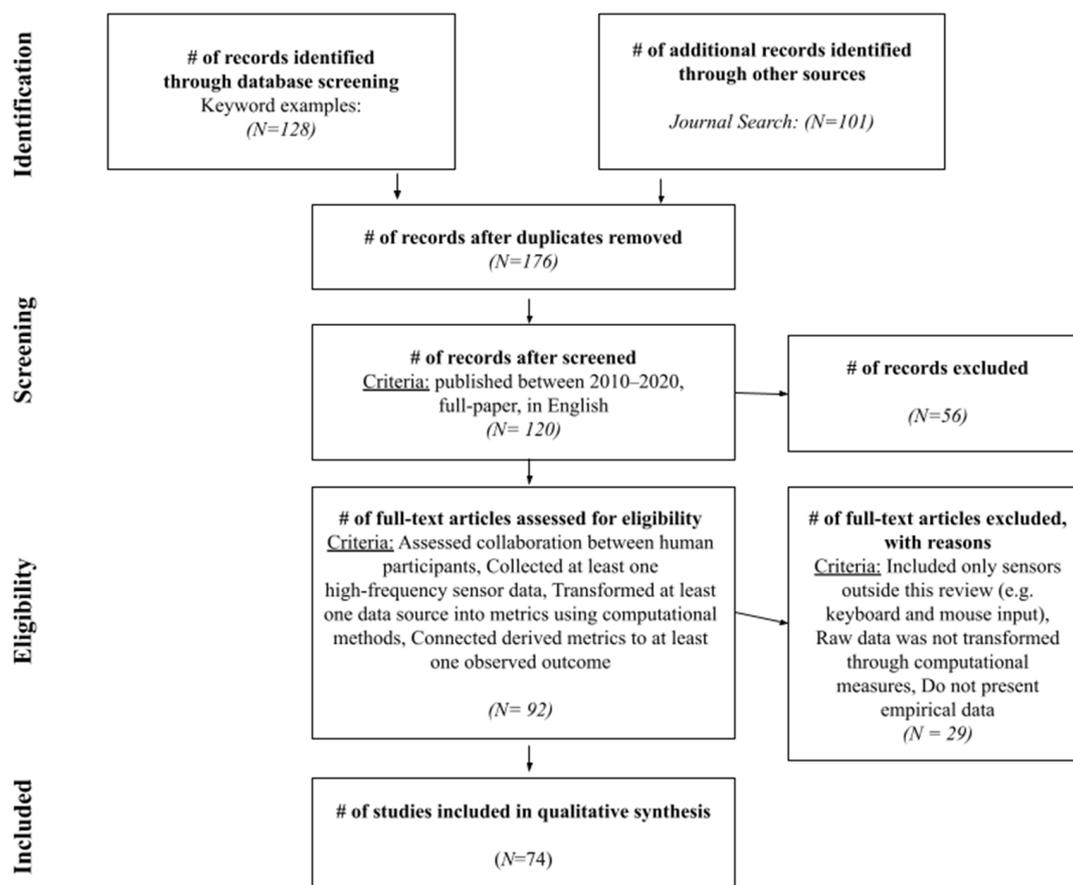


Figure 1. PRISMA diagram showing the flow of information through different phases of the review inclusion process.

2.2. Classification Framework and Process

The creation of our classification scheme and the mapping of prior findings in MMCA was guided by a simple framework for MMCA research connecting process data with different aspects of collaboration (Figure 2):

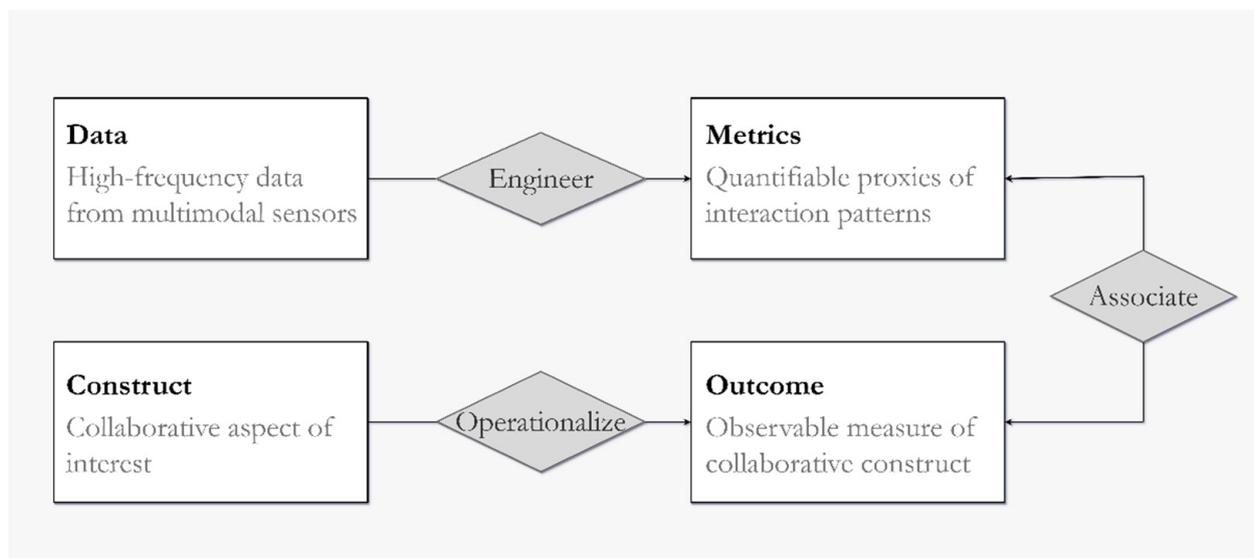


Figure 2. The MMCA data-construct framework.

This data-metric-outcome-construct framework does not apply to all MMCA research, but our paper selection criteria target the significant portion of papers that are relevant to this model. In papers that statistically link sensor data with collaborative outcomes, researchers often begin with a construct(s) of interest and a data source. At one end, **constructs** are operationalized into measurable **outcomes** of interest (e.g., the number of questions solved, collaboration quality scored by a rubric, or the perceived helpfulness of group members on a scale of 1–5). At the other end, **metrics** are generated from high-frequency **data** as a hypothesized indicator of the target outcome (e.g., joint visual attention, vocal pitch, or the prevalence of certain hand gestures). Outcomes are often measured through manual methods (e.g., self-report, survey, human coding), while metrics are usually calculated computationally. Of note is the fact that the term “outcome” has references to the role of outcome variables in statistics and may include the process and conditions of collaboration (e.g., symmetry of contribution, personality traits, prior expertise). We discuss the outcomes targeted in MMCA research more comprehensively in Section 4.2.

In one hypothetical example of MMCA research, data could be the x, y location of gaze locations captured by an eye tracker. One metric from this data could be the amount of time two people spent looking at the same on-screen location. Coordination, a collaborative construct, can be operationalized as a score given by researchers on coordination efficiency, and thus becomes an outcome.

We classified each paper in our dataset based on this MMCA data-construct framework. Specifically, for each paper, we identified the data, metrics, outcomes, and constructs from the raw text, then classified these into larger categories. When recording the raw data, researchers directly copied the language used by authors whenever available. A classification scheme was generated to classify the raw data into larger categories, and involved iterative cycles of discussion and paper coding involving all researchers. We sought internal consistency by repeatedly drawing on bottom-up categories for data and outcomes, classifying a significant (35%) proportion of papers and discussing discrepancies as a group. This process took place over a total of 20 iterations and yielded if-then rules for classification, as well as a look-up table consisting of the raw metrics or raw outcomes found in the papers (such as speech time or stress) and the respective larger category code (such as speech participation and affective state). To achieve the reliability and consistency of the synthesis process recommended by PRISMA, the four authors participated in parallel and active communication during this phase.

The final classification scheme consisted of two primary spreadsheets. The first captured general information about the paper including paper title, publication year,

authors, author affiliations, research questions, and task domain. The second spreadsheet focused on the data, metric, and outcome link assessed in the paper. Namely, information on the types and brands of instruments used to collect data, raw data source, raw metric, larger metric category, raw outcome, larger outcome category, outcome instrument, analysis methods, and significance of metric–outcome links were logged onto the spreadsheet.

2.3. Theoretical Framework for Classification

While the raw metrics and outcomes were derived from a bottom-up approach closely mirroring the paper text, the outcomes were further classified into larger categories identified through Dillenbourg's [20] seminal theoretical framework for assessing collaboration. In this work, Dillenbourg argues that, while generally the case, collaborative interactions cannot solely be assessed by their effects on learning or task performance. For one, collaboration takes place in a variety of contexts and through a variety of interactions. Exclusively addressing the outcomes places a "black box" around these important variables, which ignores the underlying mechanisms of collaboration.

Instead, Dillenbourg encourages researchers to study the conditions and processes of collaboration in addition to its effects. As an example, the conditions of collaboration detail the composition of the group, which include factors such as age, personality, or the heterogeneity of these variables. The processes of collaboration describe "the mechanisms which account for knowledge acquisition through collaboration". These mechanisms primarily draw on socio-constructive and socio-cultural approaches to learning, and include self-explanation, sharing cognitive load and mutual regulation. Finally, the effects of collaboration identify what group members can do (either individually or together) after the collaborative session. This is most commonly assessed through learning gains or task performance, but is also expanded to include conceptual change or increased self-regulation. Over the years, scholars seem to have converged on similar frameworks for assessing collaboration. For example, Ref. [21] proposed a model that categorized the input, processes and outputs of a group interaction, and [22] proposed a similar model identifying the antecedents, processes, and outcomes of collaboration. Importantly, all of these frameworks highlight the necessity of looking inside the "black box" of the collaborative process.

In the current paper, we adapted and renamed this model to classify the outcomes that multimodal metrics were designed to capture within our sample. Specifically, we categorized these into the **conditions** under which collaboration can succeed, the specific interactions that occur during the collaborative **process**, and the end **product**.

3. Research Questions for the Review

Based on the MMCA data-construct framework (Figure 2), our research questions (RQs) can be summarized as follows:

- RQ1: What sensor-based metrics have been used to capture collaborative processes?
- RQ2: What outcome measures have been used to validate sensor-based metrics?
- RQ3: Which connections between metrics and outcomes have been successful?
- RQ4: How was theory used to inform the connections between metrics and outcomes?

The results section follows this order. In Sections 4.1 and 4.2, we present a taxonomy of sensor-based metrics and collaborative outcomes. In Section 4.3, we summarize how different outcomes have so far been connected to metrics using our taxonomy. Here, we introduce an openly accessible web-based visualization where the entirety of data-construct connections in our corpora can be interactively explored [23]. Our plans are to keep updating the database so that it may serve as a central repository for MMCA research. In Section 4.4, we investigate the role of theory in the data-construct research model, summarizing how theory integration has so far been carried out in research.

4. Results

4.1. What Sensor-Based Metrics Have Been Used to Capture Collaborative Processes (RQ1)?

Our first research question addresses the type of metrics that have so far been used in MMCA research. Metrics represent a layer of abstraction created by processing raw data into measurable features, hypothesized to be proxies of different collaborative outcomes. We organized metrics into six larger categories, which are summarized in Table 1. The six larger categories reflect primary data modalities that are closely tied to data sources (e.g., ‘verbal’, ‘physiological’). Conversely, the smaller categories were generated using a ground-up approach through iterative groupings of similar metrics found in the papers (e.g., ‘vocals pitch’ and ‘energy’ into an ‘audio features’ category). The table also shows examples of each metric category, the sensors and computation methods used, and papers that use this metric type.

There are several observations we can make from Table 1. First, the table reflects the richness of the data sources and metrics used to capture collaboration. We found 457 sensor-based measures in the 74 papers analyzed in our corpus. Metrics ranged from indices of physiological synchrony, linguistic features, joint visual attention, body movement/synchronization, facial expressions, to context-specific actions captured by log files. This variety indicates that MMCA is a burgeoning field that has brought forth a rich set of metrics to investigate collaboration.

Second, we observed that there was a preference for certain types of metrics in our corpora. Verbal data were used in 35 papers (close to half of our corpus, or 47% of papers) and gaze data were used in 28 papers, or 37%. More specifically, visual attention and speech-related features were frequently utilized, while touch, eye motion, heart rate, brain activity and eye physiology metrics were less frequently explored. This may have been due to several reasons. The cost of some sensors can be inhibitive, as in the case of research-grade wearable EEG devices, while ease of analysis can also play a part—for instance, established front-to-end software exists for extracting acoustic features from speech. The necessity of a human coding step, as is often the case in semantic speech features, or the invasiveness of sensors may also have played a role. Finally, some data sources have a larger perceived distance with collaborative constructs—for instance, more meaning-making steps are required to connect heart rates to complex human behaviors compared to hand gestures.

Third, we observed that papers tended to analyze only one modality. One strength of MMCA is that collaboration can be understood through the combination of multimodal features (i.e., the complex interplay of subtle body postures, tone of voice, gaze direction, words chosen, physiological states, and other behaviors in collaborative interactions). However, the majority of the papers only analyzed one modality (55%), a few combined two modalities (33%), and only a small fraction used more than two modalities (12%). This indicates that there are obstacles to conducting truly multimodal analysis, which we elaborate further on in the discussion section.

Fourth, when describing how these metrics were computed from sensor data (5th column in Table 1), we categorized whether they were computed using a procedure created by the researcher (e.g., by generating them from the raw data, or by using a combination of self-designed computations) or if they could be generated solely by an existing tool or procedure (e.g., an eye-tracking software). We found that 405 metrics (87%) used the former. This is not surprising, given that we only studied academic sources whose main contribution is a new metric set. However, constantly inventing new features can introduce issues when trying to generalize results. In several instances, we found that metrics shared the same name but were computed in different ways; or metrics had different names but were computed in the same way. For instance, Ref. [24] discusses interruptions, while [25] discusses overlap cues, yet both papers calculate the number of events when a speaker is interrupted by another.

