

Harry Bernard Quedenfeld

**PHYSIOLOGICAL RESPONSE DATA IN  
EDUCATION: COULD THIS BE THE FUTURE  
FOR ANALYZING SOCIAL INTERACTIONS  
AND LEARNING? A SYSTEMATIC  
LITERATURE REVIEW**

Faculty of Education and Culture  
Master Thesis  
February 2022

# ABSTRACT

Harry Bernard Quedenfeld: Physiological response data in education: could this be the future for analyzing social interactions and learning? A systematic literature review

Master's Thesis

Tampere University

Master of Arts (Teaching, Learning, and Media Education)

February 2022

---

This thesis systematically reviews the literature which integrated physiological and other biodata to study human social interactions in the fields of education, physiology, neurophysiology, affective neuroscience, and psychology over the past decade (2011-2021). The aim was to identify the benefits and drawbacks of their use of physiological data and inform future directions in education. This systematic literature review pre-defined data collection methodology, which involved keyword selection, database selection and searching, title and abstract screening and sorting, methods section appraisals, and full-text screening. Studies were categorized to address distinct research questions. Qualitative-only studies were separated into a pool to address research questions 2 and 2a from the qualitative perspective. The remaining quantitative and mixed methods studies were then segregated based on the field of research: in or out of the education field to address research questions 1-1c and 2-2a respectively. In education, studies most often measured electrodermal activity (EDA) (65%) and eye movements (27%), especially interested in the synchrony of bio-physiological signals (85%), in-person (65%) with groups of two to four students (85%). The analysis revealed the plausible utility of biodata for research in education involving social interactions, particularly in learning analytics research. In combination with other data, biodata can measure prior knowledge and individual and group level emotional, cognitive, and relational components of collaborative learning. Additionally, biodata can indicate learning gains, collaboration quality, task performance, and cognitive challenge, though in a context- and time-dependent manner. Multidimensional recurrence quantification analysis, matrix analysis, and minimum width envelope were identified as promising data analysis techniques to gain insights about interpersonal cognitive, attentional, metacognitive, social, and emotional process dynamics from time-series bio-physiological data of multiple subjects. Further, the visualization of eye-tracking data was identified as a useful tool for intervention in learning as well as for qualitative content analysis. The analysis found that current methodologies in education suffer from paradigmatic ambiguities that specifically arise from experimental design, data sources, data handling, and data analysis. A 2 x 2 confusion matrix revealed methodologically based ambiguities in the reviewed literature, namely the weak ability of several studies to address true negative, false positive, and false negative results. Multimodal biodata, particularly for triangulation, can address limitations imposed by the sources of data. Standardization of protocol for signal selection, thresholding, and data processing were recommended. As well, standardization of statistical test usage would help reduce current bias of diverse approaches. Suggestions are made to clarify the tacit features of experimental paradigms such that study replicability, comparability, and hence value increases. Practically, a ramping up of interdisciplinary efforts is recommended to tackle the challenges of multimodal biodata handling and analysis. This research concludes that the inclusion of physiological and other biodata in the education field offers greater potential with these adjustments.

Keywords: Human social interactions, physiological data, biodata, collaborative learning

The originality of this thesis has been checked using the Turnitin Originality Check service.

# CONTENTS

<b>1</b>	<b>INTRODUCTION</b> .....	<b>5</b>
<b>2</b>	<b>SOCIAL INTERACTIONS AND DYNAMICS IN LEARNING ENVIRONMENTS</b> .....	<b>7</b>
2.1	Defining visual and joint attention .....	8
2.1.1	<i>Why eye-tracking would be important: joint attention in collaborative learning</i> .....	9
2.1.2	<i>Gaze data: the current state of designs and analytic methods</i> .....	10
2.2	Emotions in interactions in education .....	13
2.2.1	<i>Why study socioemotional group interactions?</i> .....	14
2.2.2	<i>What does biosensing offer? The value of multimodal data for emotions in learning</i> 15	
2.2.3	<i>Example 1: EDA as an index of emotional state in education</i> .....	16
2.2.4	<i>Example 2: Utility of facial expression recognition to measure emotional valence in education and its current challenges</i> .....	17
2.3	But what about cognitive processes? .....	19
2.3.1	<i>What is cognition?</i> .....	20
2.3.2	<i>Cognition and social metacognition in collaborative learning</i> .....	20
2.3.3	<i>Using biodata to understand cognition: physiological arousal and synchrony as indicators of metacognitive monitoring</i> .....	22
2.4	Collaborative learning .....	23
2.5	Conclusions .....	24
2.6	The present study and its research questions.....	27
<b>3</b>	<b>METHODS</b> .....	<b>28</b>
3.1	Research design and procedures .....	28
3.2	Data collection.....	29
3.2.1	<i>Databases and keywords</i> .....	29
3.2.2	<i>Inclusion and exclusion criteria</i> .....	29
3.2.3	<i>Screening process</i> .....	34
3.3	Data analysis and synthesis.....	36
3.3.1	<i>Data appraisals</i> .....	36
3.3.2	<i>Data extraction and analysis</i> .....	37
<b>4</b>	<b>RESULTS</b> .....	<b>39</b>
4.1	The use of PRTs to study social interactions .....	39
4.2	Research question one: How have biosensors and other physiological response technologies been utilized to investigate human social interactions in the field of education? 42	
4.2.1	<i>Research questions 1a and 1b, respectively: what are the focuses of research? How are different biosensors integrated?</i> .....	42
4.2.2	<i>Research question 1c: how are physiological data handled and analyzed?</i> .....	50
4.3	Research questions 2 and 2a: What are benefits and drawbacks of incorporating physiological data into education research? What are the future directions for research according to the current state of the literature? .....	55
<b>5</b>	<b>DISCUSSION</b> .....	<b>69</b>
5.1	Current strengths and future directions of physiological data in education .....	69
5.1.1	<i>Qualitative approach to embodied movement data</i> .....	76
5.2	Methodological challenges in biodata research .....	77
5.3	Data analysis: the state of the art, bias, and methods for bias reduction.....	83
5.3.1	<i>The state of the art and future potential</i> .....	83

5.3.2	<i>Data analysis bias requires recognition</i> .....	85
5.3.3	<i>Example: EDA data analysis in studies of emotion and future potential</i> .....	86
5.4	Data handling bias and methods for reduction.....	89
<b>6</b>	<b>CONCLUSION</b> .....	<b>94</b>
<b>7</b>	<b>FINAL CONSIDERATIONS</b> .....	<b>96</b>
7.1	Limitations.....	96
7.2	Reflections on ethics .....	97
	<b>REFERENCES</b> .....	<b>98</b>
	<b>APPENDICES</b> .....	<b>136</b>
	Appendix A: The title and abstract screening tool adapted from an example tool in Polanin et al. (2019). .....	136
	Appendix B: The decision-tree/algorithm from the Mixed Methods Appraisal Tool (MMAT) to categorize studies) from Hong et al. (2018).....	140
	Appendix C: The unedited Mixed Methods Appraisal Tool (MMAT) from Hong et al. (2018).....	141
	Appendix D: Summaries for studies using physiological and biodata in human social interactions in the education field in the past decade.....	143
	Appendix E: Summaries for studies using physiological and biodata in human social interactions outside of education 2011-2021 .....	162
	Appendix F: Relative frequencies of data analyses techniques .....	211

## TABLES

<b>TABLE 1.</b>	Inclusion and exclusion criteria .....	30
<b>TABLE 2.</b>	Keywords for inclusion, exclusion, and sorting .....	32
<b>TABLE 3.</b>	Revised Mixed Methods Appraisal Tool (MMAT).....	36
<b>TABLE 4.</b>	Acronyms utilized in the present text .....	39
<b>TABLE 5.</b>	Relative frequencies (RF) of major experimental paradigm characteristics in education research.....	42
<b>TABLE 6.</b>	Biodata usage outside the education field .....	55
<b>TABLE 7.</b>	Relative frequencies of data analysis techniques for studies outside education sorted by biodata type and utility.....	211

## FIGURES

<b>FIGURE 1.</b>	Flowchart for screening and study inclusion .....	35
<b>FIGURE 2.</b>	A nested model of issues related to the sources of limitations and biases in studying human social interactions with biodata in education .....	78
<b>FIGURE 3.</b>	Confusion matrix depicting the relations of a physiological (or bio) datum with a CAMMS event .....	79
<b>FIGURE 4.</b>	Physiological data indicates measures of interest <i>and</i> other phenomena.....	82
<b>FIGURE 5.</b>	The divergences of EDA data handling.....	92

# 1 INTRODUCTION

The increasing interest in utilization of physiological measures in social interactions in education has heightened the need to understand the know-how, pros, cons, and future directions of physiological response technology integration. Knowledge of physiological responses in social interactions has a great importance to develop the understanding of collaborative learning.

The use of numerous physiological measures has provided great insight into cognition, emotion, and embodied engagement during social interactions, which have contributed to further understanding social dynamics in learning situations such as the ones happening during collaborative learning. Learners participate in collaborative learning when they jointly build knowledge, which is predicated on embodied social interactions, verbal (e.g., Visschers-Pleijers, et al., 2006) and non-verbal (e.g., Brennan & Hanna, 2009), to share ideas and regulate each other to reach a shared understanding. In conjunction with behavioral data collection, physiological data has been used to inform and constrain theories of social cognition (e.g., Bartholow, 2010). Similarly, physiological responses, such as electrodermal activity (Cartaud et al., 2018) or electroencephalography (EEG) readings (Perry et al., 2017) provide insight into emotional reactions during social interactions. Additionally, biosensing methodology such as facial expression recognition (e.g., Malmberg et al., 2019a) and eye-tracking (e.g., Schneider & Pea, 2013) has benefited the understanding of embodied engagements such as visual attention in social interactions.

However, there have been no systematic literature reviews on the state of the art of the utilization of biosensors and its analysis in social interactions. Furthermore, there is particularly scarce information on the potential benefits and shortcomings of biosensors that can be implemented in and avoided by educational research. Understanding this state of the art in education with reference to literature outside education will help inform future directions of biosensor use and analysis in the education field.

A systematic literature review approach was taken because of its ability to summarize the literature comprehensively and impartially across a wide range of literature and critique the state of the art. Since the research questions concern biosensor methodology in social interactions across fields and their current and potential uses in educational research, a systematic review is an adequate approach. The present systematic literature review focuses on current research (the past 10 years) regarding the use of physiological response measurement technology and methods in social interactions. The review's discussions are oriented to address research questions related to the specific phenomenon in education: collaborative learning. The approach here draws from elements specifically germane to collaborative learning: cognitive regulations, emotional responses to a group, and embodied engagements in tasks. The conclusions describe the advances in the social cognition field that can affect group learning, incorporation of biosensors in educational research methodology, and the gaps to be filled in this field.

## 2 SOCIAL INTERACTIONS AND DYNAMICS IN LEARNING ENVIRONMENTS

In an increasingly connected world, learning often happens in a social milieu. Past decades of educational research have placed a significant importance on understanding social learning and small group cognitive and emotional processes. Piaget's work (2001) largely guided this focus, as he asserted that conflicts of a socio-cognitive nature give rise to meaningful cognitive restructuring. This focus was also brought about by Vygotsky (1978), who postulated that learning occurs at a socio-cultural level before internalization by the learner. Current thought subsumes these two theories under the umbrella of the socio-constructivist theory of learning, which places importance on negotiation of meaning for thinking and learning, and as such, researchers have focused efforts to understand group interactions and identifying features of patterns that underlie collaborative learning.

The last decade of educational research has shown an interest in capturing twenty-first century skills such as collaboration by integrating novel data collection tools as part of a multimodal data set. Advances in the accessibility and suitability of new biosensing devices grant researchers the ability to transcend ontologically flat data. Multimodal datasets provide data across a gamut of cognitive and non-cognitive processes that are otherwise invisible, such as micro-interactions or imperceptible physiological responses (Reimann et al., 2014). Such technologies novel in the field of education measure eye movements, electrodermal activity, brain oscillatory activity, heart rate, embodied movements, and other bio-physiological signals. The ontological divides between the body, mind, and environment are crossed by the processes, including cognitive processes, that occur in learning in social contexts. Thus, interdisciplinary approaches that capture data from embodied

movements, brain activity, and language grants us access to investigate them. For example, current studies have used eye-tracking to reveal the role of joint visual attention and electrodermal activity (EDA) synchrony in quality of collaboration and learning gains (Pijeira-Díaz et al., 2016; Schneider et al., 2020), a link between EDA synchrony and group affect (Mønster et al., 2016) and metacognitive regulations (Ahonen et al., 2018), as well as the relations of inter-brain synchrony (IBS) among students in a classroom to student-teacher social dynamics (Dikker et al., 2017; Bevilacqua et al., 2019).

This review explores the potential of biodata (especially in multimodal data sets) to reveal information on attentional, cognitive, and emotional processes in social learning contexts with a focus on how data triangulation can aid investigation of key features of collaborative learning and address various current methodological limitations. In sum, this review sets the stage for the systematic review by establishing (a) that the utilization of multimodal data (especially biodata) is still in the early stages in the education field, (b) the requirement for interdisciplinary expertise, (c) a gap in understanding how the state of the art intersects with education, and (d) the how and why of triangulation of data from bio-physiological, cognitive, and emotional aspects of human interactions in education research.