Table 1. Taxonomy of metrics.

Larger Category	Lower-Level Category	Metric Examples	Sensors Used	Computation Methods	References
Verbal (N = 35)	Speech Participation (N = 18)	Speech time, Silence duration, Verbal dominance, Speaking turns, Interruptions, Speech frequency, Verbal participation symmetry among group	Microphone, Microcone, Video camera	Arithmetic calculation	[24–41]
	Verbal Content (N = 10)	Dialogue acts, Sequences of verbal utterances, Linguistic features		Arithmetic calculation, Qualitative coding, Supervised machine learning	[27,32,42–49]
	Audio Features (N = 17)	Pitch, Energy, Speaking rate, Acoustic features, Mean audio level, Prosodic and tone features		Arithmetic calculation, Supervised machine learning, OpenSMILE	[28,32,33,39,41,42,48,50–59]
Physiological (N = 13)	Electrodermal Activity (EDA) (N = 11)	EDA peak detection, Galvanic skin response, Physiological synchrony	Varioport 16-bit digital skin conductance amplifier, Smart wristband, Electroencephalogram, Wearable sensor	Arithmetic calculation, Correlation, Cross-recurrence quantification analysis	[52,57,60–68]
	Heart Rate (N = 2)	Heart rate		Arithmetic calculation	[63,68]
	Neural Activity (N = 1)	Brain synchrony		Arithmetic calculation	[69]
	Mixed (e.g., EDA + Heart Rate) (N = 2)	Physiological linkage, Physiological simultaneous arousal, Physiological concordance index		Arithmetic calculation	[34,61]
Gaze (N = 28)	Gaze/Eye Direction (N = 28)	Gaze fixations, Gaze area of interest, Attention center, Count of faces looking at screen, Fraction of convergent gaze, Gaze similarity, Joint visual attention	Eye tracker, Video camera, Microsoft Kinect, Optical see-through head-mounted display	Arithmetic calculation, regression, Matrix Calculation, BeGaze, Maximum a posteriori estimation, Supervised and unsupervised machine learning	[24–26,28,30,40,44,49,55,57,70–87]
	Eye Motion (N = 3)	Gaze transitions, Gaze saccades		Arithmetic calculation, Eye-tracking softwares (e.g., BeGaze)	[70,72,75]
	Eye Physiology (N = 1)	Pupil size		Arithmetic calculation	[84]

Table 1. Cont.

Larger Category	Lower-Level Category	Metric Examples	Sensors Used	Computation Methods	References
Head (N = 11)	Facial Expression (N = 6)	Facial action units, Facial expression features, Smiling synchrony	Video camera, Microsoft Kinect	OpenFace	[43,48,51,54,64,67]
	Head Motion (N = 5)	Head movement		Arithmetic calculation	[28,32,38,58,59]
Body (N = 21)	Hand Motion (N = 8)	Gesture, Wrist movement, Total manual gestures per second, Iconic gestures per second, Deictic gestures per second, Distance between hands, Hand motion speed, Touch patterns	Video camera, Webcam, Microsoft Kinect	Arithmetic calculation, Qualitative coding, Unsupervised machine learning	[15,26,38,43,58,79,88,89]
	Gross Body Motion (N = 12)	Total Movement, Type of movement, Body synchronization, Physical synchrony, Joint movement, Joint angle		Arithmetic calculation, OpenPose, Supervised and unsupervised machine learning	[27,28,31,35,38,52,57,90–94]
	Location (N = 6)	Distance from the center of the table, Body distance, Dyad proximity		Arithmetic calculation, OpenTLD	[27,28,59,90,91,95]
Activity Log (N = 12)	Writing Action (N = 1)	Total number of pen strokes, Average stroke time, Average stroke path length, Average stroke displacement, Average stroke pressure	Digital pen, Touch screen, Interactive tabletop, Arduino IDE, Video camera, Log files	Arithmetic calculation	[27]
	Touch (N = 3)	Total number of touch actions, Symmetry of touch actions among group		Arithmetic calculation, Qualitative coding	[36,45,96]
	Task-Related (N = 9)	Object manipulation, Calculator use, Times mathematical terms were mentioned, Times commands were pronounced, Amount of exploration, Arduino measure of complexity, Arduino active hardware blocks, Arduino active software blocks		Arithmetic calculation, Qualitative coding, OpenCV, Micro-controller logs	[27,37,39,56,58,75,90,92,97]

4.2. What Outcome Measures Have Been Used to Validate Sensor-Based Metrics (RQ2)?

Our second research goal was to catalog the types of outcome measures used by researchers to validate their sensor-based metrics. To organize the types of collaborative outcomes utilized in MMCA research, we referred to the seminal work by Dillenbourg [20], as mentioned in Section 2.3. After grouping our raw outcomes by similarity, we observed a parallel distinction based on these frameworks. That is, research in our corpora aimed to connect data-derived metrics with either the conditions within a collaborative group (e.g., individual personalities, leadership); the processes integral to, or indicative of, successful collaboration (e.g., group rapport, engagement, mutual understanding); or the products of collaboration (e.g., learning gains, performance). The results of our analysis for each outcome measure are displayed in Table 2. The first column shows the larger category of the outcome, as classified by Dillenbourg's framework [20]. These overarching categories contain several lower-level outcome categories shown in the second column, which were derived using a bottom-up approach to classifying our data.

The first observation we made from Table 2 is that performance and learning, the 'product' measures of collaboration, were most often assessed (in 53% of the papers, not including repeat papers across outcomes). Among outcomes targeting the 'process' of collaboration, coordination and communication outcomes, such as information pooling, time management, and mutual participation, were studied in a number of papers. This is likely because these factors are often linked with highly available data sources such as audio and video, and often use a validated coding scheme (see [98]). Affective states were not widely studied, possibly because they were primarily individual experiences, rather than a consciously targeted outcome in collaboration (For a review of the active and ongoing multimodal research on individual affective states during learning, see [99]).

Second, Table 2 also illustrates a significant diversity in task context and measurement methods both across and within outcome measures. Column 4 indicates that the majority of the papers used tasks from science, technology, engineering and math (STEM) domains, such as programming, neuroscience, or robotics. Other domains included gameplay or group decision-making tasks. Whether the field's understanding of multi-modal analytics are transferable to less frequently studied domains, such as subjects in the humanities or creative endeavors, is an open area of research.

Third, column 6 reveals the common questionnaires and coding schemes consistently used by researchers, notably the Meier, Spada, Rummel framework [98], to assess communication and coordination, the NASA Task Load Index [100] to measure cognitive load, and the NEO-FFI questionnaire [101] to assess personality. However, papers employed a large variety of measures for affective state, interpersonal relationships or perception, and individual cognitive processes. In these outcome categories, there was a high level of variability, even among validated measurement tools.

Finally, a related observation is that process outcomes of collaboration tend to have the highest levels of variability in how they are measured. This could suggest that more theoretical discussion among the field is needed to determine (1) what constitutes a process measure and (2) to discover validated measures to assess this. We further elaborate on these points in the discussion section.

Table 2. Taxonomy of Outcomes.

Larger Category	Lower-Level Category	Outcome Examples (Data Directly from Papers)	Domain	Measurement Methods	Questionnaires or Coding Schemes (Ratio of Validated to Generated)	Reference
Product (N = 37)	Performance (N = 24)	Completion time, Success of task, Quality of task, Correctness	Pair programming, problem-solving, instruction giving, math, physics, engineering/design	Automated coding, human coding, self-report	Questionnaire (0:2): Researcher Generated	[25,27,29,32,42,51,57,58,62,64,65,69,71,74,76–78,80,83,86,91,92,95,97]
	Learning (N = 19)	Normalized learning gain, dual learning gain	Neuroscience, programming, engineering/design, nutrition	Pre-post test		[15,26,34,43,44,47,56,59,61,64,70,72–74,77,80,83,84,90]
Process (N = 25)	Communication (N = 22)	Conversational efficiency, agreement, mutual understanding, dialogue management, verbal participation	Pair programming, neuroscience, problem-solving, math, nutrition, engineering/design, gaming, naturalistic	Automated coding, human coding, self-report	Coding Scheme (15:5): Meier, Spada, Rummel [98], Researcher Generated Questionnaire (1:0): Meier, Spada, Rummel [98]	[31,34,36,37,39,45,47,49,55,56,61,63,72–74,81–83,86,87,89,91]
	Coordination (N = 28)	Information pooling, consensus reaching, socially shared regulation, synchrony, task division, time management, technical coordination, routine choice	Pair programming, neuroscience, problem-solving, math, physics, nutrition, engineering/design	Human coding, self-report	Coding Scheme (17:6): Meier, Spada, Rummel [98], Researcher Generated Questionnaire (1:1): Meier, Spada, Rummel [98], Researcher Generated	[31,34,36,38,39,41,45,47,51,55,56,58,61,63,66,67,69,72–74,78,79,81–83,86,91,96]
	Affective state (N = 6)	Stress, confidence, emotional state, empathy, frustration, perceived valence and arousal	Programming, problem-solving, physics, engineering/design, gaming	Self-report	Questionnaire (3:3): Social presence in Gaming [102], NASA Task Load Index [100], Ainley, Corrigan and Richardson [103], Hadwin & Webster [104], Researcher Generated	[40,48,56,62,92]

Table 2. Cont.

Larger Category	Lower-Level Category	Outcome Examples (Data Directly from Papers)	Domain	Measurement Methods	Questionnaires or Coding Schemes (Ratio of Validated to Generated)	Reference
Process (N = 25)	Interpersonal relationship/perception (N = 14)	Self-report quality, self-esteem in work teams, collaborative will, perception of peer (helpfulness, understanding, clarity), laughter, social presence, rapport level, team cohesion	Pair programming, neuroscience, problem-solving, instruction giving, math, physics, nutrition, engineering/design, naturalistic	Human coding, self-report	Coding Scheme (0:3): Researcher Generated Questionnaire (6:6): Sanchez-Cortes, Aran, Mast and Gatica-Perez [30], MSLQ [105], Manson et al. [106], Researcher Generated	[24,25,30,48,50–53,64,66,68,92,94,95]
	Individual Cognitive Processes (N = 11)	Mental effort, cognitive load, workload, engagement, task difficulty	Programming, problem-solving, physics, engineering/design, gaming, naturalistic	Human coding, self-report	Coding Scheme (0:2): Researcher Generated Questionnaire (7:3): Social Presence in Gaming [102], User Engagement Survey [107], NASA Task Load Index [100], Paas [108], Tapola, Veermans and Niemivirta [109], Efklides, Papadaki, Papantoniou, and Kiosseoglou [110], Researcher Generated	[43,59,60,62,65,67,68,79,86,89,96]
Condition (N = 15)	Group composition (N = 15)	Expertise, personality, assigned leadership, emergent leadership	Problem-solving, math, gaming	Automated coding, Assigned role, self-report	Coding Scheme (0:4): Researcher Generated Questionnaire (8:0): NEO_FFI [101], SYMLOG [111], GLIS [112], Manson et al. [106]	[22,24–26,30,32,37,43,50,51,72,82,85,87,91]

Notes: In the fourth column titled “Domain,” “naturalistic” domains categorized studies that assessed ecological classroom settings without explicit specification of learning domain. Under column five, “Automated coding” indicates measurements that were either automatically calculated by a computer or measures that were formulaically generated. Under the sixth column, the values in parentheses indicate the ratio of previously validated measures to measures generated by the researchers for the study. Coding schemes and questionnaires generated by researchers do not necessarily indicate that these measures have no literature backing, nor that they have not been used in other studies.

4.3. Which Connections between Sensor-Based Metrics and Collaborative Outcomes Have Been Successful (RQ3)?

In this section, we summarize the connections made between metrics and outcomes in MMCA research as per our third research question. As the main focus of this paper, these connections are investigated through different perspectives. First, we describe the types of connections (i.e., analysis methods) used in our corpus (Section 4.3.1). We then summarize the significant and non-significant connections made per each lower-level outcome category, quantitatively (Section 4.3.2) and qualitatively (Section 4.3.3). Finally, we introduce an openly accessible and interactive online visualization where data-construct connections can be explored in detail (Section 4.3.4).

4.3.1. Types of Metric–Outcome Connections

One of the most frequent types of approaches found to connect metrics to outcomes is testing the significance of individual links, often using correlation or regression analysis methods. For instance, [71] runs a series of correlation analyses between different gaze metrics and post test scores to report strength and significance. This type of approach was widely found in our data, and has the benefit of being able to quantify the power of individual metrics compared to a whole-model approach.

Another procedurally similar method builds the narrative backward from the outcome, investigating if different people or work attributes translate to divergent patterns in the data-derived metrics. *T*-tests, chi-squared tests and ANOVAs are frequent methodological choices in this approach. One example is [45], where the researchers divide less versus more collaborative groups using researcher codes, then test whether there are differences in multiple audio- and log-data-derived metrics between the two groups.

A distinctive type of approach aims to create functional prediction systems for an outcome of interest. Machine learning methods are often used, such as regression, neural networks, or support vector machines. The research by [58] generated 18 features from camera, Kinect, audio, and activity logs to build a neural network classifier, testing different metric calculation methods and combinations.

While the focus of our review is on the quantitative associations made between metrics derived from high-frequency data and collaborative outcomes, qualitative examinations between metrics and outcomes can serve an important aim for advancing the field. One such type of analytic strategy is to observe and then describe what happens during patterns of interest in the data. Qualitative methods were used to draw observations from video recordings, dialogue, or interview transcripts to shed light on metric measures, and to determine the meaning of a metric–outcome connection. For example, Ref. [26] supplements quantitative analyses by studying the dialogue of dyads who displayed high levels of joint visual attention (JVA) during a task to examine the potential causes of JVA, and hypothesizes their impact on the learning outcomes. Papers such as [64] attempt to understand electrodermal activity metrics by observing events that take place during simultaneous arousal.

These approaches are not mutually exclusive. Papers often test multiple links, or adopt several approaches to test one set of links. Papers such as [24] or [56] for instance report the strength of correlations between individual metrics and outcomes before combining metrics for outcome prediction. Combining these approaches allows for the testing of uniquely created metrics while also linking metrics multidimensionally and practically with a learning outcome of interest.

4.3.2. Quantitative Trends in the Metric–Outcome Connections

To find trends in the metric–outcome connections, we investigated which metric–outcome connections were the most and least successful. We adopt a simple definition of success, determining it by whether a particular metric–outcome link was reported to be statistically significant in a paper. The results of this survey are summarized in Figure 3.

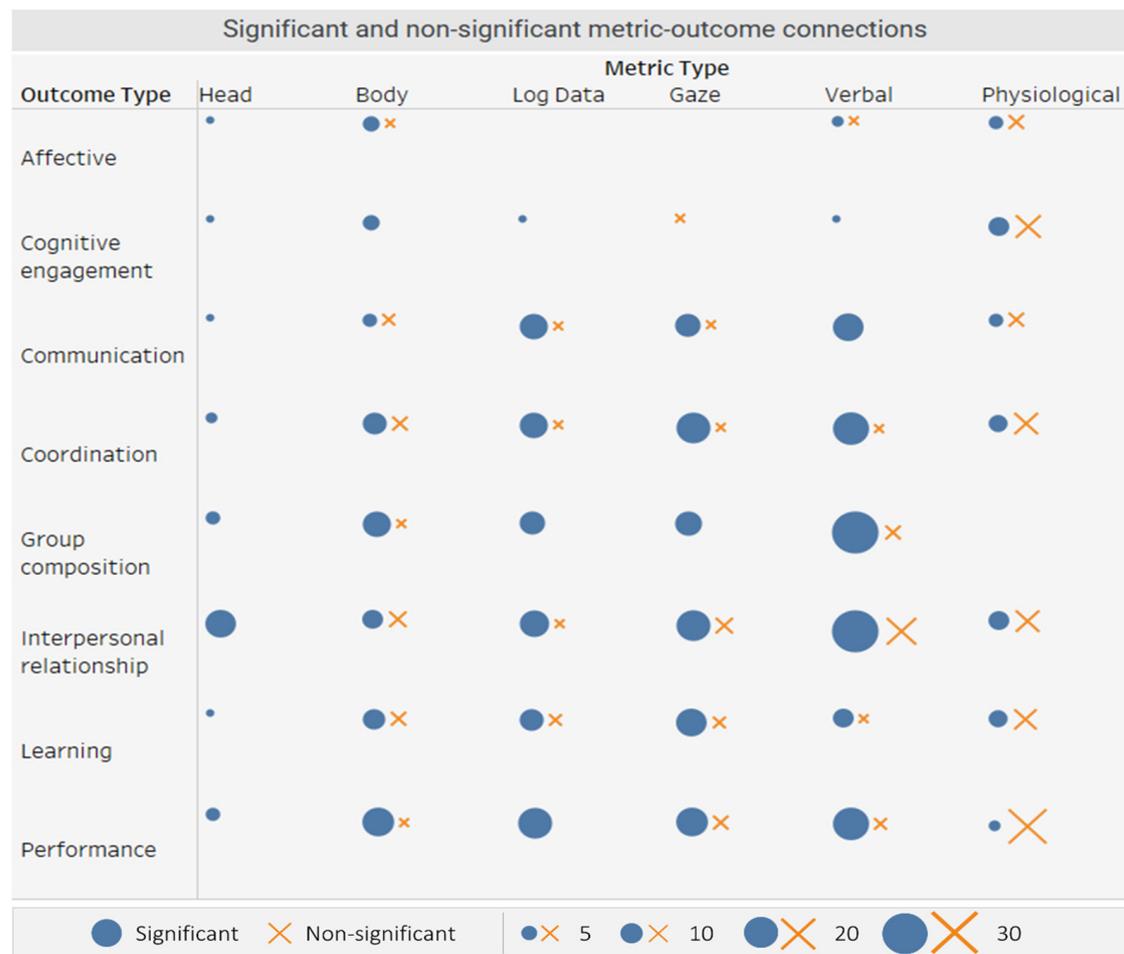


Figure 3. Success of metric–outcome connections. Circles and crosses represent the number of significant and non-significant connections, respectively.

Figure 3 shows the count of connections made between a particular outcome and metric, with the circles representing success (i.e., statistically significant connection) and the x's lack of evidence for a connection (i.e., statistically non-significant). The sizes of the markers signify the count—for example, the connection between performance and verbal metrics was found to be significant in eighteen cases, and non-significant in two cases. We note here that the statistical significance of a connection does not imply a strong connection between a metric and outcome. A statistically significant connection could be weak, but discovered due to a large sample size, choice covariates, etc., while the opposite may occur for statistically non-significant connections. In other words, our visualizations show the frequency of connections, and not their strength.

The results showed that metrics had different rates of success. Head metrics, while less frequently used, were found to have meaningful associations across all types of collaborative outcomes, in particular for interpersonal relationships and perceptions. Verbal metrics were most frequently applied among the different metric categories, and the majority were successful across all outcome types. The opposite is true for physiological metrics; while often employed, metrics failed to be associated with outcomes more often than they were successful. However, we note that the success rates of a metric type do not imply a need to move away or towards a particular metric. For instance, EDA signals, while often less straightforward for understanding natural collaborative situations, offer the unique benefits of being minimally invasive; privacy-preserving, based on universal human biology; and usable in nearly all physical environments. Instead, we take this to be a map summarizing the level of progress made in finding stable connections for

different metric–outcome combinations. For pairs with low frequency or success rates, more development and explorations of metrics are needed; for those with high frequency or success rates, fruitful endeavors might be to refine previously developed metrics or find new theory-based metric combinations and analysis methods. This has organically taken place, for example, in the study of emergent leadership using gaze metrics, where papers were built on a common understanding that a socially dominant person receives more attention from peers. Therefore, the different ways of quantifying this received attention can be tested by refining metrics to be more accurate, automated, and widely applicable to different contexts (see [30,55,85] for examples).

4.3.3. Qualitative Findings for Metric–Outcome Connections

We discuss the significant findings for each outcome category of collaboration below. We provide an overview of successful metric–outcome links per outcome type, with the goal of informing future research directions.

Product: Performance

The most common way of assessing the success of a collaborative session is by how well the team worked together to achieve a given task. The ‘product’ category of outcomes is measured to find multimodal indicators of success in process data. By making this connection, it is assumed that multimodal metrics measure one or many from a wide range of adaptive behaviors characteristic of good collaboration, such as group members being in ‘sync,’ or all members of the group being engaged in the task. As the target mechanism varies widely, the metrics explored for this outcome category also span across the entire range of metric types. Nonetheless, connecting process metrics with the products of collaboration can help understand what the process of a successful collaboration looks like, or yield practical algorithms for predicting performance or detecting underperforming groups.

Both group and individual performances were studied as outcomes in our corpora, although individual outcomes were targeted more than 80% of the time. Individual task-dependent log metrics on how someone approached a task (e.g., variety of code blocks [58], calculator use count [27]) had significant repeated links to performance. Body metrics (e.g., amount of movement [27,52,91,92], distance between members [58,59]) achieved a high rate of success for hands-on tasks while being generalizable across different contexts. Combinations of verbal (e.g., speech participation [25,27,29,30]) and gaze metrics (e.g., area of interest [78], joint visual attention [26,70,77,83,86]) were usually successfully linked to increased performance, although one paper found low-level gaze features to be unsuccessful predictors [72]. Physiological metrics, on the other hand, were largely unsuccessful and there was notable divergence in the success levels of physiological synchrony [62,63,65,66].

Product: Learning Outcomes

Another “product” of successful collaboration is increased learning for all group members involved in the collaborative process. Because successful performance and successful learning share comparable mechanisms, similar types of metrics were found to be associated with both outcomes. Gaze metrics were most frequently connected with learning in particular, with joint visual attention positively correlating with learning gains in both in-person and remote settings [44,71,73,74,77,80,82,84]. Researchers typically view gaze as a proxy for students’ focus of attention and use it to determine if students are paying sufficient attention to the critical elements of the learning material. Sophisticated measures of verbal content, such as linguistic coordination [47,49] and verbal coherence [47,49] were predictive of learning outcomes, while simple measures of speech, such as the individual length of utterances, were not correlated with learning gains [47]. In a similar vein, we would expect students who are more proficient learners in classrooms to behave differently from others who were not. Individual clustered body movements, such as hand, wrist movement [15] and active posture [90], but not group-level metrics such as body

synchronization [90], were predictive of individual learning gains. Lastly, based on the papers we reviewed, internal physiological measures such as EDA synchrony were more successful for predicting learning than performance [34,57,60,62,65], although the results were still mixed. Apart from EDA, we found that overall learning tended to have more mixed success with metrics. In other words, there were fewer metrics that were established as effective for predicting learning outcomes. One explanation is that successful metrics are task- or domain-dependent: learning tests in our corpus targeted different types of learning (e.g., memorization, conceptual understanding, transfer questions) in widely different domains (e.g., programming, physics, math, history, language).

Process: Communication

Integral to the success of any kind of collaborative activity is effective communication. Successful collaboration occurs in cases where members actively regulate their communication. For example, [113] finds that actively engaging in explaining and listening during group work leads to better retention of learning.

Communication was most often studied as a sub-domain of ‘collaborative quality,’ in particular through the Meier, Spada and Rummel coding scheme, where the quality of communication is judged by how well participants sustain mutual understanding, and perform dialogue management, e.g., balanced turn-taking. Two papers [56,64] created labels for the communication-related state of collaboration at a given time, while other outcomes included balanced verbal participation [45,50], joint visual attention [47], and speech cohesion [57]. The study by [87] was unique in its approach of using lower-level outcomes, investigating connections between pair gaze patterns (metric) and use of referential forms (outcome; e.g., a deictic pronoun such as ‘this’ or ‘those’).

Speech is the primary data type that holds information about dialogue management, the first component of communication as per the Meier, Spada and Rummel coding scheme. As such, verbal metrics were the most frequently employed metric for communication quality, with speech time [50] and symmetry of speech among the group members [33,34] found to be significant indicators. On the other hand, to gauge how well collaborators sustain a mutual understanding, the second component of communication, JVA, was frequently investigated as a metric in a number of papers. JVA was particularly found to be correlated with students’ tendencies to reach a consensus and manage dialogue [74], similar to [73] that found associations with sustaining mutual understanding, reaching a consensus, and pooling information. EDA metrics showed mixed results for link to communication measures [34,57,62,64], which may indicate that EDA metrics remain internal to the group participants and have less observable connections with communication amongst group members. ML models predicting communication outcomes most often used an extensive set of audio and/or verbal metrics, although actions recorded in log data were combined with verbal metrics in [36,37,39,56]. Network analysis-based joint gaze metrics [81] and physical synchrony metrics were used independently in [57] to some success.

Process: Coordination

Given the complex and interdependent nature of many collaborative activities, group members must coordinate their efforts in order to succeed [98]. In successful groups, individuals must contribute useful information, process it together, divide tasks, allocate enough time for subtasks, and coordinate parallel and joint activities. Coordination is achieved not only through oral contributions, but also through subtle non-verbal interactions: group members need to coordinate their attention by jointly looking at the same areas of interest and coordinate their physical actions through various body postures and gestures. In our coding, targeted activities included the coordination of communicative content (information pooling, reaching a consensus) and processes (task division, time management, technical coordination). Coordination was also often studied as a sub-domain of ‘collaborative quality’ through the Meier, Spada and Rummel coding scheme.

The verbal participation of group members was found to be consistently connected to coordination [36,41,64,83]. This included measures of verbal coherence, such as the similarity of within-group speech [47,49]; equal contribution to the discussion [45]; and various Coh-matrix metrics, such as readability, intentional verbs or using the active voice [57]. Speech features (e.g., loudness, pitch, jitter) were also heavily used in our corpus, but only in conjunction with supervised machine learning approaches [39,41,56]. Several non-verbal metrics were also significantly related to coordination. For example, a large body of evidence indicates that successful coordination is correlated with a higher occurrence of joint visual attention [70,73,78], especially for task division and reaching a consensus [74]. Gaze awareness tools (i.e., being able to see the gaze of a partner in real time) was found to significantly improve information pooling, time management, reaching a consensus [73], and verbal coherence compared to a control group [49]. This indicates that helping group members achieve joint visual attention might be beneficial to their coordination. Finally, several papers found that successful coordination is also associated with increased physiological synchrony [34,68]. Qualitative evidence suggests that episodes of high synchrony might be associated with a common response to external events, and low synchrony with joint work or moments of confusion [62]. What is more, the number of back-and-forth transitions between these two states was strongly correlated with dialogue management, and moderately with reaching consensus or information pooling [62]. This could indicate that successful coordination involves rapid transitions between states of (dis)synchrony, a finding that was replicated for measures of joint visual attention [82].