## *2.1 Defining visual and joint attention*

This section defines visual and joint attention, reviews the findings of joint attention for collaborative learning, and then explores the designs, methods, and results of studies that utilize gaze data to measure visual attention in the education field. This section will partly set the stage for a systematic review of the literature by showing the need for a deep exploration of the state of the art and identification of potentials and limitations of biodata use.

This literature review addresses visual attention in the education field for individuals and between-subjects, i.e., dyads, triads, and small groups, and notably most of the studies to-date utilize pair-level analysis. For individuals, Anderson (2005) characterized visual attention as “the behavioral and cognitive



process of selectively concentrating on one aspect of the environment while ignoring other things”.

In learning scenarios, visual attention serves as a precise indicator of selective resource processing. During learning involving multiple participants, such as collaborative learning, visual attention of dyads may synchronize, which is termed joint attention (JA). Joint attention is when social partners jointly focus on an external entity, and crucially, monitor each other’s attention to it (Tomasello, 1995). Joint attention plays a key role in the coordination in social interactions. Adult gesture and gaze initiate joint attention with toddlers (Tsuji et al., 2021). Gaze is followed and monitored in dyadic cooperation (Pfeiffer et al., 2012), and joint attention responsivity is a function of gaze communicativeness (Caruana et al., 2020). Deictic gestures have been firmly established as signals to pertinent aspects of the environment (e.g., Bates et al., 1989), yet are contextually coded to the extent that non-informative gestures are deemed as such and ignored (Kajopoulos et al., 2021). The development of JA has been shown to be highly relevant for development of language abilities (Dawson et al., 2004), social competence and information processing, intelligence, and theory of mind (Bruinsma et al., 2004; Charman, 2000; Dawson et al., 2002; Mundy & Newell, 2007; Tomasello & Farrar, 1986).

### 2.1.1 Why eye-tracking would be important: joint attention in collaborative learning

This review is mainly focused on understanding visual attention in the context of collaborative learning (CL); thus, the author turns attention to research using eye-trackers to study joint attention in CL contexts. Richardson and Dale (2005) set the stage with findings that the proportion of gaze alignment is related to listeners’ accuracy on comprehension questions. In a spontaneous dialogue, prior establishment of common ground knowledge positively impacted JA (Richardson et al., 2007). Elevated joint visual recurrence was positively associated with productive collaboration when dyads coded a single entity (Jermann et al., 2011).

Collaborators that share gaze position status could more readily reach common ground (Müller et al., 2013), which can result in more efficient

discourse (Schneider & Pea, 2013; 2015) and can be predictive of collaboration quality within dyads (Schneider & Pea, 2014). Indeed, computer-supported collaborative learning (CSCL) approaches that share gaze position transform gaze into a deictic gesture, which help engage dyad members into cycles of synchronization and desynchronization to produce high quality collaboration (Schneider & Pea, 2015). Nüssli (2009) collected data on the speed and pitch of speech, in combination with eye-tracking data, to find predictors for subject success (used with up to 91% accuracy). Chanel and colleagues (2013) found that gaze synchrony when the partner's gaze was visible can accurately predict emotion management and reaching consensus, which are both crucial for successful collaboration. Cherubini and colleagues (2008) designed an algorithm to show an increased likelihood of misunderstandings based on the distance of emitter and receiver gaze. Brennan and colleagues (2008) found that JA and further synchronization between individuals greatly improves collaboration compared to isolated speech or other collaborative conditions. Importantly, JA seems to be highly beneficial but not always sufficient for collaboration; Schneider and colleagues (2018) qualitatively compared two dyads with high JA but maximally different learning gains and found that reactions to proposals varied greatly. Taken in sum, those findings evidence the key role JA, and overall, within-dyad synchronization, have in collaboration. Additionally, the studies show the promise of eye-tracking technologies for assessing and impacting factors shaping collaboration.

### 2.1.2 Gaze data: the current state of designs and analytic methods

Investigations of visual and joint attention employ various experimental designs and methods to capture eye movements. The apparatuses most used are mobile eye-trackers (glasses) and computer-equipped eye-trackers (stationary sensors), with a single recent study employing the combined use of glasses and web camera (Thepsonthorn et al., 2016). Schneider & Pea (2013) utilized a setup combining eye-tracking and a tool to visualize partner gaze on computer screens to track dyad gaze pattern. The measures of interest gathered from the data corpus were saccades, fixations, joint attention measure (joint fixation for >2 s), and pupil dilation. Subjects worked in pairs to learn new information

from text and diagrams, for which areas of interest (AOIs) were defined. Leader and follower gaze dynamics were analyzed contextually based on qualitative analysis of student coordination, convention creation, hypothesis building, and theory sharing. Using data from that study, Schneider & Pea (2014) combined network analysis of visual attention and machine learning algorithms to predict collaboration quality. With the same data corpus, Schneider & Pea (2015) explored linguistic convergence (e.g., grammatical structure mimicking) and coherence (word repetition) of dyadic verbal exchanges over time to determine another layer of synchrony that can be analyzed with JA. Schneider and colleagues (2018) used head-mounted eye-trackers (eye-tracking glasses) to measure joint attention during a group learning session, with measures of interest being fixations and joint fixations as they relate to learning gains and quality of collaboration (as rated from Meier et al., 2007). Cross recurrence graphs aided visualization of JA synchronization, augmented with speech information. As well, the leadership component of JA was analyzed in terms of its relation to learning gains. Qualitative analysis provided insight into the role of gestures and speech on measures of interest.

A handful of research groups have also begun to collect and analyze multimodal biodata, which allows for biodata triangulation. In this way, biodata streams can be compared in relation to the same measure. For example, the relations of EDA synchrony with joint attention could be explored to better understand the dynamics of synchronicity during social learning activities. Or, they may be used in a complementary manner, such that one data stream is treated as a factor for another. In practice, eye-tracking data could inform the precise gaze location on a partner's face, and EDA data could provide arousal state data.

In recent work from Schneider and colleagues (2020), electrodermal activity (EDA), blood pulse volume (BPV), and body movements (Kinect-tracked whole-body motions) were co-collected with eye-tracking data to analyze their relations in the materialization of physiological synchrony (PS) in dyadic collaboration. Specifically, within-dyad Pearson's correlations of EDA were used as indexes of PS to identify markers of learning gains. EDA oscillation events were matched with videos and transcripts to qualitatively analyze behaviors and perceived intentions of the discourse. However, the eye-tracking data in this

study has yet to be analyzed; only qualitative accounts of attention were provided to contextualize EDA synchrony for two groups.

In a conference paper, Chanel and colleagues (2013) utilized computer-equipped eye-trackers in combination with EDA, blood pressure, respiration amplitude, and heart rate measures to assess the degree of physiological coupling and joint attention during a computer-based collaboration. From surveys of self-rated collaboration quality, eight factors of collaborative interactions were investigated in terms of relation to JA and coupling data.

Another study's goals were to examine the relations between five physiological coupling indices (PCIs) and three measures of collaboration: collaborative will, collaborative learning product, and dual learning gain (Pijeira-Díaz et al., 2016). PCIs were computed from pair-wise EDA timeseries data. Regression analyses between the PCIs and measures of collaboration were performed to reveal potential predictors of measures of collaboration. Though they concurrently collected data from eye movement, EDA, temperature, heart rate, and wrist acceleration, only EDA-based PCIs were analyzed as predictors of measures of collaboration.

Thepsoonthorn and colleagues (2016) assessed the relations between measures of mutual gaze convergence and head nodding synchrony and prior knowledge state. Head-mounted cameras and accelerometers were used in teacher-student dyads during instructional sessions. Mutual gaze convergence and head nodding synchrony were computed as the measures of interest and tested as potential predictors of prior knowledge state. However, these measures were analyzed separately in terms of their interactions with prior knowledge, with no analysis of their potential co-relations.

With the above work taken together, integration of multimodal physiological and biosensor data (Chanel et al., 2013; Pijeira-Díaz et al., 2016; Schneider et al., 2020; Thepsoonthorn et al., 2016) is at an early stage of development that often misses opportunities of triangulation, exploration of co-relations, and complementary use. Properly taking advantage of multimodal (bio)datasets will aid to achieve the aim of many of the above studies, which includes the aspiration to bring the field closer to a dashboard of real-time assessment and intervention in collaborative learning. As well, the work above also highlights the value of contextualizing quantitative gaze data with

synchronized qualitative analysis of speech and behaviors. While quantitative pre- and post-tests or reports aid informing the knowledge status and self-reported quality of collaboration, content analysis of dyads during biodata-marked events allows a deeper understanding of CL dynamics as they unfold over time (cf. Chanel et al., 2013 vs. Schneider et al., 2020).

## *2.2 Emotions in interactions in education*

Besides attention, emotion is another key element of learning in social settings and specifically collaboration. This section defines emotion, provides a background of socioemotional group interactions, elaborates on the value of multimodal data for emotion in learning research, and discusses the use, findings, and challenges of two major bio-measures (EDA and facial expression recognition) as quantitative measures of emotion.

Emotion has been conceptualized varying over many disciplines but can generally be regarded as “a feeling that is often short-lived, intense, and specific” (Artino et al., 2012). In education (and psychology) literature, emotion is an affective state, rather than a trait (Artino et al., 2012). Further, emotion can be thought of as varying in two dimensions: valence (positive or negative) and activation (activating or deactivating) (McConnell & Eva, 2012). Emotion plays crucial roles in education, especially in collaborative settings (Schutz & Pekrun, 2007; Järvenoja and Järvelä, 2009), to positive or negative effect (Imai, 2010). In individual and group scenarios, learning and performance are complexly and dynamically connected with emotion through cognition and motivation. Essentially, positive emotions allow large-scale perspectives, and negative emotions fine-scale ones, implying both valences benefit learning and performance in a context-dependent manner (McConnell & Eva, 2012).

Intensive human social interactions are required by educational activities such as collaboration. Since collaboration happens in groups, this review will turn to research in education to describe the landscape of emotional and cognitive interactions and regulations, introduce the value of quantitative, multimodal data, and explore electrodermal activity and facial expression recognition analysis in depth.

### 2.2.1 Why study socioemotional group interactions?

Attainment of shared goals requires effective coordination of group effort and resources. This coordination of group behavior is termed group regulation. Group regulation is a crucial set of cognitive interactions that include behaviors such as setting of goals and tasks, monitoring, and assessing methods and outcomes (Saab, 2012). These processes need to go smoothly for effective group functioning and are aided through socioemotional interactions (Webb & Palincsar, 1996). These interactions are defined as actions regarding emotional expression in a social setting such as “getting to know each other, committing to social relationship, developing trust and a sense of belonging, and building a sense of on-line community” (Kreijns et al., 2003). Demonstrations of respect, getting along, and support are positive socioemotional interactions.

Positive socioemotional interactions between agents facilitate effective collaborative learning (Kreijns et al., 2004), learning outcomes (Richardson & Swan, 2003), and satisfaction with collaboration (Bulu, 2012). Abilities to effectively regulate emotions in interactions are strongly tied to the success of collaborations (Lopes et al., 2005; Xolocotzin Eligio et al., 2012). During collaborative learning, students interact emotionally to build a positive group climate, trust within the group, interpersonal relationships, and a sense of community. Positive socioemotional interactions benefit group performance by allowing greater expression of disagreement; opinions that are exploratory, divergent, or critical are welcome and result in improved collaboration (Janssen et al., 2009). Feeling of belonging improves engagement in collaboration and motivation to collaborate (So & Brush, 2008), positive group climate increases helpful interactions (Kwon et al., 2013), and member accountability and commitment develop trust (Tseng & Yeh, 2013), which can be corrupted by unexcused violation of group norms. Sense of community, or the feeling of a trustworthy authority structure, awareness of the mutual benefits of togetherness, and the spirit that results from shared experiences (McMillan, 1996), is affected by socioemotional interactions. Emotional safety, feeling of belonging, and trust are socioemotional components necessary for sense of community (McMillan, 1996).

In the education settings, instructors may erroneously assume that positive socioemotional interactions happen automatically (Kreijns et al., 2003), which may lead to unattended struggle or isolation (Johnson et al., 2002). Since the quality and outcome of collaboration depend on those socioemotional factors, studies are beginning to employ tracking, assessment, understanding, and intervention in socioemotional group dynamics.

### 2.2.2 What does biosensing offer? The value of multimodal data for emotions in learning

There is a strong and rich effect to which qualitative methodologies can interpret actions and behaviors (Hsieh & Shannon, 2005; Elo & Kyngäs, 2008), yet alone, these methods used to interpret explicit behaviors present significant uncertainties (Elo et al., 2014). For example, implicit group factors such as emotional contagion and affect infusion may significantly impact cognitive processes (Okon-Singer et al., 2015), yet eludes full investigation due to data constraints and difficulties of measurement (Fujiki et al., 2002).

Of the numerous factors that influence collaboration quality, emotion is among the most salient and modulates behaviors that are measurable physiologically (Balters & Steinert, 2017). Quantitative measurements of changes in the autonomic nervous system through monitoring of physiological and neurophysiological signals including EEG, EMG, MEG, fNIRS, fMRI, ECG, and EDA can detect emotional state (e.g., Agrafioti et al., 2011; Boucsein, 2012, Ramirez & Vamvakousis, 2012; Kim et al., 2004). Multimodal data are of particular interest to capture the effects of socioemotional interactions (Heaphy & Dutton, 2008; Mønster et al., 2016). Multimodal data may be subjective or objective. For instance, repeated self-reports provide subjective data that may offer insight into student intentions or perceptions behind learning. On the other hand, objective data on emotions in interactions such as automatic facial expression recognition and physiological measures from EDA provide continuous information about affective states and their valence (Harley et al., 2015; Ahonen et al., 2018).

### 2.2.3 Example 1: EDA as an index of emotional state in education

Spikes in EDA are fast (1-3 s post stimulus) and robust and allow for automatic and unobtrusive evaluation of affective states (Picard et al., 2001). Emerging research in education (e.g., Harley et al., 2015; Harley et al., 2019) has investigated the physiological arousal with the understanding that it represents the arousal dimension (activating vs. deactivating) in the circumplex model of emotion (Russel, 1980), but with few linking it to valence (positive vs. negative) (e.g., Ahonen et al., 2018; Pijeira-Díaz et al., 2019).

The utilization of EDA to examine emotions in collaboration is growing. The accessibility of wearable, mobile physiological sensors allows investigation of physiological markers relevant to emotion during collaboration as they develop. Leveraging multimodal data from numerous sensors including wrist EDA sensors can aid in (60%) prediction of emotional state during interactions (Arroyo et al., 2009). Mønster and colleagues (2016) examined collaborative work during an experimental task and found that synchronous arousal was positively associated with self-reported negative affect, group tension, and a feeling of non-belonging to the group. Chanel and colleagues (2013) attempted to seek out relationships between explicit emotion sharing in a CSCL context and EDA synchrony but did not report results. Ahonen and colleagues (2018) examined event-locked EDA synchrony from dyads engaged in self-directed programming tasks and found emotional valence was detectable, mainly in a role-dependent manner. Specifically, anticipation of evaluative phases elicited arousal or relaxation depending on work quality, post-evaluative responses were deactivating (relief) or activating (frustrated), and leaders exhibited greater responses (Ahonen et al., 2018). To the author's knowledge, only Ahonen and colleagues have detected emotional valence from EDA signals alone, which typically only serves as an indicator of activation (Kreibig, 2010). Sub-group valence responses measured from EDA responses were highly contextual, and varied according to time, work feedback, and subject role. Study designs with similar paradigms of collaborative work followed by positive or negative feedback may also be able to uncover valence from EDA alone; nevertheless, the generalizability of this methodology is limited.



Given the importance of this dimension to identify emotions (Fontaine et al., 2007), more robust methodologies are required for determining valence across a range of paradigms. Recent work leverages facial expression recognition technologies to determine valence in conjunction with EDA analysis (Malmberg et al., 2019a).

#### 2.2.4 Example 2: Utility of facial expression recognition to measure emotional valence in education and its current challenges

Facial expression is a crucial data channel for automatic detection of emotions (Azevedo, 2015). During collaboration, social-related and task-related factors may elicit facial expressions that reflect positive (e.g., excitement, interest, engagement) or negative emotions (e.g., confusion, frustration, boredom, anxiety) depending on the context (D'Mello et al., 2011). However, only until recently have researchers investigated the quality of social interactions during collaborative learning with use of physiological sensors combined with facial recognition data (Malmberg et al., 2019a).

In collaborative settings, recent studies are adding face expression analysis to physiological signal analysis to understand the valence dimension of affect. One study explored the connections between observed affective states of groups and physiological synchrony of group members (Törmänen et al., 2021). Manual coding of videos was time-matched with EDA data in a mixed-methods approach. Subjects were more likely to display emotions during simultaneous arousal, and divergent affective states were observed due to social triggers rather than task-related factors. Manual coding of facial expressions is often reliable due to reaching 100% consensus (Linnenbrink-Garcia & Pekrun, 2011), as done by Törmänen and colleagues (2021). Inter-rater reliability analysis provides another method to boost reliability, though some expressions in videos remain hard to classify, with low inter-rater agreement (Holkamp & Schavemaker, 2014). As well, manual coding is labor intensive and lacks potential for significant iterative improvement compared to machine systems (Keskinarkaus et al., 2016).

Automatic facial expression recognition systems such as the mobile multimodal recording system (MORE) attempts to address these issues, which

processes live 360° video to extract facial valence (Keskinarkaus et al., 2016). Using the MORE system, Malmberg and colleagues (2019a) collected observational video and speech data, EDA, and facial expression data from post-processed videos on dyads collaborating to explore the relations between simultaneous arousal episodes and facial expression valence. They found that negative facial expressions occurred nearly twice as much (40%) during synchronous arousal than positive ones, and that among interactions during EDA synchrony, confusing interactions had the greatest amount of negative valence. Importantly, although Keskinarkaus and colleagues first reported a valence recognition rate of 70.8%, Malmberg and colleagues (2019a) have achieved a 96.26% recognition rate using the Cohn-Kanade facial expression database. Comparison of human and machine emotion recognition *in situ* would have provided unique insight into the MORE system's reliability.

The databases used for assessment cause for a large variation in recognition accuracy, as Huang and colleagues' (2016) training on the MAHNOB-HCI database resulted in roughly 50% recognition accuracy using spontaneous facial expression recognition alone. Practically, functional systems or models of emotion recognition remain under development and have faced obstacles of accuracy due to cross-subject variability (Yin et al., 2017) and movement when not in controlled lab settings (Xu et al., 2017). Wei and colleagues (2017) argue that to achieve successful classification of affective state, sophisticated models must rely on response induction and controlled protocols and are not yet suited to naturalistic settings of social interactions in learning. Though the machine learning systems above may not always provide accurate single-face classifications, the promise lies in iterative improvements. Though accuracy seems low, Huang and colleagues (2016) achieved similar classification rates to human raters using facial data alone, and better performance when combined with EEG data. According to Huang and colleagues, real-time valence classification can be performed with EEG data at comparable classification rates to human raters. However, it must be noted that Huang and colleagues' study used static images to compare human and machine performance, where humans are at a disadvantage due to lack of contextual information.

Crucially, multimodal data provides a more reliable interpretation compared to single-channel data (e.g., facial emotion recognition, EEG, EDA, or self-report alone) (Harley et al., 2015; Huang et al., 2016). Harley and colleagues (2015) assessed agreement between three synchronized measurements of emotion (i.e., EDA, automatic facial expression recognition, and self-report) while subjects learned in a multi-agent computerized environment. They synchronized facial expression and EDA sensor data with markers from self-reported emotional states and found high agreement between self-report and facial data but low agreement between all three data modalities. Thus, various components of emotional response may not always be tightly coupled.

### *2.3 But what about cognitive processes?*

To this point, this review has defined concepts and discussed the use, findings, benefits, and challenges of investigation of visual attention and emotional processes with biodata. But how is cognition conceptualized and investigated with physiological response technologies? How can bio-based measures of cognitive processes be disentangled from attentional and emotional ones? Unsurprisingly, these processes of interest are inextricably linked. Indeed, research involving emotion (e.g., Scherer & Moors, 2018) recognizes the role of cognitive activity in stimulating physiological arousal (e.g., EDA) in CL contexts. Further, researchers have found evidence for a direct link of cognition to emotion through interception and appraisals (Critchley & Garfinkel, 2017).

The aim of this section is to review the state of using biodata in education for investigation of cognitive processes. This section defines cognition, introduces perspectives on cognition, explores the paradigms, findings, and challenges associated with metacognitive monitoring research that utilizes biodata, focuses on how EDA arousal and synchrony are linked to metacognitive events, and ends with discussion of limitations.

### 2.3.1 What is cognition?

The definition of cognition and the search for a unique “marker” of the cognitive is a matter of philosophical debate, and perhaps, little importance to the operation of science (Allen, 2017). Broadly, it refers to the mental processes or action of knowledge acquisition and understanding through sensing, experiencing, and thinking. Processes encompass but are not limited to attention, perception, comprehension, knowledge and memory formation, reasoning, problem solving, decision making, and language production. Studies on cognitive processes in psychology and education can focus on the individual or the group, the latter of which is conceptualized as group or socially shared cognition (e.g., Szanto, 2014). Though this is a matter of philosophical debate (cf. Szanto, 2014 vs. Ludwig, 2015), it is outside the scope of this paper.

### 2.3.2 Cognition and social metacognition in collaborative learning

In educational settings and activities, students often ask questions, share perspectives, assess, and elaborate them when they collaborate (Vygotsky, 1978; Chi, 2009). One way to think about the cognitive activities in collaboration are low and high cognition (Cohen, 1994; Kempler & Linnenbrink, 2006). Low cognition includes reading, listening, and other information processing activities that acquire knowledge, whereas high cognition is marked by the generation, elaboration, and critique of ideas, as well as the relation of new ideas to prior knowledge (Volet et al., 2009). High cognition is associated with learning in CL contexts, and low cognition aids students in establishing common ground (mutual understanding) (Clark & Brennan, 1991), which supports higher cognition and hence effective CL (Volet et al., 2009).

Another perspective on cognition that has yielded substantial research in education places focus on metacognition. Cognitive activities involve task-related content. Conversely, metacognitive activities orient, plan, monitor, evaluate, and reflect on cognitive activities (Meijer et al., 2006). Monitoring gathers and processes relevant information for planning actions to guide actions of self or other collaborators. Some students have been found to monitor poorly in isolation, which can hamper planning, especially in poorly structured

environments (De Bruin et al., 2011). During collaboration, triangulation of multiple students' monitoring activities can surmount this challenge. Members of the group regulate joint cognitive processes with social metacognitive activities to improve focus, construct common ground, facilitate shared representations, and inhibit erroneous constructions and conceptualizations (Iiskala et al., 2011; Hadwin & Oshige, 2011). Metacognitive activities and (content-related) cognitive activities are interwoven, as they require the same cognitive resources and occur continuously. Hence, this feed-forward, cyclical nature of social metacognitive activities explains why groups engage in greater regulation when faced with challenge (Winne et al., 2013).

The regulation of learning is dynamic and as such, the temporal occurrence of metacognitive monitoring is not set. Metacognitive monitoring can activate after every phase of regulated learning (Sonnenberg & Bannert, 2016). In another study, when primary school student triads worked on a writing task, they showed metacognitive activities in CSCL with metacognitive scaffolding increased compared to structural or no scaffolding conditions, but metacognitive activities across groups were temporally homogenous (Molenaar & Chiu, 2014). In a similar study, Su and colleagues (2018) showed that college students collaborating with scaffolds performed better, potentially due to ordered monitoring of content, organizing, and monitoring of the process compared to groups with low performance who redundantly organized.

While sometimes successful, interventions do not equally boost metacognitive monitoring for all students (Järvelä & Hadwin, 2013). Despite its importance to support CL (Järvelä et al., 2016), the education field is still developing methods that capture "invisible" metacognitive monitoring of individuals and groups (Järvelä et al., 2021). Further, delivering helpful interventions in a timely manner is another concern, which could be addressed by approaches that use physiological data as markers of metacognitive monitoring (Järvelä et al., 2021) and machine learning algorithms for real-time application (see e.g., Huang et al., 2016).

### 2.3.3 Using biodata to understand cognition: physiological arousal and synchrony as indicators of metacognitive monitoring

In collaborative learning, metacognitive monitoring has traditionally been explored by temporal characterization of interactions, assessing interaction quality, assessing individual contributions to group cognitive activity, and utilizing think aloud methods (e.g., Malmberg et al., 2017; Volet et al., 2017). While videos non-intrusively provide data on verbal and non-verbal interactions, they may lack the ability to capture subtle non-verbal behaviors and are severely limited in capturing cognitive activities “under the hood”. In addition to eye-tracking, body movement tracking, and neurophysiological measurements, many researchers in education are turning to physiological arousal and synchrony as potential tools due to ease (EDA wristbands) and relations with mental activities (Palumbo et al., 2016) such as sharing in CL contexts (Järvelä et al., 2021) and metacognitive monitoring (Hajcak et al., 2003). Recent research using these measures is relatively unobtrusive and has advanced the field’s understanding of CL and its moderating factors (Winne, 2019).

In terms of cognitive processes, arousal has a dependent relationship with attention levels (Sharot & Phelps, 2004), engagement tied to high mental effort (Fritz et al., 2014), and task-dependent cognitive load (the cognitive demands) (Fairclough et al., 2005). Past research suggests physiological arousal due to individual (Hajcak et al., 2003) and joint (Ahonen et al., 2018) monitoring activities. Malmberg and colleagues (2019a), discussed above for their findings of emotion-based arousal, also noted the involvement of markers of metacognitive monitoring. Malmberg and colleagues (2019b) analyzed individual and synchronous arousal of university students during a collaborative exam and found no significant relationship with metacognitive monitoring events. Since video cannot capture all metacognitive processes, for example if members were confused but did not verbally or non-verbally express this, arousal synchrony would not be matched with any observable metacognitive monitoring event, which is a false negative result. Automatic facial emotion recognition data may provide insights by tracking subtle emotional variations (i.e., micro-expressions). A deeper qualitative analysis showed the groups that exhibited the highest levels of synchrony verbally expressed more difficulties.

Malmberg and colleagues hypothesized the lack of observed connection between arousal and metacognitive events could be due to arousal's stronger link with anticipation and cognitive load (Critchley, 2002) or low-level interactions, instead of high-level interactions that involve metacognitive monitoring (Malmberg et al., 2019a). Dindar and colleagues (2019) found that students with similar self-evaluations of prior task-related knowledge displayed more arousal synchrony during a collaborative task, which may be a result of similar cognitive load or other dynamics within collaboration. Another study also found that difficulties encountered during collaboration may have been responsible for synchrony (Mønster et al., 2016), and work by Haataja and colleagues (2018) showed similar results, but also a weak yet statistically significant positive correlation between physiological synchrony and monitoring.