Process: Affective State

An additional process measure integral to collaboration relates to the affective state of individual group members. Emotions play an important role in the interactions and performance of a collaborative team. Notably, individuals can perpetuate both positive and negative emotions to their team members [114], and the forms of non-verbal communication triggered by certain emotional states impact team decisions and interactions [115]. Within our sample, outcomes were mostly individual affective states similar to those found in the wider field, such as frustration, boredom, interest, and valence (from positive to negative).

Due to the biological responses connected with affective states, physiological metrics, and, in particular, physiological linkage were frequently studied in collaboration studies. The study by [60] found that the physiological linkage between participants during a gameplay scenario was correlated with feelings of empathy, suggesting support for “emotional contagion”, which posits that individuals simulate similar biological responses in order to understand the other person. This may be irrespective of the type of emotion, since [66] found that physiological synchrony was not associated with emotional valence. Both [43] and [95] assessed feelings of frustration using body movement data during learning. While [43] found that students expressed a higher frustration after receiving help from a tutor, based on movement data, Ref. [95] showed that the transitional probability of working with an instructor, compared to working individually, was negatively associated with frustration. The study by [43] postulates that the increase in frustration could be due to a lingering misconception, while [95] interpret their data to mean that instructors are often effective in helping their students with problems. These discrepancies highlight a future direction for using movement data to understand emotion, using observable contextual information to form hypotheses that could then be verified in subsequent research. Finally, using a combination of speech, body movement and skin response metrics across time, Ref. [52] found that lower levels of team regularity corresponded to a more positive valence. The authors suggest that this indicates that teams with more repetitive patterns felt that the collaboration was more unpleasant.

Process: Interpersonal Relationship/Perception

During group interactions, interpersonal relationships and perceptions can play a mediating role in the products of collaboration as well as other process measures. For example, rapport between individuals is linked with higher learning gains [116]. On the other hand, group conflict (both task-related and social) was shown to negatively affect group performance [117]. Understanding individual and interaction-level metrics that contribute to interpersonal perceptions can help researchers develop interventions to foster mutual rapport and support collaboration.

Examples of this outcome within our sample include group members' perceptions of the contributions and helpfulness of others, sense of rapport, and perceptions of group cohesion. Verbal metrics were studied in the majority of interpersonal perception papers. Verbal dominance and speech length was significantly correlated with perceived contribution [25]; however, individual features of speech such as speech rate and voice features were not significantly linked with interpersonal perception outcomes such as helpfulness and understanding [48,52]. It may be that verbal cues are too granular to impact high-level interpersonal relationships at the group level, or, as the authors of [52] propose, participants may rely more heavily on visual metrics, such as facial expressions, to inform interpersonal perceptions. In line with this conjecture, Ref. [48] found that facial expressions were correlated with perceptions of peer helpfulness, understanding and clarity.

Process: Individual Cognitive Processes

Individual cognitive processes during collaboration such as engagement and participation can be powerful indicators of the success of collaboration. In particular, imbalanced cognitive engagement can be seen as a deterrent for effective collaboration [20]. Research in this area has shown that mental effort is related to a group's performance [118] and that misunderstanding team members' individual cognitive processes hinders successful collaboration [110]. Thus, building an understanding of which metrics are indicative of individual cognitive processes can enable teams to better direct and monitor their progress together. Of the papers that measured individual cognitive processes, engagement, mental effort, and cognitive load were among the most studied subdimensions. Commonly used questionnaires were the Social Presence in Gaming questionnaire, and the NASA Task Index.

Physiological metrics were studied most frequently for cognitive processes, and physiological synchrony between participants was shown to significantly predict engagement and mental effort [60,61,66,68], but not perceived workload [63]. The study by [66] presents their findings as showing that episodes of continuous physiological synchrony among group members during collaborative problem solving refer to the groups' increased mental investment in the tasks. Other notable findings revealed that body distance was significantly related to physical engagement [79]. Additionally, during voice-based communication between collaborators, the presence of shared gaze reduced cognitive workload, potentially because it provided a shared referent [86]. However, the authors also suggest that, in text-based communication, shared gaze may increase cognitive load since participants must divide their attention.

Condition: Stable Personal Attributes

Stable personal attributes, such as personality type and individual expertise, were long studied as important contributors to group success. For instance, studies found that too many high achievers in a group can actually hinder collaboration [119] or that social loafing tendencies of individuals can have a negative impact on group performance [120]. In MMCA, three types of personal attributes seem to have emerged as key research areas: the 'Big Five' personality traits (extroversion, agreeableness, openness, conscientiousness, and neuroticism), leadership, and expertise.

For personality traits, Ref. [40] showed that the amount of speech activity and gaze attention in a group in the presence of an individual are good indicators of extroversion.

The study by [54] explores verbal, nonverbal and motion metrics to list key predictors for each trait, e.g., voice pitch for agreeableness. The study by [28] refines these metrics by studying the co-occurrence of verbal and nonverbal metrics, improving prediction accuracy. Lastly, Ref. [46] showed a high classification accuracy (around 80%) using a combination of inter- and intra-personal vocal behaviors, such as pitch and intensity. Their approach differs in that they use a time-aware machine learning method (i.e., BLSTM).

For leadership, visual dominance ratio (VDR) was found to be a popular metric that took advantage of the fact that dominant individuals receive more attention from peers. The early study by [55] automatically measured VDR through gaze and speech cues. Two papers spearheaded by Beyan [35,85] then used head pose to measure the attention received and successfully predicted leadership. Verbal metrics were also commonly used to predict leadership, such as verbal dominance and speech length [24,25]. The study by [24] explains that these findings are intuitive in light of their data, where leaders typically drove the conversation, proposed new topics, and summarized group decisions. Prosodic features, such as articulation rate or low pause durations were also indicative of leadership among student groups [33].

In the prediction of expertise, there is less convergence in the metrics used, but a range of promising metrics emerged as grounds for future investigation. The study by [33] found that students with higher expertise had a shorter pause duration in speech, and spent less time writing. The study by [29] achieved preliminary success in detecting domain experts in math using vocal features, in particular the patterns of transitions between speakers. The study by [27] looks at more context-dependent features for predicting math expertise, namely the time spent using calculators and mentions of numbers or mathematical terms, along with the more generalizable feature of speed of writing or drawing. The study by [88] finds that experts gesture less, but dynamically gesture more in hard problems, and differ in the type of gestures they make. Lastly, Ref. [75] somewhat uniquely predicted the expertise of Tetris papers with gaze and action metrics such as gaze fixation duration and the use of effective in-game actions.

We make three overarching observations across findings for different outcomes. Firstly, several established metrics were employed across all outcome types. Joint visual attention, physiological synchrony, acoustic features, verbal participation, and turn-taking, and visual attention, either given to a group member or an area of interest for a task, were among the most popularly explored metrics. Secondly, not all metrics from the same data source were predictive of the same outcome. For example, sophisticated verbal metrics such as linguistic coordination, but not speech length, were successfully correlated with learning gains. Thirdly, there were very few cases in which two metric–outcome connections could be said to be exactly alike. Homonymous metrics varied in their calculations, and outcomes were frequently dependent on task context or based on study-specific researcher codes. Connections were, in many cases, only many to one, or one to one, rather than both. This all points to the necessity of being principled and explicit about connecting metrics and outcomes in MMCA research. We explore this issue further in the discussion section.

4.3.4. Interactive Visualizations of the State of the Field

While prior connections between multimodal metrics and collaborative outcomes are important to keep in mind when designing new MMCA studies, it is difficult for individual researchers to keep track of this ever-growing landscape. Thus, our metric-construct mappings are available online [23] as a series of interactive visualizations, seen below in Figure 4. The graph starts on the far left with higher-level outcome categories, and ends on the far right with the specific types of metrics that were used. The edges represent the count of successful connections that were made, or the prevalence of certain lower-level categories within a higher-level category.

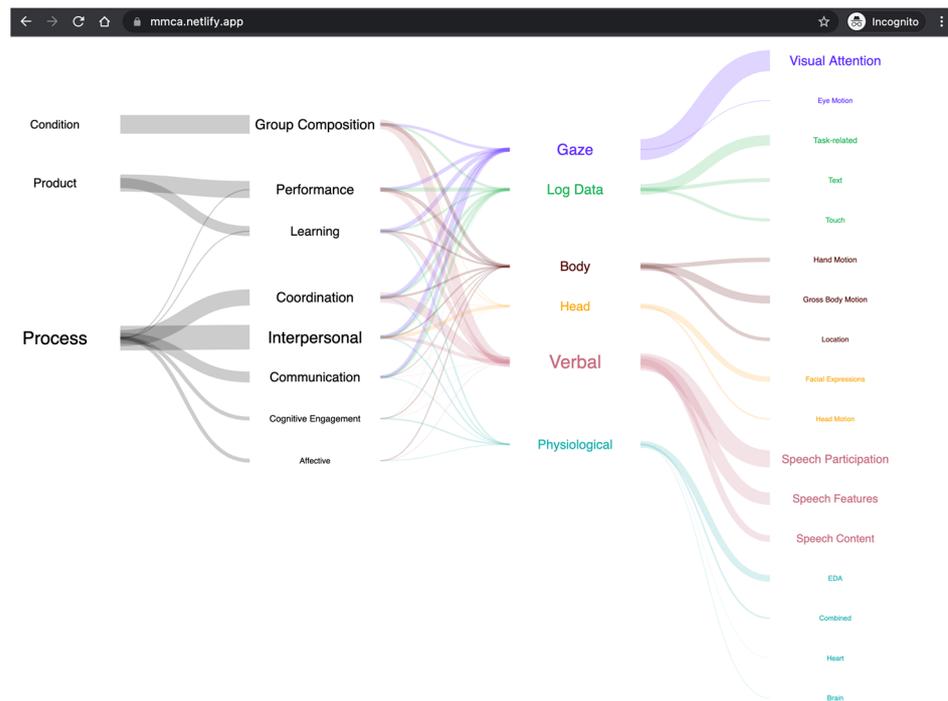


Figure 4. The Multimodal Collaboration Analytics (MMCA) website.

The nodes of edges of the graph are clickable, which allow users to drill down into lower-level outcomes and metrics, as in Figure 5. When users click on a lower-level outcome or metric, a Google Scholar page opens the related paper(s).

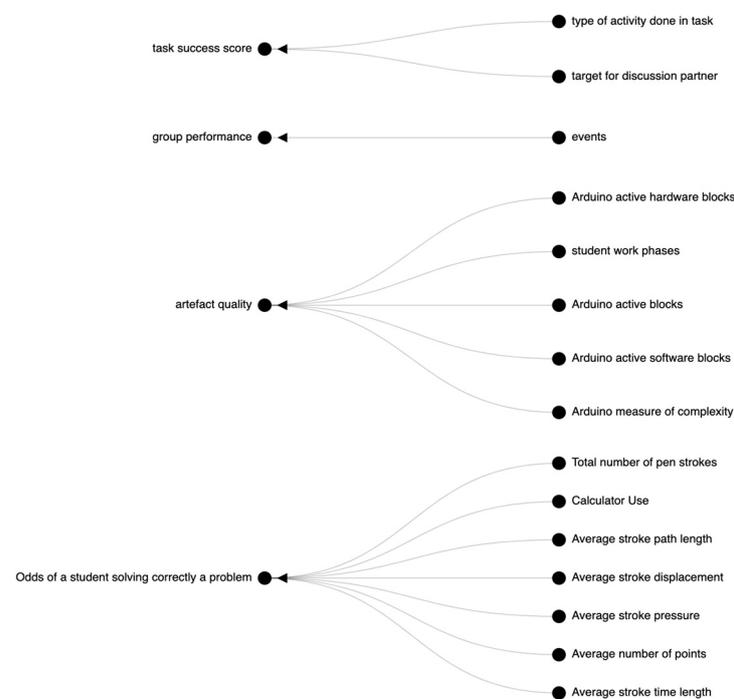


Figure 5. After clicking on the edge between “Log Data” and “Performance”, users can explore the outcomes and measures used to connect the two. By further clicking on the nodes, users can open the actual paper on Google Scholar.

This website has several intended functions. For non-academic users, it provides an entry point into the field of Multimodal Collaboration Analytics (MMCA). It shows what

the most studied metrics and outcomes are, and newcomers can access detailed instructions for replicating the approach used by the authors with the paper links. For academics, this website can help identify established and under-explored areas of study. For example, it seems that very few researchers have studied the group effect using multimodal sensing technology. Additionally, it provides a bird's eye view of approaches that could be used for capturing an outcome of interest. For example, an expert researcher interested in group coordination might be knowledgeable in eye-tracking methodologies, but less aware of other approaches that use verbal or body movement data. The website provides a map that could be used to make informed choices on metrics and data types, supporting a more holistic approach to studying collaborative processes. Our long-term vision is to continuously update the database and keep developing the visualizations to provide an easy-to-use website for understanding multimodal collaboration analytics. Ultimately, this could become part of a project such as the human genome project, for collaborative metrics and outcomes.

4.4. How Was Theory Used to Inform the Connections between Metrics and Outcomes (RQ4)?

Our final research question aims to understand the use of theory in past MMCA research. There is a growing consensus on the importance of theory in multiple stages of research for quantitative social sciences research (e.g., [17,121]). While this is often argued for, there is less empirical evidence on the ways or degree to which prior work has actually employed theory. We attempt to fill this gap by looking at how theory has so far been integrated into research. We then identify the references used repeatedly in MMCA research, presenting a list of key theories that can inform research for each of the collaborative sub-dimensions in our taxonomy.

4.4.1. How Is Theory Used in MMCA?

We reviewed papers in our corpora to understand how theory was actually utilized. While the degree to which theory was integrated forms a continuum, we were able to observe three distinctive categories: (1) no theory was explicitly used; (2) a theory was used to justify an outcome or metric; (3) a theory was used for more than simply justifying a metric or outcome.