These findings indicate that physiological synchrony may be most informative to investigate monitoring during collaborative difficulties. Also noteworthy is the possibility of false positive results, where arousal accompanied a measure of interest, but was confounded by another stimulus. For example, an arousal event could have been caused by the cognitive load, not the metacognitive monitoring, of a verbalization of a complex concept to the group. Triangulation with indices of cognitive load, such as pupillary dilation (Matthews et al., 1991) and heart rate variability (Solhjoo et al., 2019), would improve the interpretability of results. Future studies that utilize more suited data channels and apt paradigms may disentangle factors (Azevedo, 2015).

## *2.4 Collaborative learning*

This thesis orients the discussion of biodata in education towards applications in collaborative learning contexts. The conceptualization and landmark findings related to collaborative learning are covered.

The concept of collaborative learning in this paper borrows from the framework put forth by Roschelle (1992) of convergent conceptual change. From this view, collaboration involves grounding, a process wherein participants share their construction of meaning for concepts, experiences, and conversations, which has been an area of deep investigation (Clark & Wilkes-Gibbs, 1986). For example, from the psycho-linguistic perspective, constructing

a common ground ensures that agents achieve mutual understanding and shared meaning and conceptualization of language used. The education field takes grounding further to focus on shared meaning making (Stahl, ,2007), which involves a transition from mutual understanding to understanding “the meanings of the semiotic tools that constitute the mediators of interpersonal interaction” (Baker et al., 1999, p. 31). Through this process, new meanings and understandings are co-constructed and lead to conceptual change. Research has identified concrete mechanisms that enable grounding and joint meaning making, such as apt articulation and clarification of thinking (Webb et al., 1995), restructuring of understandings to reveal the extent and limits of their knowledge (Cooper, 1999), partaking in processing to explain ideas through augmentation of their partner’s (Damon, 1984), jointly constructing ideas (Webb & Palincsar, 1996), and providing detailed arguments to negotiate meaning and resolve disagreements (Baker, 2003). Crucially, collaborating group members must contribute equally for these behaviors to take effect, as free riders prevent high quality collaboration (Salomon & Globerson, 1989). Damon and Phelps (1989) theorized that mutuality, or equal participation among group members, is a prerequisite for successful collaboration.

Recent educational research has employed technologies that gather physiological response and biodata to understand and support collaboration. Advances in these technologies and research methodologies that employ them may provide accurate and reliable means to capture the processes such as cognitive and emotional processes in social learning settings.

## *2.5 Conclusions*

There are multiple viable tacks to improve the current understanding of biodata (i.e., physiological response data, measurements of embodied movements, etc.) in relation to cognitive and emotional processes in CL, and more generally social interactions in education. First, other measurements of the body related to emotional and cognitive activity, from eye-movements to facial expressions to neurophysiology, should be further investigated at individual and group levels. Thus, research focusing on unpacking the relations between biodata and



cognitive and emotional factors in CL will contribute to the field's understanding (see Schneider et al., 2020).

Second, scant application of theoretical models of social interactions in CL results in an oversimplified interpretation of results. Take the popular concept of physiological synchrony as an example. Recent studies suggest a complex relationship between physiological synchrony and emotional and cognitive processes and CL measures (e.g., quality, learning outcome) that varies over time and is at least group-, task- and context-dependent. Synchrony may undulate in a cyclical manner in groups with high quality collaboration (Schneider et al., 2020), be high in groups with low quality collaboration (Schneider et al., 2020), or be related to the phases in regulated learning processes (Malmberg et al., 2019b). Of the articles reviewed, only a few utilized a theoretical model, the most frequent of which was self-regulated learning (Hadwin et al., 2017), to ground investigations of physiological measures in CL (Haataja et al., 2018; Järvelä et al., 2021). Use of video to qualitatively identify learning phases, while difficult, could contextualize arousal during learning in terms of relevant theoretical models (Malmberg et al., 2019b).

Additionally, exploratory research that utilizes novel combinations of multichannel data such as speed and pitch of speech (Nüssli, 2009) and natural language processing (NLP, Schneider & Pea, 2015) are necessary to contextualize biodata. Multimodal data is necessary because novel data modalities alone fall short of offering direct measurement of cognitive or emotional activity.

While multimodal data offers promise, there are concurrent challenges, which require systematic empirical research to resolve. Multimodal data trace a gamut of processes, from cognitive to emotional and more (e.g., motivational, behavioral), which coincide and overlap. Robust theoretical approaches and nuanced understanding of concepts are prerequisites for effectively analyzing and inferring information from the data (Wise & Shaffer, 2015). The highly interdisciplinary nature of the field coupled with the big data that accompanies biodata collection combine to create large hurdles in handling and analysis (D'Mello et al., 2017). To progress the field with the use of big, multimodal data, multidisciplinary collaboration with experts in data-driven analytical methods, such as data mining or learning analytics is crucial (Gašević et al., 2015).

Examples of recent research that has considered these constraints and prerequisites are machine learning technologies that give potential for real-time scaffolding for learners (D'Mello et al., 2017) and new bio-physiological indices of intersubjective processes to research the multi-leveled group regulation processes (Knight et al., 2016; Wallot et al., 2016).

In sum, the integration of interdisciplinary methodologies to investigate processes and regulations of learning that trace features such as physiological measures, neurophysiological measures, eye-movement, body movement, self-report data, and video data of CL processes make visible otherwise nearly imperceptible processes. The novel approach is to leverage the triangulation of multimodal data to characterize and evaluate individual, intersubjective and group-level cognitive and non-cognitive dynamics in collaborative learning contexts objectively and subjectively.

However, many current paradigms fall short of convincingly accounting for false positive and false negative findings, biodata interpretation dissimilarities (e.g., valence, cf. Ahonen et al., 2018 vs. Malmberg et al., 2019a), lacking application of theoretical models, and inconsistent findings that may result from differing contextual factors or data handling and analysis (Winne, 2019). Though there are reviews of eye-tracking technology in certain fields (e.g., Sharafi et al., 2015) or of physiological elements to education (e.g., Roos et al., 2021), a systematic review of physiological response technologies utilization, benefits and shortcomings in the education field has yet to be conducted.

A systematic review of the current literature in education will give more insight into the focuses of research, biosensor integration, and data handling and analysis. Fields outside of education, such as psychology, neurophysiology, and affective neuroscience have thus far provided tools and techniques for biosensor work in the education field, yet the remaining potential contributions are much greater (Baker & Siemens, 2014), and integration is at an early developing stage (Winne, 2019). A systematic review of literature inside and outside education that lays out the state of the art, then, is an adequate means to situate the discussion on the benefits and drawbacks of biodata use and its future in the education field.

## 2.6 *The present study and its research questions*

Understanding the importance of the body in human social interactions during learning has increased as research interest and theories multiply. Indeed, the recent intersection of physiological response technologies with education enables even richer comprehension of the situation. Thus, this study employs a critical theoretical approach to understand the uses, integration, focuses, and analysis methods of biosensing in education, and the challenges and advantages of incorporating physiological data into education research.

1. How have biosensors and other physiological response technologies been utilized to investigate human social interactions in the field of education?
  - 1a. What are the focuses of research?
  - 1b. How are different biosensors integrated?
  - 1c. How are physiological data handled and analyzed?
2. What are the benefits and drawbacks of incorporating physiological data into education research?
  - 2a. What are future directions for research according to the current state of the literature?

This systematic literature review can provide an overview of the conceptualizations, methodologies, and findings in the education field that utilize biodata in human social interactions. This study may help to identify the future directions for research according to the state of the art in this multi and interdisciplinary sphere.

# 3 METHODS

## *3.1 Research design and procedures*

The present study is a systematic literature review (SLR). An SLR is a review that systematically identifies, appraises, and synthesizes the conclusions of primary research relevant to the research questions (Boland et al., 2017; Zawacki-Richter et al., 2020). A SLR can summarize and identify common concepts in the state of the art by choosing literature that meet established criteria (Oxman, 1994; Eriksson & Lindstrom, 2005). The analysis and subsequent conclusions of a systematic literature review tackle research questions in a holistic and objective manner through rigorous and systematic methods to address pre-determined research questions (Zawacki-Richter et al., 2020). Thus, a SLR was identified as an adequate approach to describe biosensor research methodology in social interactions, identify its strongest and weakest points for application in the education field and current gaps.

. The present systematic literature review made use of the PRISMA guidelines and flow diagram. The evidence-based guidelines put forth in the PRISMA statement (Liberati et al., 2009) include a 27-item checklist and four-part flow chart that indicate elements required to conduct a transparent literature review. The search was narrowed according to inclusion and exclusion criteria to retrieve and analyze only literature relevant to the research questions of this review. The author established the systematic review protocol to address the research questions before conducting the review to minimize researcher bias. The data collection was performed in November 2021 according to the process described next. The research questions guided literature investigations to understanding the methods and focuses of biosensors and other physiological response technologies in human social interaction in education, as well as their benefits and drawbacks.

## 3.2 Data collection

### 3.2.1 Databases and keywords

The following electronic databases were utilized for a literature search: Education Collection (ProQuest), Education Resources Complete (ProQuest), Teacher Reference Center (ProQuest), PsycINFO (Ovid), PubMed (NCBI), and ScienceDirect (Elsevier). Literature searches were also conducted in the American Journal of Physiology – Advances in physiology education to capture the recent uses of eye-tracking in education. To ensure thoroughness, Google Scholar was also utilized to safeguard against missing relevant articles. According to the recommendation of Haddaway et al. (2015), the abstracts of the first 300 articles were screened. Three article(s) was/were identified that had not resulted from other database searches.<sup>1</sup> The databases selected encompass a multidisciplinary and interdisciplinary sphere including education, psychology, and physiology that pertain to this study's inquiry to the utilization of physiological response technologies in human social interactions in the last decade.

The search terms were gathered through preliminary literature searches to determine the keywords used in the fields. The present study investigated the physiological response research in human social interactions in education and in other fields. Thus, two groups of terms were used. The keywords that relate to physiological response research were *eye-tracking*, *emotion recognition*, *electrodermal activity*, *electroencephalography*, *frontal EEG asymmetry*, and *skin conductance*. To form a search string, all keywords were individually searched while paired with *social interaction* via the Boolean operator AND. The study set an expansive range of search terms to retrieve a wide scope of articles related to its aims.

### 3.2.2 Inclusion and exclusion criteria

To give grounds for interpretation and weight to this study's findings, the inclusion and exclusion criteria are communicated (Boland et al., 2017). The strength and accuracy of the conclusions are directly impacted by the quality of the studies

---

<sup>1</sup> The last search was run on November 19, 2021.

included (Oxman, 1994). To set a minimum standard of quality that allowed for adequate rigor of methodology, only peer-reviewed articles were included (Wilder, 2014). To capture the state of the art of physiological response technologies and data in social interactions, only articles published on or after 2011 were included. Only studies with data that measured or addressed physiological states or responses were included. Only studies that included human social interactions were included, including rare cases where humans were deceived in believing their partner was human. Studies of humans interacting with machines (hardware and/or software) or non-human animal subjects were excluded. Non-empirical studies were excluded.

At this point, studies were categorized to address distinct research questions. Qualitative-only studies were separated into a pool to address research questions 2 and 2a from the qualitative perspective. The remaining quantitative and mixed methods studies were then segregated based on the field of research: in or out of the education field to address research questions 1-1c and 2-2a respectively (see Table 1 for inclusion criteria).

**TABLE 1.** Inclusion and exclusion criteria

<b>Criterion</b>	<b>Included</b>	<b>Excluded</b>
<b>Resource Type</b>	Peer-reviewed articles	Books, grey literature, peer-reviewed articles that cannot be accessed as full texts, peer-reviewed articles that are corrections
<b>Publication Date</b>	2011 or later	2010 or earlier
<b>Data Source</b>	Physiological responses and states	
<b>Experimental Subjects</b>	Humans, humans paired with machines	Non-human animals, non-human animals paired with humans, machine only
<b>Setting for Subjects</b>	Social interactions, multiple subjects interacting	Individual, isolated human subjects without interaction
<b>Study Methodology</b>	Quantitative, mixed methods, and qualitative studies	Non-empirical studies

---

<b>Research Field</b>	All; education field literature analyzed separately
-----------------------	---

---

Exclusion key words were used to indicate non-human studies because human studies often did not indicate human to human interactions in the abstract. Manual checks were performed on the stage to exclude studies including non-human subjects to ensure exclusion was warranted. Table 2 indicates the exact keywords used for inclusion, exclusion, and sorting processes of the gathered literature via the abstract screening tool. Manual checks were performed on sorting into the qualitative-only pool because mixed methods studies also may have contained these key words. Manual checks were performed on sorting into the education field pool because key words (e.g., tutor) may not guarantee the study was in the education field.