The first category consists of studies that do not explicitly mention a theoretical framework for their analyses. A significant portion of papers belonged to this category (e.g., [39,41,49,79]). Studies concentrated on reporting the results of predictive analyses (e.g., supervised machine learning models) for an outcome of interest. For example, Ref. [49] adopts a data-driven approach to predict types of collaboration, generating thousands of gesture- and audio-based metrics and using strengths of correlations and a best first search method (similar to stepwise feature selection) to select a subset of metrics.

In the second (and largest) category, a theory was mentioned in the introduction or related work section of the paper to justify the choice of a particular collaborative outcome or data-derived feature (e.g., [48,79,86,91]). Studies often drew upon theory to argue for the significance of a particular sub-dimension of collaboration. For example, Ref. [79] draws upon various theoretical research (e.g., [20,122]) to establish the meaning and importance of collaborative problem solving, a process they then capture using multimodal data. Conversely, papers such as [91] draw from theoretical work (namely [123], on the role of gestures in teaching and learning) to justify the use of gestures in studying collaboration quality.

In the last (and smallest) category, a theory was used for more than simply justifying the use of a particular variable or construct. One example is the study of joint attention [124], the mechanism by which a shared reference helps collaborators build common ground (i.e., grounding theory [125]). Joint visual attention (JVA) was extensively studied using eye-trackers in MMCA studies (e.g., [24,70,71,73,77,82,84,86]). In this case, theory did more than inform the choice of collaborative features; it created a space where researchers could go back and forth between theoretical constructs and data-generated metrics to obtain a clearer

understanding of the relationship. For example, Ref. [26] identified which collaborator initiated and responded to offers of joint visual attention from the eye-tracking data, and showed that group members who shared this responsibility achieved more effective interactions in a collaborative learning task. Ref. [126] combined JVA with cognitive load information (generated from participants' pupil sizes). Combining these two measures provided deeper insights about the collaborative processes than JVA alone, suggesting that moments of joint visual attention *and* joint mental effort are crucial for high quality collaboration. Finally, theory was also used to inform interventions: Ref. [84] designed a controlled experiment where participants could see the gaze of their remote partner in real-time while solving a collaborative learning task. They found that this "gaze awareness tool" helped groups achieve more joint visual attention compared to a control group, which was positively and significantly correlated with collaboration quality and learning gains. In summary, a small subset of our corpora engaged in data-driven theory building (by clarifying and refining the relationship between the outcome and metrics), and designing interventions to support collaboration.

4.4.2. What Are the Core Theories for MMCA Research?

While collaboration is a multi-faceted construct, we expected a set of theories to be cross-referenced across MMCA papers. To test this assumption, we extracted 4278 references from the corpus, and found only ten references that were cited more than five times. Among these ten papers, four were empirical studies of joint visual attention using multiple eye-trackers [84,127–129]; two papers were reviews (of the use of multimodal data in education [4] and non-verbal activity in small groups [11]); two papers were empirical studies of leadership [30,130]; one paper was a collaboration coding scheme [67]; and one empirical paper studied physiological synchrony [65]. In short, we did not find a core set of theoretical references collaboration in our corpus. This finding reflects the diversity, or, alternatively, the lack of cohesive theoretical perspectives used to study collaboration with sensor data.

Even though we did not find a set of theories that were widely cited in our corpus, we observed a variety of frameworks that were used to inform the connection between data and constructs. First, there were holistic frameworks that integrated different collaborative dimensions. For example, McGrath [21] provides a conceptual framework that describes the relationship between individual characteristics, group structures, the task at hand, the technological setting, and how these factors influence the interactions within the group. This framework was useful for papers that attempted to combine multiple data sources and collaborative factors. Another framework that was used across multiple papers is the coding scheme developed by Meier, Spada and Rummel [98]. While this is not a theory paper per se, it describes a taxonomy of different collaborative dimensions, provides links to related theories, and is helpful in breaking down collaboration into nine distinct constructs (i.e., sustaining mutual understanding, dialogue management, information pooling, reaching a consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation). In Table 3 we provide a more exhaustive list of theories for each of our outcomes.

Table 3. Theories commonly used in the study of collaborative outcomes.

Product: Group performance
<ul style="list-style-type: none"> - Studies connecting group performance with data-derived metrics used the largest number of theories across all collaborative outcomes, from Convergent conceptual change [131], to Leadership styles [132], Interactive alignment [133], Joint attention [124], Grounding theory [125] and Personality theories [134]. These theories are described in more detail below.
Product: Learning outcomes
<ul style="list-style-type: none"> - Convergent conceptual change [131]: In this framework, collaboration is seen as the process of constructing shared meanings for conversations, concepts and experiences. Markers of collaborative learning are captured through iterative cycles of interactions that converge toward a shared set of behaviors and terms. - Shared meaning making [135] is associated with “the increased cognitive-interactive effort involved in the transition from learning to understand each other to learning to understand the meanings of the semiotic tools that constitute the mediators of interpersonal interaction”.
Process: Affective state
<ul style="list-style-type: none"> - Emotion contagion [136]: This theory postulates that group members tend to be affected by each other’s emotional state. In MMLA, this kind of “ripple effect” can be captured from observable behaviors (e.g., facial expressions, body postures). - Some researchers go further and posit the existence of emotion cycles [137], where emotions from an individual affect other group members, and subsequent reactions influence future interactions, creating emotion cycles.
Process: Interpersonal relationship
<ul style="list-style-type: none"> - Bion-Thelen Interaction theory [138]: the Bion-Thelen Interaction theory posits a “work” versus “emotion” distinction, but assumes that these are parallel processes. Thus, the Bion-Thelen Interaction theory predicts that interpersonal relationships should positively correlate with group performance. - Interpersonal relationships over time: Tickle-Degen and Rosenthal [139] proposed a dynamic structure of rapport, including mutual attentiveness, positivity and coordination and predicted that in early interactions, positivity and attentiveness weighed more heavily, whereas coordination and attentiveness weighed more heavily in later interactions. Less research in MMLA attempted to assess the temporal aspect of group dynamics, though the nature of high-frequency sensors is well-suited to lend empirical support for, or opposition to, these theories. - Nonverbal correlates of rapport [139]: In theorizing about measures of rapport, Tickle-Degen and Rosenthal described two levels of non-verbal correlates for interpersonal relationships. One was the molecular level, which included duration-level measurement, such as head nodding or eye-contact, and the other was the molar level, which combined the discrete acts of many participants in interaction. These dual levels of measurement correspond well to MMLA analysis. For example, correlations assess individual metrics, while machine learning often utilizes complex combinations of metrics together.
Process: Communication
<ul style="list-style-type: none"> - Interactive alignment: Pickering and Garrod [133] call for a mechanistic account of the dialogue—the interactive alignment account, which posits that the linguistic representations used by participants gain alignment through an automatic process. - Joint attention: Tomasello [124] postulates that the presence of joint attention is a necessary prerequisite for the understanding of other individuals as intentional agents, and for development of social cognition. - Media synchronicity theory [140] claims that the effectiveness of communication is affected by matching between the capabilities of the communication medium and the needs of the communication process.

Table 3. *Cont.*

Process: Coordination
<ul style="list-style-type: none"> - Grounding theory [125]: This theory describes how group members coordinate their actions to establish and maintain a common ground. For example, participants orient their body posture so that they attend to the same thing; they might use deictic (i.e., pointing) gestures to make sure that the group’s attention is correctly directed to a location of interest; and use particular linguistic constructions to create and sustain mutual understanding. Grounding ensures that collaborators are on the same page and share a common definition of the terms used.
Process: Individual Cognitive Processes
<ul style="list-style-type: none"> - Metacognitive monitoring [141]: Metacognitive monitoring describes the process of recognizing and adapting an individual’s cognitive or behavioral activity in order to better achieve one’s aims. Within the context of collaboration, this theory predicts that an individual’s perception and regulation of the team’s collaboration should influence the group’s final performance. As evidence for the validity of this theory, Järvelä and Hadwin [142] demonstrated that joint attention and mutual efforts were facilitated by the sharing of metacognitive judgements and feelings.
Condition: Stable personal attributes
<ul style="list-style-type: none"> - Personality: The big-five personality traits framework [134] is frequently applied in papers that aim to measure personal traits relevant to group work. - Expertise: Cognitive load theory [143] was used to hypothesise that experts are able to downshift to unimodal communication and show less gesturing in complex tasks.

Our survey suggests that theory might be underused in MMCA research; researchers primarily used theory to choose which features should be used to predict collaboration, but there are rich opportunities for MMCA to contribute to theory building. Finally, we note that the deep integration of theory is not a universal prerequisite for a paper to make meaningful contributions to the study of collaboration. In some fields conducting MMCA research, the norm is to draw implications ground-up from the data (e.g., computer science). Rather than making normative arguments on ‘what ought to be’ for any single paper, we hope to illustrate the range of ways theory can, and were used in MMCA.

The interest from various disciplines (e.g., social psychology, education, and communication, but also human–computer interaction and artificial intelligence) in studying collaboration with new analytical methods is reflected in the rich number of theories described above and how they are used to guide data collection and analysis. While this interest offers new opportunities for multidisciplinary collaborations and for advancing the field, it also presents challenges for creating shared methodologies and conceptualizations of collaboration analytics. We discuss these challenges and opportunities in more detail below.

5. Discussion

The overall takeaways of Section 4.1 (metrics), Section 4.2 (outcomes), Section 4.3 (metric-outcome connections), and Section 4.4 (the use of theory) are summarized into current strengths, potential challenges and future opportunities in Table 4.

Table 4. Takeaways from this review of the field (strengths, challenges, opportunities).

Dimension	Current Strengths	Potential Challenges	Future Opportunities
Metrics	<ul style="list-style-type: none"> - Data sources (e.g., sensors and computer vision algorithms) have become more accessible - They generate a rich range of metrics for capturing collaborative outcomes (see Table 1) 	<ul style="list-style-type: none"> - Many metrics have some level of divergence (same name, different computation; or vice-versa) - Researchers tend to specialize in unimodal data, and/or create their own metrics 	<ul style="list-style-type: none"> - To converge on agreed upon definitions and standardized data collection/cleaning/feature generation processes - To make multimodal data collection and analysis easier
Outcomes	<ul style="list-style-type: none"> - There is a theoretically grounded way of categorizing outcomes into products, processes, and conditions - There are emerging gold standards for generating ground truth measures of collaborative outcomes and processes (see Table 2) 	<ul style="list-style-type: none"> - Terms such as “collaboration” are used interchangeably for distinct collaborative outcomes - Different ways of capturing collaborative outcomes, even when they share the same label - There are understudied outcomes (e.g., affective) 	<ul style="list-style-type: none"> - To develop a finer taxonomy of collaborative outcomes, with associated definitions and operationalizations - To investigate outcomes and processes that are understudied (e.g., effects in groups)
Metric-outcome connections	<ul style="list-style-type: none"> - Some trends are emerging; for example, stable connections between: verbal metrics/head motion and interpersonal relationships, joint attention and coordination, etc. (see Figures 3 and 4) 	<ul style="list-style-type: none"> - Methodologies and types of connections vary (e.g., 1:1 t-tests, n:1 machine learning), which makes it difficult to compare connections - Some modalities have mixed results (e.g., physiological data, in particular, physiological synchrony) 	<ul style="list-style-type: none"> - To develop the best practices for computing and reporting results, to facilitate meta-analyses - To share data for easier replication - To connect a wider variety of group-level metrics with outcomes
Theory	<ul style="list-style-type: none"> - MMCA attracts researchers from very different fields, who use a rich variety of theories (see Table 3) - These different perspectives are useful for painting a holistic picture of collaboration (i.e., viewed from different angles) 	<ul style="list-style-type: none"> - Reconciling different theoretical perspectives is difficult, sometimes impossible - Theories tend to be modality-specific (e.g., emotion contagion for physiological data), which makes it challenging to integrate them 	<ul style="list-style-type: none"> - To integrate different theoretical perspectives by developing meta models of collaboration - To use theory to inform confounds, which results in interpretation and generalization
Overall	<ul style="list-style-type: none"> - This is an exciting time. There is a lot of innovative work and momentum in MMCA. - This momentum is attracting researchers from very different fields (e.g., psychology, education, engineering, etc.) 	<ul style="list-style-type: none"> - As a field, MMCA is both old (in terms of the theories and framework used) and young (in terms of the metrics generated, and their connection to outcomes) 	<ul style="list-style-type: none"> - To create multidisciplinary collaborations, so that the field can generate shared definitions, data collection tools, methodologies, and the reporting of results

As the current strengths and challenges were described in the Results section, we dedicate the discussion section to propose opportunities for the future of multimodal collaboration analytics. We structure this section according to our framework, and suggest opportunities for metrics, outcomes, metric–outcome connections, and the integration of theory.

5.1. Opportunities for Improving Sensor-Based Metrics

While sensor-based metrics are becoming more accessible, we observed a lack of common terminology, definitions, and methods for computing them. For instance, in determining what qualifies as joint visual attention, Refs. [78,83] employ different computations and parameters (e.g., thresholds for the amount of time delay and distance between two gazes). While varied approaches contribute to building momentum in MMCA, the field

would likely benefit from building up common ground across disciplines by explicitly sharing, replicating, and converging towards principled data collection, cleaning and metric creation processes.

5.2. Opportunities for Improving the Use of Collaborative Outcomes

Collaborative outcomes play a major role in validating sensor-based metrics by providing ground truths for collaborative constructs. Because the validity of results is contingent upon the validity of outcomes, researchers need to use validated, reliable instruments to measure outcomes whenever possible. From this perspective, there are many opportunities for improvement.

First, many papers tend to use generic terms such as “collaboration”, “communication”, or “coordination” when they are capturing different constructs. Oftentimes, these terms are not defined or operationalized precisely, which presents challenges for aggregating findings across studies. A first step towards pursuing rigor in outcomes is the taxonomy of outcomes we proposed in Section 4.2 and the related list of validated measures used to capture them (Table 2), intended to aid researchers in choosing appropriate instruments when validating sensor-based metrics. In the long term, however, MMCA would likely benefit from developing explicit “best practices” for measuring outcomes across collaborative dimensions. For example, in the field of social-emotional learning (SEL), researchers aimed to define rigorous domains of development, including cognitive, emotion, values and perspectives. These domains have specific skills and frameworks used to assess each one, which helps to build clarity and allows a greater precision for the field [6].

5.3. Opportunities for Improving the Connections between Metrics and Outcomes

Promising metric–outcome connections seem to be emerging, based on our review. For example, joint visual attention was repeatedly found to relate to coordination; verbal and head-based metrics seemed to reflect interpersonal relationships; and so on. However, we also noted some challenges. Findings are likely biased because researchers tend to report significant results and omit non-significant ones. Most metrics seem to be successful for this reason, which makes a fair evaluation challenging. There is also a variance in how researchers connect metrics to outcomes (from using simple correlations to supervised machine learning), and how they report their results. Effect sizes or variances were often omitted from results, which makes it difficult to conduct statistical meta-analyses. There is an opportunity to develop guidelines so that findings across studies can more easily be integrated (e.g., include non-significant results whenever possible and report effect sizes). Alternatively, the field would benefit from having a central repository where data are shared and can be jointly analyzed, so that meta-analyses can test the same metric across papers. Other fields benefited from this kind of data sharing practice (e.g., [144]). This would also allow researchers to build meta-models of collaboration and prediction models for different outcomes that are built from larger datasets. The website described in Section 4.3.4 is a first step in this direction. Finally, there is an opportunity for more researchers to study a wider variety of group-level metrics, particularly for verbal, body head and log modalities.

5.4. Opportunities for Integrating Theory

Lastly, our review suggests that theory might currently be underused. There are opportunities to use theory not just to select which aspect of collaboration to focus on, but also to inform micro-decisions in the MMCA pipeline. For example, Wise and Shaffer [17] organized the role of theory into distinct functions: giving guidance on variable choice, on potential confounds, subgroups, or covariates, and serving as a framework when choosing to attend to certain results, interpreting and generalizing them to new contexts. This suggests there are opportunities to use theory for more than simply informing the choice of sensor-based metrics and collaborative constructs. A final challenge is that there is a vast

number of collaborative theories available, and not all of them are suitable for informing sensor-based metrics. Thus, an additional opportunity is to organize a list of suitable theories, and how they can be used in MMCA. Table 3 is a first step in this direction.

5.5. Limitations and Future Steps

We acknowledge several limitations to our review. Firstly, we adopt a particular framework to understand prior work in MMCA, which may magnify or obscure certain issues within the field. Secondly, we are unable to carry out a quantitative meta-analysis, because more often than not effect sizes were not reported. As such, we do not have statistical evidence for our observations beyond count. Thirdly, we made simplifications in Section 4.3 to determine the success of a metric–outcome connection, particularly ambiguous in the case of machine learning papers. We chose models that had a higher-than-chance rate, and were highlighted as the best-performing models by the authors, but this required several judgement calls.

Future steps for the current project revolve around updating and enhancing the MMCA website. We plan to expand the database for the website and provide more fine-grained visualizations. Additionally, this platform would benefit from other features, such as the ability for other researchers to add their own data; provide the code used to generate metrics; or even a wiki-like platform associated with each outcome and metric so that researchers can discuss and agree on definitions, operationalizations and measurements. In the future, we also hope to create some of the standards mentioned in our opportunities section, such as the best practices for measuring outcomes, or guidelines for sharing MMCA results.

6. Conclusions—Why This Work Matters

While summarizing metric–outcome connections might not seem groundbreaking, it is the foundation on which innovative research and applications can be developed. Not only can this advance our scientific knowledge of collaboration by rigorously defining and operationalizing constructs from sensor-based metrics, it can also tell us, precisely, the strength of these connections. This matters in the context of developing valid and robust assessment tools. Cutting-edge Bayesian frameworks, such as the Evidence-Centered Design (ECD, sometimes referred to as Stealth Assessment [145]), require an evidence model where weights are given to particular behaviors. In the age of Big Data, one can imagine such an evidence model being fueled not just by limited, anecdotal, context-specific data, but by the entire academic literature, taking into account contexts, as well as individual and cultural differences, to build a generic assessment tool that can adapt to most common collaborative scenarios.

Other fields of research benefited from this kind of quantification. In medicine, for example, diagnostics were based on accumulated evidence of particular measures (e.g., white blood cells, cholesterol, glucose) crossing a particular threshold and connecting these values to particular health issues. Sensors had to be designed to capture these measures, and then the accumulated evidence was able to connect a set of values with particular diagnostics. When applied to collaboration, a similar approach can generate collaborative diagnostics. These diagnostics could provide information about which specific collaborative outcome is lacking, and which metrics were used to identify it.

If this kind of model of collaboration is created, however, the hope is that it will be used to create experiences that “augment the human potential” [146]. For example, by allowing Human–Computer Interaction (HCI) researchers to design playful, life-enhancing learning experiences through various media (e.g., Virtual/Augmented Reality, simulations, and video games, to name a few) to help individuals improve their collaborative skills. These examples provide a justification for tediously building the kind of metric–outcome connections described in this review.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Levy, F.; Murnane, R.J. Education and the changing job market. *Educ. Leadersh.* **2004**, *62*, 80.
- General Education Maps and Markers: Designing Meaningful Pathways to Student Achievement. Available online: <https://www.aacu.org/publications/gems/maps-and-markers> (accessed on 17 August 2021).
- Greiff, S.; Holt, D.V.; Funke, J. Perspectives on Problem Solving in Educational Assessment: Analytical, Interactive, and Collaborative Problem Solving. *J. Probl. Solving* **2013**, *5*, 5. Available online: <http://cogprints.org/9041/> (accessed on 17 August 2021). [[CrossRef](#)]
- Blikstein, P.; Worsley, M. Multimodal Learning Analytics and Education Data Mining: Using computational technologies to measure complex learning tasks. *J. Learn. Anal.* **2016**, *3*, 220–238. [[CrossRef](#)]
- Schneider, B.; Dowell, N.; Thompson, K. Collaboration Analytics—Current State and Potential Futures. *J. Learn. Anal.* **2021**, *8*, 1–12. [[CrossRef](#)]
- Jones, S.; Bailey, R.; Brush, K.; Nelson, B. Introduction to the taxonomy project: Tools for selecting & aligning SEL frameworks. *Collab. Acad. Soc. Emot. Learn.* **2019**, *3*, 1–13.
- Collins, F.S.; Morgan, M.; Patrinos, A. The Human Genome Project: Lessons from Large-Scale Biology. *Science* **2003**, *300*, 286–290. [[CrossRef](#)]
- Van Essen, D.C.; Smith, S.M.; Barch, D.M.; Behrens, T.E.J.; Yacoub, E.; Ugurbil, K. The WU-Minn Human Connectome Project: An overview. *NeuroImage* **2013**, *80*, 62–79. [[CrossRef](#)]
- Martinez-Maldonado, R.; Gašević, D.; Echeverria, V.; Nieto, G.F.; Swiecki, Z.; Shum, S.B. What Do You Mean by Collaboration Analytics? A Conceptual Model. *J. Learn. Anal.* **2021**, *8*, 126–153. [[CrossRef](#)]
- Sharma, K.; Giannakos, M. Multimodal data capabilities for learning: What can multimodal data tell us about learning? *Br. J. Educ. Technol.* **2020**, *51*, 1450–1484. [[CrossRef](#)]
- Gatica-Perez, D. Automatic nonverbal analysis of social interaction in small groups: A review. *Image Vis. Comput.* **2009**, *27*, 1775–1787. [[CrossRef](#)]
- Rasipuram, S.; Jayagopi, D.B. A comprehensive evaluation of audio-visual behavior in various modes of interviews in the wild. In *PETRA'19, Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments, Rhodes, Greece, 5–7 June 2019*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 94–100.
- Mitri, D.D.; Schneider, J.; Specht, M.; Drachler, H. From signals to knowledge: A conceptual model for multimodal learning analytics. *J. Comput. Assist. Learn.* **2018**, *34*, 338–349. [[CrossRef](#)]
- Shankar, S.K.; Prieto, L.P.; Rodríguez-Triana, M.J.; Ruiz-Calleja, A. A Review of Multimodal Learning Analytics Architectures. In *Proceedings of the 2018 IEEE, Proceedings of the 18th International Conference on Advanced Learning Technologies (ICALT)*, Mumbai, India, 9–13 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 365–369.
- Worsley, M. (Dis)engagement matters: Identifying efficacious learning practices with multimodal learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK'18)*, Sydney, NSW, Australia, 7–9 March 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 365–369.
- Mu, S.; Cui, M.; Huang, X. Multimodal Data Fusion in Learning Analytics: A Systematic Review. *Sensors* **2020**, *20*, 6856. [[CrossRef](#)] [[PubMed](#)]
- Wise, A.F.; Shaffer, D.W. Why Theory Matters More than Ever in the Age of Big Data. *J. Learn. Anal.* **2015**, *2*, 5–13. [[CrossRef](#)]
- Shaffer, D.W. *Quantitative Ethnography*; Cathcart Press: Madison, WI, USA, 2017; ISBN 978-0-578-19168-3.
- Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.A.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *J. Clin. Epidemiol.* **2009**, *62*, e1–e34. [[CrossRef](#)] [[PubMed](#)]
- Dillenbourg, P. *What Do You Mean by Collaborative Learning?* Elsevier: Amsterdam, The Netherlands, 1999; pp. 1–19.
- McGrath, J.E. *Groups: Interaction and Performance*; Prentice-Hall: Hoboken, NJ, USA, 1984; Volume 14.
- Thomson, A.; Perry, J. Collaboration processes: Inside the black box. *Public Adm. Rev.* **2006**, *66*, 20–32. [[CrossRef](#)]
- Multimodal Collaboration Analytics—A Review of the Literature. Available online: [Mmca.gse.harvard.edu](https://mmca.gse.harvard.edu) (accessed on 24 September 2021).
- Bhattacharya, I.; Foley, M.; Zhang, N.; Zhang, T.; Ku, C.; Mine, C.; Ji, H.; Riedl, C.; Welles, B.F.; Radke, R.J. A multimodal-sensor-enabled room for unobtrusive group meeting analysis. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI'18)*, Boulder, CO, USA, 16–20 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 347–355.
- Jayagopi, D.; Sanchez-Cortes, D.; Otsuka, K.; Yamato, J.; Gatica-Perez, D. Linking speaking and looking behavior patterns with group composition, perception, and performance. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI'12)*, Santa Monica, CA, USA, 22–26 October 2012; Association for Computing Machinery: New York, NY, USA, 2012; pp. 433–440.
- Schneider, B.; Sharma, K.; Cuendet, S.; Zufferey, G.; Dillenbourg, P.; Pea, R. *Detecting Collaborative Dynamics Using Mobile Eye-Trackers*; International Society of the Learning Sciences: Singapore, 2016.