**TABLE 2.** Keywords for inclusion, exclusion, and sorting

Abstract screening step	Screening Aim	Manual check (Y/N)	Indication	Keywords
3	Inclusion	N	Physiological response or state	“eye tracking”, “gaze tracking”, “gaze convergence” “visual attention”, “joint attention”, “joint visual”, “pupillary diameter”, “face emotion”, “facial emotion”, “facial expression”, biosensor, biosensing, biodata, “physiological response”, “physiological measures”, “physiological markers”, “physiological data”, “physiological linkage”, “physiological state”, “physiological activation”, “interpersonal physiology”, “physiological arousal”, “arousal contagion”, “physiological synchrony”, “physiological coupling”, “physiological compliance” ECG, EDA, GSR, EEG, IBS, INS, HR, HRV, IBI, “heart rate”, “inter-beat interval”, “inter-brain synchrony”, “inter-neural synchrony”, “neural synchronization”, fNIRS, “functional near-infrared spectroscopy”, hyperscanning, “electrodermal activity”, “electroencephalography”, “skin conductance”, “trace gesture”, “gesture detection”, “gesture-based”, “multimodal data”
4	Exclusion	Y	Non-human subjects	mice, mouse, rat, monkey, primate, macaque, ape, “model organism”, worm, dog, cat, animal, machine, AI, artificial



Abstract screening step	Screening Aim	Manual check (Y/N)	Indication	Keywords
5	Inclusion	N	Social interactions	“social interaction”, “social group”, intersubjectivity, “sensorimotor communication”, socialization, socialisation, “social behaviour”, “social behavior”, “social relation”, “social group”, “social communication”, social embodiment”, “non-verbal communication”, collaborator, “shared representation”, “group process”, “social process”, “social information” “student-teacher”, “teacher-student”, “student-student”, “group interactions”, “social factor”, co-regulation, face-to-face, “shared attention”, “collaborative learning”, “shared monitoring”, “joint monitoring”, collaboration, “group regulation”, dyad, triad, pair, “interpersonal synchrony”, “group work”, “joint action”, “joint learning”, “joint effort”, CSCL, “computer-supported collaborative learning”, “collaborative process”, “dyadic process”, “group member”, “cognitive interaction”, “emotional interaction”
6	Sorting	Y	Qualitative-only	qualitative, ethnography, “action research”, “social observation”, “focus group”, “case study”, “content analysis”, “discourse analysis”, ethnographic, ethnography, “grounded theory”, narrative, observational, phenomenological, phenomenology, genetic
7	Sorting	Y	Education field	education, educate, student, teacher, teach, learn, collaboration, collaborate, collaborative, academic, class, classroom, “knowledge construction”, “construction of knowledge”, curriculum, pedagogy, pedagogical, self-assessment, self-regulation, study, tutor, lecture

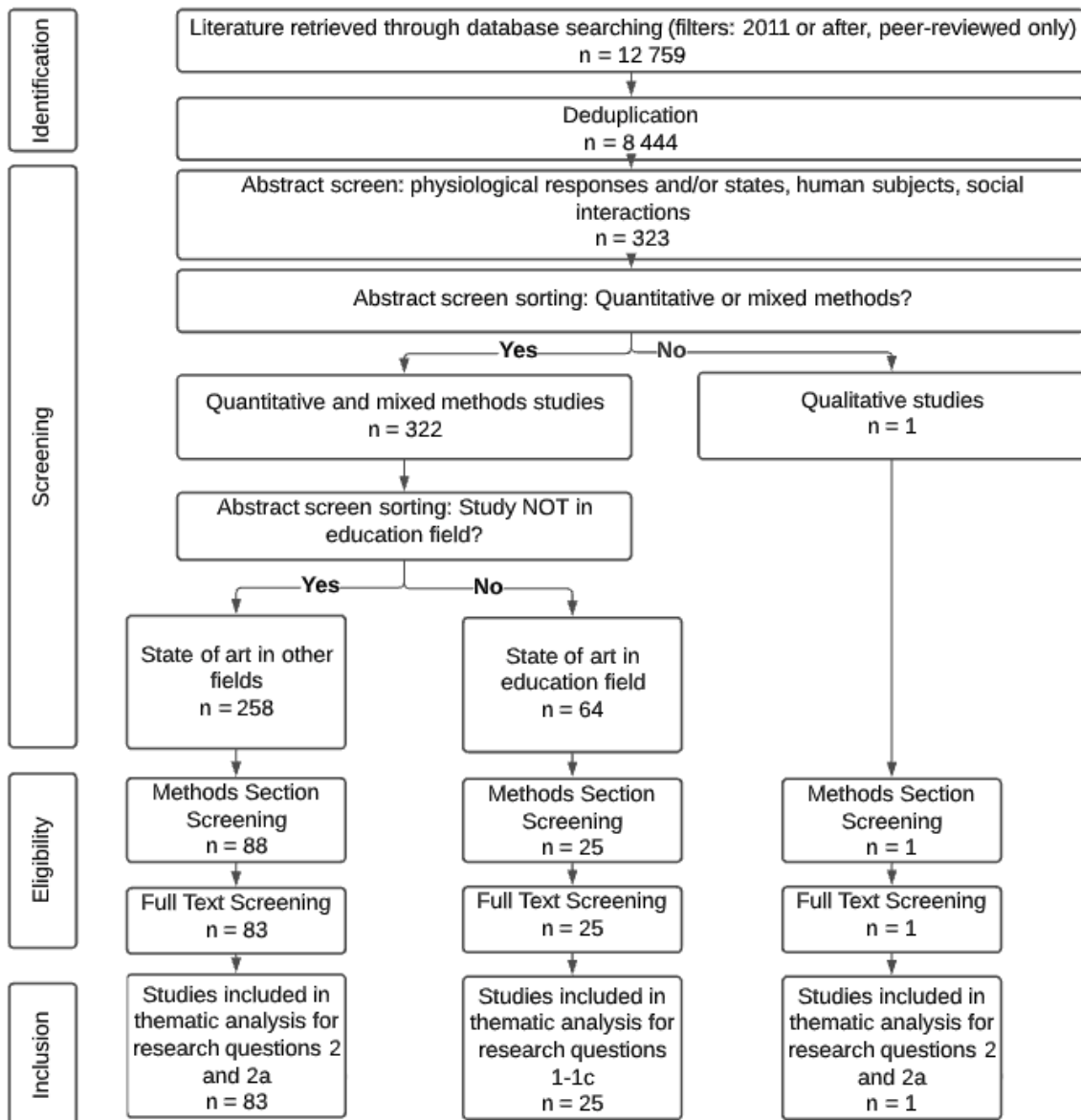
### 3.2.3 Screening process

The literature was screened first using a title and abstract screening tool (Appendix A) elaborated specifically for this study based on the guidelines and example by Polanin and colleagues (2019). Researcher bias was reduced by applying uniform screening criteria to the retrieved literature, and by recording the results, the basis for the findings is transparent.

Guided by the research questions, this SLR designed the screening procedure to reach two distinct scopes of analysis. The first scope addressed the utilization of physiological response data to understand human social interaction in the field of education (research questions 1, 1a, 1b, and 1c). The second scope considered the potential future directions of the education field according to the state of the art in physiological measurement in social interactions literature and the potential pros and cons of incorporating physiological data into education research (research questions 2 and 2a). Thus, the screening process was progressively specific on the initial dataset of literature. The first set of literature included studies that were (1) conducted within the past ten years to capture the current state of the art across fields and (2) peer-reviewed journal articles. Research that met these criteria was (3) empirical research that was experimental, quasi-experimental, qualitative, observational studies, evaluation studies, systematic literature reviews, and meta-analyses, of which systematic literature reviews and meta-analyses were used to cross-validate findings of this study. The Mixed Methods Appraisal Tool (MMAT) was used to categorize objectively studies based on experimental design (see Appendix B) (Hong et al., 2018). Additionally, these studies necessarily investigated human social interactions. Qualitative studies helped understand the current state of the art, pros, and cons of physiological data incorporation into education (research questions 2 and 2a) and were sorted into a separate pool. However, to focus quantitatively on the utilization of physiological response data, the author screened remaining quantitative and mixed methods studies. At this point, the pool of literature was divided into *in* and *out* of the education field to address research questions 1-1c and 2-2a

respectively. Then, the method sections were evaluated to determine whether the studies fulfilled all relevant criteria. Full-text screenings were next conducted to confirm relevance. A diagram of the screening process is displayed in Fig. 1 according to PRISMA guidelines (Liberati et al., 2009) that displays the process categories (left sidebar), and stepwise elaboration of the process (diagram elements) noting study quantities at each step.

**FIGURE 1.** Flowchart for screening and study inclusion



### 3.3 Data analysis and synthesis

#### 3.3.1 Data appraisals

After screening for relevance according to inclusion and exclusion criteria, the MMAT (for original, see Appendix C) was used to categorize and assess the quality of full-text articles. The MMAT was developed and updated by Hong and colleagues (2018) to provide a transparent and unbiased approach to appraise the quality of a wide spectrum of empirical study designs. Because this review included mixed methods, qualitative, and quantitative studies of various types (e.g., randomized controlled, non-randomized, and descriptive), the MMAT served appropriately. The MMAT comprised of five study design types with five criteria each. The present SLR adapted the MMAT with minor revisions to suit the literature appraised (Table 2). Criteria regarding non-empirical studies were discarded since these types of articles had been previously excluded. Across the screened literature, studies variably put forward research questions, hypotheses, and stated aims. Thus, the MMAT criteria were revised to deal with the diversity of experimental approach. Without revision, high-quality articles lacking explicit research questions would have been discarded and weakened the conclusions of this review. The possible responses on each criterion were ‘yes’, ‘no’, and ‘can’t tell’. The screened studies passed the quality assessment, so none were excluded.

**TABLE 3.** Revised Mixed Methods Appraisal Tool (MMAT)

Category of study designs	Methodological quality criteria	Responses		
		Yes	No	Can't tell
1. Qualitative	1.1. Is the qualitative approach appropriate to answer the research question/test the hypothesis/achieve the aim?			
	1.2. Are the qualitative data collection methods adequate to answer the research question/hypothesis/aim?			
	1.3. "Are the findings adequately derived from the data?" (Hong et al., 2018, p.2)			
	1.4. "Is the interpretation of results sufficiently substantiated by data?" (Hong et al., 2018, p.2)			
	1.5. "Is there coherence between qualitative data sources, collection, analysis, and interpretation?" (Hong et al., 2018, p.2)			
2. Quantitative randomized controlled trials	2.1. "Is randomization appropriately performed?" (Hong et al., 2018, p.2)			
	2.2. "Are the groups comparable at baseline?" (Hong et al., 2018, p.2)			

	2.3. "Are there complete outcome data?" (Hong et al., 2018, p.2)			
	2.4. "Are outcome assessors blinded to the intervention provided?" (Hong et al., 2018, p.2)			
	2.5. "Did the participants adhere to the assigned intervention?" (Hong et al., 2018, p.2)			
3. Quantitative non-randomized	3.1. Do the participants represent the target population?			
	3.2. "Are measurements appropriate regarding both the outcome and intervention (or exposure)?" (Hong et al., 2018, p.2)			
	3.3. "Are there complete outcome data?" (Hong et al., 2018, p.2)			
	3.4. "Are the confounders accounted for in the design and analysis?" (Hong et al., 2018, p.2)			
	3.5. "During the study period, is the intervention administered (or exposure occurred) as intended?" "Are there complete outcome data?" (Hong et al., 2018, p.2)			
4. Quantitative descriptive	4.1. Is the sampling strategy relevant to address the research question/hypothesis/aim?			
	4.2. "Is the sample representative of the target population?" (Hong et al., 2018, p.2)			
	4.3. "Are the measurements appropriate?" (Hong et al., 2018, p.2)			
	4.4. "Is the risk of nonresponse bias low?" (Hong et al., 2018, p.2)			
	4.5. Is the statistical analysis appropriate to answer the research question/test the hypothesis/achieve the aim?			
5. Mixed methods	5.1. Is there an adequate rationale for utilizing a mixed methods design to address the research question/hypothesis/aim?			
	5.2. Are the different components of the study effectively integrated to answer the research question/test the hypothesis/achieve the aim?			
	5.3. "Are the outputs of the integration of qualitative and quantitative components adequately interpreted?" (Hong et al., 2018, p.2)			
	5.4. "Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?" (Hong et al., 2018, p.2)			
	5.5. "Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?" (Hong et al., 2018, p.2)			

### 3.3.2 Data extraction and analysis

Data extraction was performed on all 109 studies systematically, which resulted in over 40,000 words of data. Extracted data included Study (Author(s), year), Study design, Participants, Physiological Response Technology (PRT) apparatus, PRT procedure, PRT utility, Quantitative data handling and analysis, Qualitative data handling and analysis, and Conclusions (including limitations and future directions). These extracts were further condensed into the summarized data presented in tables ( which tables?) and appendices (mention which appendices) present in this thesis.

After summarizing the information from each study, the findings were categorized according to the type of biodata used and the study aim. In sequence, information particularly relevant for research questions 1-1c and 2-2a were highlighted. It is important to mention that priority was given to the 26

studies in the field of education. The remaining 83 studies outside of education presented a wide array of aims, integration methodologies, and findings, the most pertinent of which are presented in the results section. However, due to the substantial diversity, depth, and complexity of this literature, further work beyond this thesis is required to provide a comprehensive assessment of the state of the art from diverse fields.

# 4 RESULTS

The present SLR investigated how PRTs have been used in scientific studies addressing social interactions, as well as identified the strengths and weaknesses of this technology and methods to inform future directions in educational studies. This chapter begins with a summary of all results of studies utilizing PRTs in human social interactions in different fields of science, and in the sequence presents the results organized to respond to the research questions. Synthesis of the information to determine pros and cons of integration into education research and its future directions is conducted in the discussion section.