27. Ochoa, X.; Chiluiza, K.; Méndez, G.; Luzardo, G.; Guamán, B.; Castells, J. Expertise estimation based on simple multimodal features. In Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI'13), Sydney, Australia, 9–13 December 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 583–590.
28. Okada, S.; Aran, O.; Gatica-Perez, D. Personality Trait Classification via Co-Occurrent Multiparty Multimodal Event Discovery. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI'15), Seattle, DC, USA, 9–13 November 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 15–22.
29. Luz, S. Automatic identification of experts and performance prediction in the multimodal math data corpus through analysis of speech interaction. In Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI'13), Sydney, Australia, 9–13 December 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 575–582.
30. Sanchez-Cortes, D.; Aran, O.; Jayagopi, D.B.; Mast, M.S.; Gatica-Perez, D. Emergent leaders through looking and speaking: From audio-visual data to multimodal recognition. *J. Multimodal User Interfaces* **2013**, *1*, 39–53. [[CrossRef](#)]
31. Starr, E.L.; Reilly, J.M.; Schneider, B. *Toward Using Multi-Modal Learning Analytics to Support and Measure Collaboration in Co-Located Dyads*; International Society of the Learning Sciences, Inc. [ISLS]: New York, NY, USA, 2018.
32. Miura, G.; Okada, S. Task-independent Multimodal Prediction of Group Performance Based on Product Dimensions. In Proceedings of the 2019 International Conference on Multimodal Interaction (ICMI'19), Suzhou, China, 14–18 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 264–273.
33. Scherer, S.; Weibel, N.; Morency, L.-P.; Oviatt, S. Multimodal prediction of expertise and leadership in learning groups. In Proceedings of the 1st International Workshop on Multimodal Learning Analytics (MLA'12), Santa Monica, CA, USA, 26 October 2012; Association for Computing Machinery: New York, NY, USA, 2012; pp. 1–8.
34. Dich, Y.; Reilly, J.; Schneider, B. Using Physiological Synchrony as an Indicator of Collaboration Quality, Task Performance and Learning. In *Artificial Intelligence in Education*; Penstein Rosé, C., Martínez-Maldonado, R., Hoppe, H.U., Luckin, R., Mavrikis, M., Porayska-Pomsta, K., McLaren, B., du Boulay, B., Eds.; Springer: Berlin, Germany, 2018; pp. 98–110.
35. Beyan, C.; Katsageorgiou, V.-M.; Murino, V. Moving as a Leader: Detecting Emergent Leadership in Small Groups using Body Pose. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1425–1433.
36. Martinez-Maldonado, R.; Kay, J.; Yacef, K. An Automatic Approach for Mining Patterns of Collaboration around an Interactive Tabletop. In *Lecture Notes in Computer Science, Proceedings of the 16th International Conference on Artificial Intelligence in Education, Memphis, TN, USA, 9–13 July 2013*; Lane, H.C., Yacef, K., Mostow, J., Pavlik, P., Eds.; Springer: Berlin, Germany, 2013; pp. 101–110. ISBN 978-3-642-39111-8.
37. Martinez, R.; Wallace, J.R.; Kay, J.; Yacef, K. Modelling and Identifying Collaborative Situations in a Collocated Multi-display Groupware Setting. In *Lecture Notes in Computer Science, Proceedings of the 15th International Conference on Artificial Intelligence in Education, Auckland, NZ, USA, 28 June–2 July 2011*; Biswas, G., Bull, S., Kay, J., Mitrovic, A., Eds.; Springer: Berlin, Germany, 2011; pp. 196–204. ISBN 978-3-642-21868-2.
38. Ponce-López, V.; Escalera, S.; Baró, X. Multi-modal social signal analysis for predicting agreement in conversation settings. In Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI'13), Sydney, NSW, Australia, 9–12 December 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 495–502.
39. Viswanathan, S.A.; VanLehn, K. Using the tablet gestures and speech of pairs of students to classify their collaboration. *IEEE Trans. Learn. Technol.* **2017**, *11*, 230–242. [[CrossRef](#)]
40. Lepri, B.; Subramanian, R.; Kalimeri, K.; Staiano, J.; Pianesi, F.; Sebe, N. Employing social gaze and speaking activity for automatic determination of the extraversion trait. In Proceedings of the International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI'10), Beijing, China, 8–10 November 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 1–8.
41. Bassiou, N.; Tsiartas, A.; Smith, J.; Bratt, H.; Richey, C.; Shriberg, E.; D'Angelo, C.; Alozie, N. Privacy-Preserving Speech Analytics for Automatic Assessment of Student Collaboration. In Proceedings of the Interspeech 2016, San Francisco, CA, USA, 8–12 September 2016; pp. 888–892.
42. Murray, G.; Oertel, C. Predicting Group Performance in Task-Based Interaction. In Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI'18), Boulder, CO, USA, 16–20 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 14–20.
43. Grafsgaard, J.F.; Wiggins, J.B.; Vail, A.K.; Boyer, K.E.; Wiebe, E.N.; Lester, J.C. The additive value of multimodal features for predicting engagement, frustration, and learning during tutoring. In Proceedings of the 16th International Conference on Multimodal Interaction (ICMI'14), Istanbul, Turkey, 12–16 November 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 42–49.
44. Sharma, K.; Leftheriotis, I.; Noor, J.; Giannakos, M. *Dual Gaze as a Proxy for Collaboration in Informal Learning*; International Society of the Learning Sciences: Philadelphia, PA, USA, 2017.
45. Martinez-Maldonado, R.; Dimitriadis, Y.; Martinez-Monés, A.; Kay, J.; Yacef, K. Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. *Intern. J. Comput.-Support. Collab. Learn.* **2013**, *8*, 455–485. [[CrossRef](#)]

46. Lin, Y.-S.; Lee, C.-C. Using Interlocutor-Modulated Attention BLSTM to Predict Personality Traits in Small Group Interaction. In Proceedings of the 2018 on International Conference on Multimodal Interaction (ICMI'18), Boulder, CO, USA, 16–20 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 163–169.
47. Schneider, B.; Pea, R. Does seeing one another's gaze affect group dialogue? A computational approach. *J. Learn. Anal.* **2015**, *2*, 107–133. [[CrossRef](#)]
48. McDuff, D.; Thomas, P.; Czerwinski, M.; Craswell, N. Multimodal analysis of vocal collaborative search: A public corpus and results. In Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI'17), Glasgow, UK, 13–17 November 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 456–463.
49. Schneider, B.; Pea, R. The Effect of Mutual Gaze Perception on Students' Verbal Coordination. In Proceedings of the 7th International Conference on Educational Data Mining, London, UK, 4–7 July 2014; pp. 138–144.
50. Bachour, K.; Kaplan, F.; Dillenbourg, P. An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *IEEE Trans. Learn. Technol.* **2010**, *3*, 203–213. [[CrossRef](#)]
51. Kantharaju, R.B.; Ringeval, F.; Besacier, L. Automatic Recognition of Affective Laughter in Spontaneous Dyadic Interactions from Audiovisual Signals. In Proceedings of the 2018 on International Conference on Multimodal Interaction (ICMI'18), Boulder, CO, USA, 16–20 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 220–228.
52. Eloy, L.; Stewart, A.E.B.; Amon, M.J.; Reinhardt, C.; Michaels, A.; Chen, S.; Shute, V.; Duran, N.D.; D'Mello, S.K. Modeling Team-level Multimodal Dynamics during Multiparty Collaboration. In Proceedings of the 2019 International Conference on Multimodal Interaction (ICMI'19), Suzhou, China, 14–18 October 2018; Association for Computing Machinery: New York, NY, USA, 2019; pp. 244–258.
53. Lubold, N.; Pon-Barry, H. Acoustic-Prosodic Entrainment and Rapport in Collaborative Learning Dialogues. In Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge (MLA'14), Istanbul, Turkey, 12 November 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 5–12.
54. Fang, S.; Achard, C.; Dubuisson, S. Personality classification and behaviour interpretation: An approach based on feature categories. In Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI'16), Tokyo, Japan, 12–16 November 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 225–232.
55. Hung, H.; Jayagopi, D.B.; Ba, S.; Odobez, J.-M.; Gatica-Perez, D. Investigating automatic dominance estimation in groups from visual attention and speaking activity. In Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI'08), Chania, Crete, Greece, 20–22 October 2008; Association for Computing Machinery: New York, NY, USA, 2008; pp. 233–236.
56. Viswanathan, S.A.; Vanlehn, K. *High Accuracy Detection of Collaboration from Log Data and Superficial Speech Features*; International Society of the Learning Sciences: Philadelphia, PA, USA, 2017.
57. Reilly, J.M.; Schneider, B. Predicting the Quality of Collaborative Problem Solving Through Linguistic Analysis of Discourse. In Proceedings of the 12th International Conference on Educational Data Mining, Montreal, QC, Canada, 2–5 July 2019.
58. Spikol, D.; Ruffaldi, E.; Dabisias, G.; Cukurova, M. Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *J. Comput. Assist. Learn.* **2018**, 366–377. [[CrossRef](#)]
59. Spikol, D.; Ruffaldi, E.; Landolfi, L.; Cukurova, M. Estimation of success in collaborative learning based on multimodal learning analytics features. In Proceedings of the 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), Timisoara, Romania, 3–7 July 2017; pp. 269–273.
60. Järvelä, S.; Kivikangas, J.M.; Kätsyri, J.; Ravaja, N. Physiological linkage of dyadic gaming experience. *Simul. Gaming* **2014**, *45*, 24–40. [[CrossRef](#)]
61. Malmberg, J.; Haataja, E.; Seppänen, T.; Järvelä, S. Are we together or not? The temporal interplay of monitoring, physiological arousal and physiological synchrony during a collaborative exam. *Int. J. Comput. Collab. Learn.* **2019**, *14*, 467–490. [[CrossRef](#)]
62. Schneider, B.; Dich, Y.; Radu, I. Unpacking the Relationship between Existing and New Measures of Physiological Synchrony and Collaborative Learning: A Mixed Methods Study. *Int. J. Comput. Collab. Learn.* **2020**, *15*, 89–113. [[CrossRef](#)]
63. Montague, E.; Xu, J.; Chiou, E. Shared experiences of technology and trust: An experimental study of physiological compliance between active and passive users in technology-mediated collaborative encounters. *IEEE Trans. Hum.-Mach. Syst.* **2014**, *44*, 614–624. [[CrossRef](#)]
64. Malmberg, J.; Järvelä, S.; Holappa, J.; Haataja, E.; Huang, X.; Siipo, A. Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Comput. Hum. Behav.* **2019**, *96*, 235–245. [[CrossRef](#)]
65. Pijera-Díaz, H.J.; Drachler, H.; Järvelä, S.; Kirschner, P.A. Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK'16), Edinburgh, UK, 25–29 April 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 64–73.
66. Dindar, M.; Järvelä, S.; Haataja, E. What does physiological synchrony reveal about metacognitive experiences and group performance? *Br. J. Educ. Technol.* **2020**, *51*, 1577–1596. [[CrossRef](#)]
67. Mønster, D.; Håkansson, D.D.; Eskildsen, J.K.; Wallot, S. Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiol. Behav.* **2016**, *156*, 24–34. [[CrossRef](#)] [[PubMed](#)]
68. Liu, Y.; Wang, T.; Wang, K.; Zhang, Y. Predicting Collaborative Learning Quality through Physiological Synchrony Recorded by Wearable Biosensors. *bioRxiv* **2020**. [[CrossRef](#)]

69. Dikker, S.; Wan, L.; Davidesco, I.; Kaggen, L.; Oostrik, M.; McClintock, J.; Rowland, J.; Michalareas, G.; Van Bavel, J.J.; Ding, M.; et al. Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Curr. Biol.* **2017**, *27*, 1375–1380. [[CrossRef](#)] [[PubMed](#)]
70. Sharma, K.; Nüssli, M.-A.; Jermann, P.; Dillenbourg, P. *Understanding Collaborative Program Comprehension: Interlacing Gaze and Dialogues*; International Society of the Learning Sciences, Inc. [ISLS]: New York, NY, USA, 2013.
71. Sharma, K.; Caballero, D.; Verma, H.; Jermann, P.; Dillenbourg, P. *Looking AT Versus Looking through: A Dual Eye-Tracking Study in MOOC Context*; International Society of the Learning Sciences, Inc. [ISLS]: New York, NY, USA, 2015.
72. Andrist, S.; Ruis, A.R.; Shaffer, D.W. A network analytic approach to gaze coordination during a collaborative task. *Comput. Hum. Behav.* **2018**, *89*, 339–348. [[CrossRef](#)]
73. Schneider, B.; Pea, R. *Using Eye-Tracking Technology to Support Visual Coordination in Collaborative Problem-Solving Groups*; International Society of the Learning Sciences, Inc. [ISLS]: Madison, WI, USA, 2013.
74. Schneider, B.; Sharma, K.; Cuendet, S.; Zufferey, G.; Dillenbourg, P.; Pea, R. Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *Int. J. Comput. Collab. Learn.* **2018**, *13*, 241–261. [[CrossRef](#)]
75. Li, W.; Nüssli, M.-A.; Jermann, P. Gaze quality assisted automatic recognition of social contexts in collaborative Tetris. In Proceedings of the ICMI-International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (MLMI'10), Beijing, China, 8–10 November 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 1–8.
76. Higuch, K.; Yonetani, R.; Sato, Y. Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16), San Jose, CA, USA, 7–12 May 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 5180–5190.
77. Schneider, B.; Sharma, K.; Cuendet, S.; Zufferey, G.; Dillenbourg, P.; Pea, R. Using mobile eye-trackers to unpack the perceptual benefits of a tangible user interface for collaborative learning. *ACM Trans. Comput.-Hum. Interact.* **2016**, *23*, 1–23. [[CrossRef](#)]
78. Vrzakova, H.; Amon, M.J.; Stewart, A.E.; D'Mello, S.K. Dynamics of visual attention in multiparty collaborative problem solving using multidimensional recurrence quantification analysis. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19), Glasgow Scotland, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–14.
79. Spikol, D.; Ruffaldi, E.; Cukurova, M. *Using Multimodal Learning Analytics to Identify Aspects of Collaboration in Project-Based Learning*; International Society of the Learning Sciences: Philadelphia, PA, USA, 2017; p. 8.
80. Sharma, K.; Olsen, J. *An Alternate Statistical Lens to Look at Collaboration Data: Extreme Value Theory*; International Society of the Learning Sciences, Inc. [ISLS]: New York, NY, USA, 2019; pp. 400–407.
81. Schneider, B.; Pea, R. Toward collaboration sensing. *Int. J. Comput. Collab. Learn.* **2014**, *9*, 371–395. [[CrossRef](#)]
82. Schneider, B. *Unpacking Collaborative Learning Processes during Hands-On Activities Using Mobile Eye-Trackers*; International Society of the Learning Sciences, Inc. [ISLS]: New York, NY, USA, 2019.
83. Schneider, B.; Sharma, K.; Cuendet, S.; Zufferey, G.; Dillenbourg, P.; Pea, R.D. *3D Tangibles Facilitate Joint Visual Attention in Dyads*; International Society of the Learning Sciences, Inc. [ISLS]: New York, NY, USA, 2015; Volume 1, pp. 156–165.
84. Schneider, B.; Pea, R. Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *Int. J. Comput. Collab. Learn.* **2013**, *8*, 375–397. [[CrossRef](#)]
85. Beyan, C.; Carissimi, N.; Capozzi, F.; Vascon, S.; Bustreo, M.; Pierro, A.; Becchio, C.; Murino, V. Detecting emergent leader in a meeting environment using nonverbal visual features only. In Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI'16), Tokyo, Japan, 12–16 November 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 317–324.
86. Kütt, G.H.; Tanprasert, T.; Rodolitz, J.; Moyza, B.; So, S.; Kenderova, G.; Papoutsaki, A. Effects of Shared Gaze on Audio-Versus Text-Based Remote Collaborations. *Proc. ACM Hum.-Comput. Interact.* **2020**, *4*, 1–25. [[CrossRef](#)]
87. Gergle, D.; Alan, C. See What I'm Saying? Using Dyadic Mobile Eye Tracking to Study Collaborative Reference. In Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW'11), Hangzhou, China, 19–23 March 2011; pp. 435–444.
88. Sriramulu, A.; Lin, J.; Oviatt, S. Dynamic Adaptive Gesturing Predicts Domain Expertise in Mathematics. In Proceedings of the 2019 International Conference on Multimodal Interaction (ICMI'19), Suzhou, China, 14–18 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 105–113.
89. Yamashita, N.; Kaji, K.; Kuzuoka, H.; Hirata, K. Improving visibility of remote gestures in distributed tabletop collaboration. In Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW'11), Hangzhou, China, 19–23 March 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 95–104.
90. Schneider, B.; Blikstein, P. Unraveling students' interaction around a tangible interface using multimodal learning analytics. *J. Educ. Data Min.* **2015**, *7*, 89–116.
91. Reilly, J.M.; Ravenell, M.; Schneider, B. Exploring Collaboration Using Motion Sensors and Multi-Modal Learning Analytics. In Proceedings of the International Educational Data Mining Society, Raleigh, NC, USA, 16–20 July 2018.
92. Vrzakova, H.; Amon, M.J.; Stewart, A.; Duran, N.D.; D'Mello, S.K. Focused or stuck together: Multimodal patterns reveal triads' performance in collaborative problem solving. In Proceedings of the Tenth International Conference on Learning Analytics &