## *4.1 The use of PRTs to study social interactions*

To address research question 1 and its sub-questions and provide context for research questions 2 and 2a, Appendix D summarizes the studies in education within the past decade that use physiological data in human social interactions through describing: the study design type, PRT/biosensors used, bio data collected, utility, physiological measures of interest (PMOIs), other MOIs, participants, task, quantitative data analysis, and if applicable, qualitative data analysis. Because the author identified only one study using a qualitative-only methodology to investigate human social interactions in education, it is included in Appendix D. The remaining studies systematically reviewed were quantitative or mixed methods.

To address research questions 2 and 2a, Appendix E provides summary information of studies outside the field of education and specifically includes biodata, data analysis and conclusions.

**TABLE 4.** Acronyms utilized in the present text

<b>Acronym</b>	<b>Meaning</b>
%DET	Percent determinism
%REC	Percent recurrence
ADL	Average diagonal length
ANCOVA	One-way analysis of covariance
ANOVA	Univariate analysis of variance
AOI	Area of interest
AQ	Autism quotient
ASD	Autism spectrum disorder
BPV	Blood pulse volume
CAMMS	Cognitive, affective, metacognitive, motivational, and/or social
CC	Cross correlation
CFEn	Cross-fuzzy entropy
CL	Collaborative learning
CMC	Computer-mediated communication
CNV	Contingent negative variation
CO	Cardiac output
CRQA	Cross-recurrence quantification analysis
CS	Corrugator supercillii
CSCL	Computer-supported collaborative learning
DA	Directional agreement
DMCA	Detrending moving-average cross-correlation
ECG	Electrocardiogram
EDA	Electrodermal activity
EEG	Electroencephalography
EMG	Electromyography
ERD	Event-related desynchronization
ERN	Event-related negativity
ERP	Event-related potential
FF	Follower-follower
fMRI	Functional magnetic resonance imaging
fNIRS	Functional near-infrared spectroscopy
FRN	Feedback-related negativity
FT	Facial temperature
FTF	Face-to-face
GEE	Generalized estimating equations
GPA	Grade point average
GSR	Galvanic skin response
Hbo	Oxygenated hemoglobin
Hbr	Deoxygenated hemoglobin
HR	Heart rate
HRV	Heart rate variability
HRV-HF	Heart rate variability-high frequency



HRV- RMSSD	Heart rate variability root mean square of the successive differences
HRV-SDNN	Heart rate variability standard deviation of the normal-to-normal
IBI	Inter-beat interval
IBS	Inter-brain synchrony
ICC	Intraclass correlation
IDM	Instantaneous derivative matching
INS	Inter-neural synchrony
JA	Joint attention
JVA	Joint visual attention
LBP-TOP	Local binary patterns-three orthogonal planes
LF	Leader-follower
MdRQA	Multidimensional recurrence quantification analysis
MEG	Magnetoencephalography
MLA	Machine learning algorithms
MMAT	Mixed methods appraisal tool
MOI	Measure of interest
MORE	Mobile multimodal recording system
MWE	Minimum width envelope
Nc	Negative central
NLP	Natural language processing
NSSCR	Nonspecific skin conductance responses
OO	Orbicularis oculi
PBC	Physio-behavioral coupling
PC	Physiological coupling
PCI	Physiological coupling index
PD	Prisoner's dilemma
PEP	Pre-ejection period
PMOI	Physiological measure of interest
PRT	Physiological response technology
PS	Physiological synchrony
PS	Physiological synchrony
RF	Relative frequency
SCL	Skin conductance level
SCR	Skin conductance response
SLR	Systematic literature review
SM	Signal matching
SPC	Social physiological compliance
SSI	Single session index
TPR	Total peripheral resistance
WC	Weighted coherence
WPLI	Weighted phase lag index
WTC	Wavelet transform coherence
ZM	Zygomaticus major

**4.2 Research question one: How have biosensors and other physiological response technologies been utilized to investigate human social interactions in the field of education?**

**4.2.1 Research questions 1a and 1b, respectively: what are the focuses of research? How are different biosensors integrated?**

There were 32 distinct focuses of the 26 studies identified, which are systematically described below. This review defines a ‘focus’ as a specific combination of PRTs or biosensors to gather biodata for a specific utility. Thus, studies that employed more than one biosensor were deemed to have more than one focus. For example, Malmberg and colleagues (2019a) utilized EDA sensors and facial expression recognition software. Montague and colleagues (2014) and Ahonen and colleagues (2018) used EDA sensors and ECG systems. Studies that use the same biosensor for different utilities were also deemed to have different focuses. For example, Malmberg and colleagues (2019a; 2019b) used EDA sensors to analyze both arousal events and synchronous activity. Additionally, the experimental paradigms, especially the tasks, are explained to provide insight into PRT integration. Table 4 provides an overview of major study characteristics’ frequencies in the education field studies included in this systematic review. Note that some feature values may add to >100% due to multiple feature types per study.

**TABLE 5.** Relative frequencies (RF) of major experimental paradigm characteristics in education research

Study feature	Type	Occurrence frequency
Setting	Laboratory	50%
	Classroom	50%
Study design	Quantitative	50%
	Qualitative	4%
	Mixed methods	46%
Group	2-4 group members	85%

	Large group (e.g., class)	15%
Interaction	In-person interaction	65%
	Virtual interaction	27%
	Both in-person and virtual	8%
Biodata	EDA	42%
	EEG	8%
	Eye-tracking	27%
	Facial emotion recognition	4%
	fNIRS	15%
	HR	15%
	Movement (acceleration)	4%
Utility	Synchrony	85%
MOIs	Learning gains	38%
	Task performance	8%
	Collaboration quality	15%
	Metacognitive monitoring	19%
	Prior knowledge	8%
Quantitative data analysis	ANOVA	27%
	Pearson correlation	35%
	Linear model	15%
	Descriptive statistics	15%
	Cross correlation	8%
	WTC	12%
	MdQRA	8%
	t-test	15%
	Matrix analysis	12%
	MWE	4%
Qualitative data analysis	Qualitative content analysis	50%

### **Electrocardiogram (ECG)-based HR and HRV synchrony**

Four studies (and focuses) were identified gathering heart rate (HR) data with an ECG system (Ahonen et al., 2016; 2018; Montague et al., 2014; Sobocinski et al., 2020). Ahonen and colleagues (2016) investigated heart rate variability synchrony among dyads in a turn-taking, pair-programming task. They processed HR data to observe HRV-RMSSD, specifically determining the

synchrony of this signal at the group level as a measure of social physiological compliance (SPC) and investigating its relations with collaborative task type and task load. Ahonen and colleagues (2018) ran a similar experiment with turn-taking rate decided by the dyad, then extracted HRV-SDNN from ECG signals, determined their synchronicity as a measure of SPC, and investigated the relationship between SPC and group-level valence and engagement.

Montague and colleagues (2014) investigated synchrony of dyad members during a Multi-attribute Task Battery (MATB), which involved three simultaneous tasks of monitoring, tracking, and resource management. They indexed IBI synchrony to 'physiological compliance' (PC), which they decomposed into 5 indices (PCIs) of SM, IDM, DA, CC, and WC, and explored their relations to task demand, technology reliability group performance, passive user rating of active user's workload, and shared perception of technology trustworthiness. Sobocinski and colleagues (2020) explored synchrony of physiological state transitions of groups of 3 or 4 during a collaborative physics exam. The researchers computed HR synchrony as aggregated HR state to measure physiological state transitions, and the relations between transitions and monitoring targets, valence, and phase, as well as reaction to monitoring, were analyzed.

### **Electrodermal activity (EDA) synchrony**

There were 13 focuses on EDA sensors gathering EDA data, with 2 focused on individual sympathetic autonomic arousal and 11 focused on synchrony of arousal. The studies that focused on arousal events explored the relations between them and metacognitive events that were qualitatively analyzed to provide other measures of interest. Eleven studies focused on synchrony, though they used diverse terms and concepts: 6 used 'physiological synchrony' (PS), 5 of which were investigated in terms of metacognitive monitoring events from qualitatively analyzed utterances (Dindar et al., 2019; Haataja et al., 2018; 2021; Malmberg et al., 2019a; 2019b) and 1 in terms of task performance, learning gains, and collaboration quality (Schneider et al., 2020). In the latter, dyads coded robots collaboratively with a 2 x 2 intervention design. Interventions were visualization of proportion of utterances and an explanation of the benefits of collaboration.

Dindar and colleagues (2019) examined 12 groups of 3 or 4 students during a two-part collaborative learning session, which included a collaborative essay followed by experimentation and report writing. Haataja and colleagues (2018) carried out investigations with 16 groups of 3 high school students who collaborated to design a healthy breakfast with 5-phased information. The next study aimed to specifically examine the valence of monitoring utterances during triad collaboration on a company-running simulation with 6 periods with intermittent 1-min transitions (Haataja et al., 2021). Malmberg and colleagues (2019a; 2019b) conducted studies with groups of 3 or 4 collaborating on a complex, open-ended task, and a physics exam, respectively. Notably, all studies were conducted in classroom settings.

Another study investigated 'group activation level', the relations of which were explored with group affective state and task- and socially-related factors which were qualitatively analyzed for 3-4 person groups during collaborative designing of a heat-efficient house (Törmänen et al., 2021). Ahonen and colleagues (2018) investigated 'social physiological compliance' in terms of its relations to group valence and engagement through a pair-programming task described above. Pijera-Díaz and colleagues (2019) avoided adherence with a pre-determined concept and instead discussed synchrony as 'within-triad arousal', which was investigated during collaborative physics experiments. The relations between within-triad arousal levels, directional agreement, and contagion during collaboration were investigated. Gillies and colleagues (2016) focused on 'EDA synchrony', an index for synchronous engagement, during a 1-hr lesson when 20 year-6 students collaborated to construct visual, auditory, and kinesthetic representations of learning. Finally, Montague and colleagues (2014), in addition to IBI synchrony, simultaneously examined EDA synchrony as PC through 5 PCIs of SM, IDM, DA, CC, and WC to task demand, technology reliability group performance, passive user rating of active user's workload, and shared perception of technology trustworthiness during dyadic multi-attribute task battery sessions.

### **Electroencephalogram (EEG)-based synchronous electrical brain activity**

In 2 studies (Bevilacqua et al., 2019; Dikker et al., 2017) EEG systems were employed to measure electrical brain activity with an aim to measure synchrony between pairs. Both studies conceptualized of the inter-brain synchrony as total

interdependence and conducted experiments in classrooms. Dikker and colleagues (2017) sought to explore the relationship between 3 instances of TI: group, student-to-group, and student-to-student and other measures of interest: student teaching style preferences, focus, group affinity, empathetic disposition, student-teacher closeness, and student-student closeness across differing FTF conditions. Specifically, EEG signals were recorded for 50 min during alternating classroom conditions of video, lecture, reading, and discussion from students who faced the wall, group, or peer for 2 min before and after class. Bevilacqua and colleagues (2019) collected biodata from 20 min lessons of interleaved lecture and video teaching modes (5 min each, alternating). They investigated the relations between PMOIs: student-group and student-teacher total interdependence and other MOIs: teaching style, student knowledge retention, pre- and post-study teacher closeness and content likeability, and pre- and post-session likability of lesson and experiment.

### **Eye-tracking technologies (glasses, computer-equipped, or cameras) measured eye movements and pupillary dilation**

Six studies employed eye-tracking devices to track eye movements to compute gaze direction and movement. Four of these studies defined their physiological measure of interest (PMOI) as JVA (Schneider & Pea, 2013; 2014; 2015, Schneider et al., 2018), Shvarts and Abrahamson (2019) analyzed data as scanpaths, Molinari (2017) focused on gaze fixations, and Thepsoonthorn and colleagues (2016) mutual gaze convergence.

Of these studies, two leveraged multimodal physiological data (Schneider & Pea, 2013; Thepsoonthorn et al., 2016). During 5 min FTF lectures of teacher-student dyads, Thepsoonthorn and colleagues (2016) utilized eye trackers and accelerometers (placed on the head) to measure eye and head movements to generally track dyadic non-verbal interactional behavior and specifically measure head nodding synchrony and mutual gaze convergence. The relationship between the latter PMOIs and prior knowledge was investigated.

In addition to JVA, Schneider and Pea (2013) also measured pupillary dilation as an indicator of arousal and hence cognitive load. Schneider and Pea explored JVA through an intervention with a tool that made their partner's gaze visible and hence increased JVA compared to the no-visible condition in 3 studies (2013; 2014; 2015). With these interventions, dyads collaboratively explained

diagram-based contrasting cases for 12 min. Using this paradigm, Schneider & Pea (2013) examined the relations between JVA and learning gain, GPA (prior course performance), quality of collaboration, amount of speech production, pupillary dilation, fixations, saccades, and the qualitative measures of student coordination, convention creation, hypothesis building, and theory sharing. Schneider & Pea (2014), using the same data corpus as the previous study, assessed the relations between JVA (in a gaze awareness context) and quality of collaboration, learning gain, fixations, saccades, as potentially modulated by 8 (qualitatively defined) social collaboration quality dimensions: “sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, reciprocal interaction, and individual task orientation.” Schneider & Pea (2015), from the same data corpus, analyzed how the gaze awareness intervention altered collaboration quality through changes in linguistic features of dyadic communication: convergence and coherence (qualitatively determined measures). Schneider and colleagues (2018) studied how dyads’ JVA (no gaze awareness intervention) was related to learning gain, task performance, and quality of collaboration during a collaborative construction task.

Work by Shvarts and Abrahamson (2019) is the only qualitative study included in this SLR. Dyads consisted of a tutor, who guided, and a student peer, who manipulated a triangle on a computer to change its color according to its vertex position. The study focused on patterns across students’ actions, specifically describing the connection between scanpaths from eye-movement data and qualitatively assessed features of dyadic interactions: embodied movements (e.g., hand gestures), utterances, attention coordination, and JVA.

Molinari (2017) tracked eye movements of dyads set up in different rooms to work alone then collaborate via computer on concept maps. To understand the influence of shared and different prior knowledge and co-construction of shared representations, they focused on the linkage between gaze fixation and number of visual elements in concept maps, shared and unshared, and how they were transferred to the collaborative concept map.

### **Facial expression recognition software measured LBP-TOP features**

In addition to EDA measurements of arousal and synchrony, Malmberg and colleagues (2019a) also analyzed emotion valence as groups of 3 or 4 students

collaboratively composed of a midterm plan. To do this, they extracted LBP-TOP (local binary patterns-three orthogonal planes) features from videos with facial expression recognition software to analyze valence expressions, categorized as positive, negative, or neutral are related to work phases and collaboration quality in terms of levels of interaction (high or low).

### **Functional near-infrared spectroscopy (fNIRS): oxygenated hemoglobin (Hbo) measurements to evaluate inter-brain synchrony**

fNIRS was utilized to measure Hbo and deoxygenated hemoglobin (Hb/Hbr) concentrations to determine synchrony of brain activity in various regions in 3 studies (Liu et al., 2019; Pan et al., 2018; Zheng et al., 2018). Liu and colleagues (2019) studied the influence of prior knowledge and teaching mode on brain synchrony in a 2 x 2 fashion with teacher-student dyads. Namely, they investigated potential correlation and mediation of perceived teacher-student interaction, familiarity with teaching materials (prior knowledge), and students' post-test scores (learning outcome) on teacher-student left prefrontal cortex and right temporal-parietal junction synchrony in identified channels of interest. Pan and colleagues (2018) explored the role of differing learning modes of music (part- and whole-learning) and learning performance on interpersonal brain synchronization (IBS) dynamics. In teacher-student dyads, teachers taught students 2 similar songs phrase by phrase or all at once. Particularly, their aim was to investigate how learning modes affect IBS, how IBS during vocal or non-vocal interaction affects learning performance, and how IBS associated with observation or imitation phases during the learning predicts learning performance. IBS in the bilateral inferior frontal cortex was the focus, as a center involved in interactive communication, singing, and game playing, and is an important mirror neuron system hub (Pan et al., 2018). Zheng and colleagues (2018) aimed to understand how inter-neural synchrony (INS), synonymous with IBS and brain-to-brain synchrony, varies by teaching style and numerical reasoning pre- and post-test scores. Teacher-student dyads engaged in teaching and learning via FTF lecture, a FTF interactive mode, or a video lecture over 13-26 minutes. Interested in whether teachers can predict the students' knowledge state (the prediction-transmission hypothesis), student brain activity was 10s time lagged as a new physiological measure of interest (PMOI). They also identified frequency bands of interest (0.5-0.7 Hz and 0.06-0.07 Hz) and the left anterior



superior temporal cortex as a representational hub in the semantic system and right temporal-parietal junction for its exclusive involvement in high-level mentalizing as regions of interest.

In research from Holper and colleagues (2013), teachers led students (in dyads) through the Meno dialog adaptively until students come to the correct conclusion. They studied Hbo and Hb levels and synchrony to indicate prefrontal cortex activity and examined how it related to phase of dialog (task), percent agreement with the original Meno dialog, and transfer ability. As well, the relations between teacher and student Hbo levels (synchrony, though not referred to as such) were analyzed. The prefrontal cortex was an area of focus due to its ability to mark mental effort.

### **Biodata and social interactions: dependencies and indicators**

Half of the studies took place in laboratory and half in classroom settings, and they focused mostly on the synchrony of bio-physiological signals (85%) in groups with two to four participants (85%). In-person interaction (only) was the most common across studies (65%), followed by virtual interaction (only) at 27% and both combined at 8%. The most often used biodata streams came from electrodermal activity (EDA) readings (65%), followed by eye-tracking (27%), fNIRS (15%), HR (15%), EEG (8%), facial emotion recognition (4%) and bodily movement (4%). In terms of measures of interest, learning gains was the most common (38%), followed by metacognitive monitoring (19%), collaboration quality (15%), task performance (8%), and prior knowledge (8%).

Several studies investigated physiological synchrony through physiological coupling indices (PCIs) due to various associations between distinct indices and outcomes of interest such as learning gains, task performance, and social interactions. As well, the nuances of physiological synchrony were further explored: group-dependence, temporal variation (e.g., EDA synchrony cycling and momentary brain synchronization), and role-dependence in the group. A significant number of studies used biodata to indicate cognitive regulations. EDA arousal, EDA synchrony and aggregate heart rate were investigated as potential indicators of whether groups were facing cognitive challenges or were “off-track”.

#### 4.2.2 Research question 1c: how are physiological data handled and analyzed?

##### **Electrocardiogram (ECG)**

ECG apparatuses captured signal that was processed and computed into HR, HRV-SDNN, HRV-RMSSD, and IBI. In all studies, group-level synchrony was the physiological measure of interest (PMOI). Sobocinski and colleagues (2020) utilized descriptive statistics of group-level frequency and distribution of HR synchrony, an index to physiological state transitions, and related them to qualitative measures of metacognitive monitoring. Ahonen and colleagues (2016; 2018) utilized two varying computations of heart rate variability (RMSSD and SDNN respectively) synchrony to measure social physiological compliance (SPC). Both studies computed average cross-pair correlations to estimate SPC, with significance determined by bootstrapping. To study the association between SPC and self-report items, the latter article fit a linear regression model (Ahonen et al., 2018). Montague and colleagues (2014) looked at IBI synchrony as an estimation of physiological compliance (PC) through five indices: signal matching (SM)<sup>2</sup>, instantaneous derivative matching (IDM)<sup>3</sup>, directional agreement (DA)<sup>4</sup>, cross correlation (CC)<sup>5</sup>, and weighted coherence (WC)<sup>6</sup>. PC indicators (PCIs) were compared to each other with correlation coefficients. PCIs were compared with baseline with a linear mixed effects model. The relations between PCIs and other MOIs were tested with linear mixed effects models.

##### **Electrodermal activity (EDA)**

In all research on EDA, the major analyses of interest dealt with synchrony of sympathetic autonomic arousal measured through electrodermal activity. Malmberg and colleagues (2019a; 2019b) both analyzed EDA peak (nonspecific skin conductance responses; NSSCRs) synchrony between students as

---

<sup>2</sup> SM compares the differences between the areas under two subjects' biodata curves (Montague et al., 2014)

<sup>3</sup> IDM compares the slopes of the curves through averaging instantaneous derivatives of corresponding points (Montague et al., 2014)

<sup>4</sup> DA is calculated in two steps: first, each point on the curve is compared to the previous point to compute directional movement; then, a percentage of the two curves' DA is computed (Montague et al., 2014)

<sup>5</sup> A CC coefficient determines the covariance of corresponding data points along two data curves (Montague et al., 2014)

<sup>6</sup> WC quantifies the band-specific similarity of physiological responses while disregarding phase differences (Montague et al., 2014)

physiological synchrony, but the former performed a descriptive statistical analysis on 1-min segments, while the latter computed single session indices (SSIs) through Pearson correlations on 15-s windows of data, determined significance through Monte Carlo shuffling, and temporal variation was measured by comparison to the mean in 120-s windows in 1-s steps. Haataja and colleagues (2018) calculated physiological concordance as an index of PS. It was computed the same as Malmberg and colleagues (2019b), and a detrending moving-average cross-correlation (DMCA) coefficient was utilized to probe relations between group monitoring and physiological concordance. A later study by Haataja and colleagues (2021) quantified the synchrony between two group member signals through multidimensional recurrence quantification analysis (MdRQA); percent determinism (%DET)<sup>7</sup>, percent recurrence (%REC)<sup>8</sup>, and average diagonal length (ADL)<sup>9</sup> were compared and generalized estimating equations (GEEs) modeled arousal and synchrony as predictors for valence, equality of participation, and task performance. Dindar and colleagues (2019) employed MdRQA as well for describing synchrony of triad EDA signals in 1-min windows. Every minute was visualized in recurrence plots. Correlations were calculated between monitoring durations and MdRQA indices. T-tests were conducted to reveal potential differences in PS between CL conditions with and without monitoring.

Pijeira-Díaz and colleagues (2019) used 1-min windows with 250-ms moving steps to rate individual EDA into low, medium, or high arousal based on peak frequency. DA was computed for 1-s windows, level agreement on 1-min windows. Descriptive statistics were employed to compare high-arousal intervals between subjects, and latency was analyzed to explore a possible contagion effect. Törmänen and colleagues (2021) similarly categorized arousal levels across 30-s segments, but medium and high were “activating” and low was “deactivating”; chi-square tests of independence explored affective state and physiological activation relations and observed relations between valence of interactions and task- and socially-related factors. Gillies and colleagues (2016)

---

<sup>7</sup> %DET quantifies the repetition of these components in terms of the greater patterns of synchrony (Wallot et al., 2016)

<sup>8</sup> %REC quantifies the sharedness of signals' individual components (Wallot et al., 2016)

<sup>9</sup> ADL measures the mean magnitude of the repeated synchrony patterns (Wallot et al., 2016)

computed EDA synchrony using a between-student correlation matrix using group-level time-bin averaged EDA amplitude values. They computed Pearson correlations for every possible pair across conditions to create connectivity networks for each condition, network analysis computed facets of connectivity between students.

Montague and colleagues (2014) applied the same data analysis to EDA data as HR-IBI data. Schneider and colleagues (2020) ran correlations to test the relations between physiological data and MOIs, as well as the PCIs: Pearson correlation, DA, SM, and IDM (in 2-s moving windows). Pearson correlation was identified as the best suited index for PS, aggregated in 30-s windows, and computed high/low activity via comparison to the mean and ran correlations between them and MOIs.

### **Electroencephalogram (EEG)**

EEG total interdependence was the measure of signal synchrony and was calculated the same in both Dikker and colleagues (2017) and Bevilacqua and colleagues (2019). Total interdependence was calculated as spectral coherence with the Welch method. Dikker and colleagues' study conducted repeated-measures univariate analyses of variance (ANOVAs) to detect differences of brain synchrony across teaching styles. Then, multilevel models were created to find the relationship between student-to-group total interdependence and questionnaire metrics. Bevilacqua and colleagues (2019) computed multilevel models with days nested within students were used to address RQs, namely repeated-measures multilevel regression analyses: independent variable x factor on dependent variable (e.g., teaching style X quiz scores on student-group total interdependence).

### **Eye-tracking**

Thepsoonthorn and colleagues (2016) detected mutual gaze convergence in 1-s windows, average percent mutual gaze convergence was computed, and t-tests observed differences between knowledge conditions. Molinari (2017) ran correlations between eye movement measures (i.e., fixations and shifting on concept mapping elements) and learning performance, and transitions between AOIs in differing knowledge conditions were statistically described.

Schneider and Pea (2013) ran ANOVAs to test the effects of joint attention on learning gains, quality of collaboration, and cognitive load (estimated by pupil

dilation). ANOVAs observed effects of gaze condition/intervention on JVA, individual fixations & saccades on learning outcomes, and gaze condition on amount speech production. A Pearson correlation was run between speech production and JVA. Model for potential mediators of student learning were tested: collaboration, JVA, cognitive load, with GPA as the covariate.

Schneider and Pea (2014) constructed novel network graphs for each subject where fixations are nodes and saccades are edges; dyad-level network graphs where nodes are screen areas and edges are saccades. ANOVAs computed network metrics (node quantity, node size, edge quantity, reciprocated edge quantity) based on gaze condition. Pearson correlations were run on betweenness centrality, JVA, and matrix metrics and qualitative sub-dimensions of collaboration quality and learning outcomes.

Schneider and Pea (2015) computed unigram, bigram, and trigram probabilities with Pearson correlations between n-grams and measures of interest. Convergence (mimicking grammatical structure of interlocutor) for numerous grammatical features compared between visible gaze and no-gaze conditions (ANOVAs). Coherence (word repetition between dyad members) was calculated via cosine similarity over a 5-exchange window. The predictor strength of coherence for collaboration quality and learning outcome was calculated by cosine similarity matrix. Similarity over time was computed with correlations of 5 vs. 5 utterances using 1-exchange sliding window.

Schneider and colleagues (2018) ran correlations between joint attention and learning gains and performance during tasks, also divided by year of student (1st, 2nd, or 3rd year). Correlations were run between quality of collaboration, its constitutive components, and percentage of joint attention. On 2 dyads that both showed high JVA (both high task performance but differing learning gains): augmented (with spatial and speech duration information) cross-recurrence graphs visualized the synchronization of JVA. A measure of imbalance of 'visual leadership' was computed and plotted against learning gains per task type.

### **Facial expression emotion**

A study utilized an automatic facial expression recognition algorithm estimated valence expressions; descriptive statistics were applied to the expression frequency per minute in simultaneous arousal episodes, and a one-

way ANOVA observed differences in valence over interaction types (high-level, low-level, and confusing) (Malmberg et al., 2019a).