- Knowledge (LAK'20), Frankfurt, Germany, 23–27 March 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 295–304.
93. Behoora, I.; Tucker, C.S. Machine learning classification of design team members' body language patterns for real time emotional state detection. *Des. Stud.* **2015**, *39*, 100–127. [[CrossRef](#)]
 94. Dale, R.; Bryant, G.A.; Manson, J.H.; Gervais, M.M. Body synchrony in triadic interaction. *R. Soc. Open Sci.* **2020**, *7*, 200095. [[CrossRef](#)]
 95. Chng, E.; Seyam, M.R.; Yao, W.; Schneider, B. Using Motion Sensors to Understand Collaborative Interactions in Digital Fabrication Labs. In *Lecture Notes in Computer Science-Artificial Intelligence in Education*; Bittencourt, I.I., Cukurova, M., Muldner, K., Luckin, R., Millán, E., Eds.; Springer: Berlin, Germany, 2020; pp. 118–128. ISBN 978-3-030-52236-0.
 96. Evans, A.C.; Wobbrock, J.O.; Davis, K. Modeling Collaboration Patterns on an Interactive Tabletop in a Classroom Setting. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW'16), San Francisco, CA, USA, 27 February–2 March 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 858–869.
 97. Martinez, R.; Yacef, K.; Kay, J. Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In *Educational Data Mining 2011*; CEUR-WS: Aachen, Germany, 2011; pp. 111–120.
 98. Meier, A.; Spada, H.; Rummel, N. A rating scheme for assessing the quality of computer-supported collaboration processes. *Int. J. Comput. Collab. Learn.* **2007**, *2*, 63–86. [[CrossRef](#)]
 99. D'Mello, S. Emotional learning analytics. In *Handbook of Learning Analytics*; SOLAR: New York, NY, USA, 2017; Volume 115.
 100. Hart, S.; Staveland, L. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*; Elsevier: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183. ISBN 978-0-444-70388-0.
 101. Scandell, D.J.; Wlazelek, B. The relationship between self-perceived personality and impression management on the neo-ffi. *Personal. Individ. Differ.* **1999**, *27*, 147–154. [[CrossRef](#)]
 102. Yvonne, A.W.; de Kort, I.; Jsselsteijn, W.A.; Poels, K. Digital games as social presence technology: Development of the Social Presence in Gaming Questionnaire (SPGQ). In Proceedings of the 10th Annual International Workshop on Presence, Barcelona, Spain, 25–27 October 2007; pp. 1–9.
 103. Ainley, M.; Corrigan, M.; Richardson, N. Students, tasks and emotions: Identifying the contribution of emotions to students' reading of popular culture and popular science texts. *Learn. Instr.* **2005**, *15*, 433–447. [[CrossRef](#)]
 104. Hadwin, A.F.; Webster, E.A. Calibration in goal setting: Examining the nature of judgments of confidence. *Learn. Instr.* **2013**, *24*, 37–47. [[CrossRef](#)]
 105. Pintrich, P. *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*; ERIC: Washington, DC, USA, 1991.
 106. Manson, J.H.; Gregory, B.A.; Gervais, M.M.; Kline, M.A. Convergence of speech rate in conversation predicts cooperation. *Evol. Hum. Behav.* **2013**, *34*, 419–426. [[CrossRef](#)]
 107. O'Brien, H.L.; Toms, E.G.; Kelloway, E.K.; Kelley, E. Developing and evaluating a reliable measure of user engagement. *Proc. Am. Soc. Inf. Sci. Technol.* **2008**, *45*, 1–10. [[CrossRef](#)]
 108. Paas, F.G. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J. Educ. Psychol.* **1992**, *84*, 429. [[CrossRef](#)]
 109. Tapola, A.; Veermans, M.; Niemivirta, M. Predictors and outcomes of situational interest during a science learning tasks. *Instr. Sci.* **2013**, *41*, 1047–1064. [[CrossRef](#)]
 110. Salonen, P.; Vauras, M.; Efklides, A. Social Interaction—What Can It Tell Us about Metacognition and Coregulation in Learning? *Eur. Psychol.* **2005**, *10*, 199–208. [[CrossRef](#)]
 111. Joann, K.; Wall, V.D. SYMLOG: Theory and method for measuring group and organizational communication. *Manag. Commun. Q.* **1989**, *2.4*, 544–567.
 112. Lord, R.; Foti, R.; Vader, C.D. A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organ. Behav. Hum. Perform.* **1984**, *34*, 343–378. [[CrossRef](#)]
 113. van Blankenstein, F.M.; Dolmans, D.H.J.M.; van der Vleuten, C.P.M.; Schmidt, H.G. Which cognitive processes support learning during small-group discussion? The role of providing explanations and listening to others. *Instr. Sci.* **2011**, *39*, 189–204. [[CrossRef](#)]
 114. Kelly, J.R.; Barsade, S. Mood and emotions in small groups and work teams. *Organ. Behav. Hum. Decis. Process.* **2001**, *86*, 99–130. [[CrossRef](#)]
 115. Le Dantec, C.A.; Do, E.Y.-L. The mechanisms of value transfer in design meetings. *Des. Stud.* **2009**, *30*, 119–137. [[CrossRef](#)]
 116. Ogan, A.; Finkelstein, S.; Walker, E.; Carlson, R.; Cassell, J. Rudeness and rapport: Insults and learning gains in peer tutoring. In *Intelligent Tutoring Systems*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 11–21.
 117. Barry, B.; Stewart, G.L. Composition, process and performance in self-managed groups: The role of personality. *J. Appl. Psychol.* **1997**, *82*, 62–78. [[CrossRef](#)] [[PubMed](#)]
 118. Dindar, M.; Sanna, J.; Hanna, J. Interplay of metacognitive experiences and performance in collaborative problem solving. *Comput. Educ.* **2020**, *154*, 103922. [[CrossRef](#)]
 119. Groysberg, B.; Polzer, J.T.; Elfenbein, H.A. Too Many Cooks Spoil the Broth: How High-Status Individuals Decrease Group Effectiveness. *Organ. Sci.* **2011**, *22*, 722–737. [[CrossRef](#)]
 120. Ying, X.; Li, H.; Jiang, S.; Peng, F.; Lin, Z. Group Laziness: The Effect of Social Loafing on Group Performance. *Soc. Behav. Personal. Int. J.* **2014**, *42*, 465–471. [[CrossRef](#)]
 121. Stroebe, W.; Strack, F. The alleged crisis and the illusion of exact replication. *Perspect. Psychol. Sci.* **2014**, *9*, 59–71. [[CrossRef](#)]

122. Damon, W.; Phelps, E. Critical distinctions among three approaches to peer education. *Int. J. Educ. Res.* **1989**, *13*, 9–19. [CrossRef]
123. Roth, W.-M. Gestures: Their Role in Teaching and Learning. *Rev. Educ. Res.* **2001**, *71*, 365–392. [CrossRef]
124. Tomasello, M. Joint attention as social cognition. In *Joint Attention: Its Origins and Role in Development*; Moore, C., Dunham, P.J., Eds.; Lawrence Erlbaum Associates, Inc.: Hillsdale, NJ, USA, 1995; pp. 103–130. ISBN 0-8058-1437-X.
125. Clark, H.H.; Brennan, S.E. Grounding in communication. In *Perspectives on Socially Shared Cognition*; Resnick, L.B., Levine, J.M., Teasley, S.D., Eds.; American Psychological Association: Washington, DC, USA, 1991; pp. 127–149. ISBN 1-55798-121-3.
126. Sharma, K.; Olsen, J.; Himanshu, V.; Caballero, D.; Jermann, P. Challenging Joint Visual Attention as a Proxy for Collaborative Performance. In *Proceedings of the International Conference on Computer Supported Collaborative Learning*; International Society of the Learning Sciences, Inc. [ISLS]: Philadelphia, PA, USA, 2021; Volume 1, pp. 91–98.
127. Richardson, D.C.; Dale, R. Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cogn. Sci.* **2005**, *29*, 1045–1060. [CrossRef] [PubMed]
128. Jermann, P.; Nüssli, M.-A. Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW'12)*, Seattle, DC, USA, 11–15 February 2012; ACM: New York, NY, USA, 2012; pp. 1125–1134.
129. Cherubini, M.; Nüssli, M.-A.; Dillenbourg, P. Deixis and gaze in collaborative work at a distance (over a shared map): A computational model to detect misunderstandings. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications (ETRA'08)*, Savannah, GA, USA, 26–28 March 2008; ACM: New York, NY, USA, 2008; pp. 173–180.
130. Sanchez-Cortes, D.; Aran, O.; Gatica-Perez, D. An audio visual corpus for emergent leader analysis. In *Proceedings of the ICMI-MLMI'11, Multimodal Corpora for Machine Learning: Taking Stock and Road mapping the Future*, Alicante, Spain, 14–18 November 2011; pp. 1–6.
131. Roschelle, J. Learning by collaborating: Convergent conceptual change. *J. Learn. Sci.* **1992**, *2*, 235–276. [CrossRef]
132. Fiedler, F.E.; House, R.J. Leadership theory and research: A report of progress. In *International Review of Industrial and Organizational Psychology*; John Wiley & Sons: Oxford, UK, 1988; pp. 73–92, ISBN 978-0-471-91844-8.
133. Pickering, M.J.; Garrod, S. Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* **2004**, *27*, 169–190. [CrossRef]
134. McCrae, R.R.; Costa, P.T., Jr. The five-factor theory of personality. In *Handbook of Personality: Theory and Research*, 3rd ed.; The Guilford Press: New York, NY, USA, 2008; pp. 159–181, ISBN 978-1-59385-836-0.
135. Baker, M.; Hansen, T.; Joiner, R.; Traum, D. The role of grounding in collaborative learning tasks. In *Collaborative Learning: Cognitive and Computational Approaches*; Pergamon: Oxford, UK, 1999; Volume 31, p. 63.
136. Barsade, S.G. The Ripple Effect: Emotional Contagion and its Influence on Group Behavior. *Adm. Sci. Q.* **2002**, *47*, 644–675. [CrossRef]
137. Hareli, S.; Rafaeli, A. Emotion cycles: On the social influence of emotion in organizations. *Res. Organ. Behav.* **2008**, *28*, 35–59. [CrossRef]
138. Bion, W.R. *Experiences in Groups, and Other Papers*; Routledge: London, UK, 1961.
139. Tickle-Degnen, L.; Rosenthal, R. The Nature of Rapport and Its Nonverbal Correlates. *Psychol. Inq.* **1990**, *1*, 285–293. [CrossRef]
140. Dennis, A.R.; Fuller, R.M.; Valacich, J.S. Media, Tasks, and Communication Processes: A Theory of Media Synchronicity. *MIS Q.* **2008**, *32*, 575–600. [CrossRef]
141. Metamemory: A Theoretical Framework and New Findings. Available online: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=gz0dRLgAAAAJ&citation_for_view=gz0dRLgAAAAJ:2osOgNQ5qMEC (accessed on 20 November 2021).
142. Järvelä, S.; Hadwin, A.F. New frontiers: Regulating learning in CSCL. *Educ. Psychol.* **2013**, *48*, 25–39. [CrossRef]
143. Sweller, J. Cognitive load theory: Recent theoretical advances. In *Cognitive Load Theory*; Cambridge University Press: Cambridge, UK, 2010; pp. 29–47. ISBN 978-0-521-67758-5.
144. Stamper, J.C.; Koedinger, K.R. Human-Machine Student Model Discovery and Improvement Using DataShop. In *Proceedings of the Artificial Intelligence in Education*; Biswas, G., Bull, S., Kay, J., Mitrovic, A., Eds.; Springer: Berlin, Germany, 2011; pp. 353–360.
145. Shute, V.J.; Kim, Y.J. Formative and Stealth Assessment. In *Handbook of Research on Educational Communications and Technology*; Spector, J.M., Merrill, M.D., Elen, J., Bishop, M.J., Eds.; Springer: New York, NY, USA, 2014; pp. 311–321, ISBN 978-1-4614-3185-5.
146. Engelbart, D.C. *Augmenting Human Intellect: A Conceptual Framework*; SRI International: Menlo Park, CA, USA, 1962.