### **Functional near-infrared spectroscopy (fNIRS)**

Three studies calculated interpersonal brain/neural synchrony (IBS/INS) with WTC analysis of Hbo signal (Pan et al., 2018; Zheng et al., 2018; Liu et al., 2019). Pan and colleagues (2018) compared conditions to baseline with paired sample t-tests across frequency bands, and coupling directionality was computed with granger causality analysis. Zheng and colleagues (2018) averaged INS across all channels per dyad to calculate INS increase per teaching mode. One-way ANCOVAs for all channel combinations for all frequencies were used to compare INS across teaching modes and over time for mode and outcome. Analyses were conducted between INS increases with 2-14-s time lags before and after (step = 2 s). Liu and colleagues (2019) performed one-sample t-tests for all channels' task-related INS, generated a t-map, and a mixed 2 x 2 ANOVA was run on channels with significant INS for each condition.

### **Body movements**

Nodding synchrony (in 1.8-s windows) was calculated as the time lag between teacher and student's head acceleration with Spearman's Rank Correlation (Thepsoonthorn et al., 2016). T-tests determined differences of average percentage of head nodding synchrony per condition.

### **Qualitative analysis of eye movements**

Shvarts & Abrahamson (2019) conducted a micro-ethnography to search for patterns across student actions (e.g., hand gestures), dyad utterances, and the gaze parameters of students and tutors. They analyzed coordination dynamics of intradyadic attention and between student action and tutor attention. Joint attention was coded through non-verbal and verbal contexts (i.e., to establish gaze overlap was not by chance). The analysis and findings are discussed in depth in the discussion section.

### **ANOVA and Pearson correlation predominate, and novel analyses emerge**

Across biodata types, researchers often use ANOVAs to determine differences between means, and Pearson correlations to measure the linear correlation between two data sets (see Table 4 for relative frequencies and other less frequent analyses). To analyze biodata in social contexts, researchers employed three data analysis techniques novel to the education field. MdRQA

gauged the coordination of multiple data streams over time, matrix analysis assessed the impact of factors on connectivity measures between subjects, and minimum width envelope (MWE) allowed comparison of difference curves over time under varying conditions.

**4.3 Research questions 2 and 2a: What are benefits and drawbacks of incorporating physiological data into education research? What are the future directions for research according to the current state of the literature?**

First, all the studies are sorted by biodata and specific utility (e.g., to assess synchrony of signals) to list the associated variables of interest and the relative frequency (RF) of biodata type (Table 5). Next, the author presents the most relevant and significant data analysis methods and conclusions of all included studies outside of education, grouped by biodata type. Table 6 displays the RF of major data analysis methods. Condensed yet detailed summaries for all included literature outside the education field are presented in Appendix E.

**TABLE 6.** Biodata usage outside the education field

<b>Biodata</b>	<b>RF</b>	<b>Specific Utility</b>	<b>Associated variables</b>
EDA	12.04%	Sympathetic autonomic arousal	Affiliation (Stevanovic et al., 2019)  Dominance (Stevanovic et al., 2019) Social anxiety (Shalom et al., 2015) Guilt-induced emotional arousal (Yu et al., 2017) Joint decision-making (proposal, non-acceptance, acceptance) (Stevanovic et al., 2021) Physical presence (Hietanen et al., 2020) Sympathetic-adrenal-medullary (SAM)-associated trust (Potts et al., 2019) Gaze perception (Prinsen et al., 2019)
		Synchrony	Competition vs. Cooperation (Chanel et al., 2012) Emotional team dynamics (Mønster et al., 2016) Cooperative success (Behrens et al., 2020)

HRV-HF	3.61%	Arousal	Social and emotional regulation (Sariñana-González et al., 2019) Regulatory effort (Vanderhasselt et al., 2018) Cooperation (Sariñana-González et al., 2019)
		Synchrony	Emotional team dynamics (Mønster et al., 2016)
HRV-RMSSD	1.20%	Arousal	Acute stress, traumatic stress, gendered psychosocial stress (Birze et al., 2020)
HR	3.61%	Arousal	Social anxiety (Shalom et al., 2015) Cooperation (Sariñana-González et al., 2019) Competition (Sariñana-González et al., 2019)
		Synchrony	Interpersonal trust (Mitkidis et al., 2015)
OO/ZM and CS activity	3.61%	Valence expressions	Positive & negative emotions (Hietanen et al., 2020)  Physical presence (Hietanen et al., 2020)
		Valence synchrony	Emotional team dynamics (Mønster et al., 2016)  Competition vs. Cooperation (Chanel et al., 2012)
Eye movements	28.92%	Gaze direction & movement	Turn-taking (Freeth et al., 2013; Ho et al., 2015)  Cognitive Load (De Lillo et al., 2021; Ho et al., 2015) Visual attention (Fu et al., 2019; Nelson & Mondloch, 2019) Visual bias for theory of mind (Ferguson & Breheny, 2011) Correlational gaze behavior (Hessels et al., 2017) Joint visual attention (Anaya et al., 2021; Caruana et al., 2020)  Shared attention (Pfeiffer et al., 2012) Gesture response (Kajopoulos et al., 2021) Anxiety (Azriel et al., 2020; Hessels et al., 2018) Face scanning (Haensel et al., 2020) Social attention (Barzy et al., 2020; De Lillo et al., 2021; Freeth et al., 2013; Hanley et al., 2015)  Response preference (Kendrick & Holler, 2017) Affiliative responses (Vranjes et al., 2019) Sociality of choice (Cañigüeral & Hamilton, 2019) Emotional expression perception (Nelson & Mondloch, 2019)  Cognitive processing (Vanderhasselt et al., 2018)



			<p>Emotion processing (Vanderhasselt et al., 2018)</p> <p>Eye contact as a primer for accessing an emerging semantic knowledge system (Hoehl et al., 2014)</p> <p>Guilt-driven gaze aversion (Yu et al., 2017)</p> <p>Social Preference conceptualization (Rahal et al., 2020)</p> <p>Creation and manipulation of mental representations (Lenoble &amp; El Haj, 2021)</p> <p>Future thinking (Lenoble &amp; El Haj, 2021)</p>
		Pupillary dilation	<p>Eye-contact perception (Honma et al., 2012)</p> <p>Autonomic activity (Honma et al., 2012)</p>
HbO & Hb concentration	1.20%	Individual brain activity	(Left prefrontal cortex) cooperation (Balconi et al., 2017)
Magnetic fields from electrical brain activity	2.41%	Synchrony	(Alpha and gamma) rapid turn-taking speech (Ahn et al., 2017)
Electrical brain activity	31.12%	Synchrony	<p>Peer gaze synchrony (JVA) (Anaya et al., 2021)</p> <p>Leader-follower dynamics (Zamm et al., 2021a)</p> <p>(Alpha band) infant-mother emotional state (Santamaria et al., 2020)</p> <p>Approach-avoidance motivation (Anaya et al., 2021)</p> <p>Emotion regulation (Anaya et al., 2021)</p> <p>(P300) Active goal-directed processing of stimulus (Apanovich et al., 2018)</p> <p>Alpha band desynchronization: joint attention and/or semantic processing (Hoehl et al., 2014)</p> <p>Emotional processing and arousal (theta, delta, alpha, beta, gamma bands) (Kraus et al., 2020)</p> <p>Observation of affective, social, and informative gestures (Balconi &amp; Fronza, 2020)</p> <p>Level of cooperation (Hu et al., 2018)</p> <p>Shared representations (Ruissen &amp; de Bruijn, 2015)</p>

	<p>Mirroring and empathetic responses (Balconi et al., 2020)</p> <p>Interpersonal motor coordination (Varlet et al., 2020)</p> <p>(Cortical beta band's event-related desynchronization) Voluntary action preparation and execution (Zamm et al., 2021b)</p> <p>(Gamma band) shared intentionality (Barraza et al., 2020)</p> <p>(Theta band) joint motor action (Barraza et al., 2020)</p> <p>(Alpha band) Coupled brain-to-speech entrainment (Pérez et al., 2019)</p> <p>(Alpha band) rapid turn-taking speech (Ahn et al., 2017)</p>
Individual event related potential (ERP)	<p>(P300) Active goal-directed processing of stimulus (Apanovich et al., 2018)</p> <p>(ERN) Social feedback processing (Fishman &amp; Ng, 2013)</p> <p>(ERN) Error monitoring in shared representations (de Brujin et al., 2011)</p> <p>(P300 and FRN) Response to other's aggression (Chen et al., 2017)</p> <p>(P300) Social distance (Chen et al., 2017)</p> <p>(Contingent negative variation [CNV]) Cognitive load: processing of own statements (Wagner-Altendorf et al., 2020)</p> <p>(FRN) Social feedback processing (Czeszumski et al., 2019)</p> <p>(Negative central [Nc]) attentional control (Soto-Icaza et al., 2019)</p> <p>(N400 in centro-parietal regions) semantic processing (Hernández-Gutiérrez et al., 2018)</p>
Individual brain activity	<p>(Left prefrontal cortex) cooperation (Balconi et al., 2017)</p> <p>(Infant alpha band) joint attention (St. John et al., 2016)</p> <p>(Temporoparietal) mentalization (Soto-Icaza et al., 2019)</p>

			(Late posterior positivity) motivated social attention (Hernández-Gutiérrez et al., 2018)
Haemodynamic response	2.41%	Blood-oxygen-level dependent (BOLD) response	Social-communicative stimuli processing, i.e., joint attention (Oberwelland et al., 2016)  (Temporal and medial prefrontal regions) mentalizing Rauchbauer et al., 2019  (Hypothalamus and amygdala) social motivation and reward (Rauchbauer et al., 2019)
Pre-ejection period (PEP)	1.20%		Sympathetic activation (Peters et al., 2014)
Cardiac Output (CO)	1.20%		Approach-motivated challenge (Peters et al., 2014)
Total Peripheral Resistance (TPR)	1.20%		Avoidance-mediated threat (Peters et al., 2014)
Respiration rate	1.20%	Synchrony	Competition vs. Cooperation (Chanel et al., 2012)
IBI	2.41%	Synchrony	Physiological behavioral coupling (Strang et al., 2013) Competition vs. Cooperation (Chanel et al., 2012)
Embodied movements	3.61%	Postural sway synchrony	Physiological behavioral coupling (Strang et al., 2013)
		Finger movement synchrony	Neural self-other integration (Varlet et al., 2020)
		Hand movement synchrony	Interpersonal synchrony during joint task (Wallot et al., 2016)
fNIRS	3.61%	Synchrony	Leader-follower and follower-follower IBS dynamics (Jiang et al., 2015) Creativity (Xue et al., 2018) Cooperation (Xue et al., 2018, Sun et al., 2021)

---

Facial temperature	1.20%	Arousal	Social attention (gaze direction) (Ioannou et al., 2014)
--------------------	-------	---------	--

Proximity (Ioannou et al., 2014)

### **Electrodermal activity (EDA)**

Outside of education, 13 of 83 articles investigated EDA. A significant portion (5 of 13) of these studies investigated a form of EDA synchrony, which was termed “psychophysiological synchrony” (Stevanovic et al., 2021), “physiological linkage” (Järvelä et al., 2014), “SCL synchrony” (Vanutelli et al., 2017), “synchrony of phasic skin conductance” (Mønster et al., 2016), or “physiological compliance” (Chanel et al., 2012).

Synchrony within and between dyads was assessed with a variety of statistical analyses, with two-tailed t-tests being the most common (30% of EDA synchrony analyses), followed by repeated measures ANOVAs (20%) and Pearson correlations (20%). The following tests were run: two-tailed t-test and Pearson correlation (Stevanovic et al., 2021), two-tailed t-test and LMM (Järvelä et al., 2014), repeated measure ANOVA (Vanutelli et al., 2017), CRQA and false pair surrogate analysis (Mønster et al., 2016), and two-tailed t-test, repeated measures ANOVA and Pearson correlation (Chanel et al., 2012).

Stevanovic and colleagues (2021) found elevated EDA and dyadic synchrony during proposals and emergence of joint decision. Järvelä and colleagues (2014) showed social interaction and competitive/cooperative context impacted synchrony in a multiplayer game. Vanutelli and colleagues (2017) found elevated EDA synchrony after social feedback. Mønster and colleagues (2016) higher synchrony was observed during negative group affect. Chanel and colleagues (2012) observed higher physiological compliance during competitive gameplay.

For EDA arousal, 12 instances of data analysis were carried out in 8 studies, with One-way ANOVAs (21.43%) and paired t-tests (21.43%) being the most frequent, followed by mixed model ANOVAs (14.29%), LMMs (14.29%), and ANCOVAs (14.29%). Other analyses utilized were a repeated measures ANOVA (7.14%) and a general estimating equation (GEE) (7.14%).

Various components of EDA signal (SCL, SCR, NSSCR) were analyzed in diverse paradigms to provide insight into socially-related factors such as stress,

anxiety, guilt, or affiliation. Specifically, studies provided evidence for skin conductance level's (SCL) elevation in response to acute stress (Potts et al., 2019), skin conductance response (SCR) as affective arousal to reciprocal gaze (Hietanen et al., 2020; Prinsen et al., 2019), higher NSSCR frequency as emotional arousal and apprehension (Kraus et al., 2020), elevated SCL as emotional arousal of guilt (Yu et al., 2017), increased SCL as a physiological arousal cost of social anxiety (Rösler et al., 2021), higher skin conductance in relation to subjective measures of anxiety and an objective measure of HR (Shalom et al., 2015), and decreased EDA (coupled with HR, HRV, and EMG signal analyses) was positively associated with intersubjective affiliation (Stevanovic et al., 2019).

### **Electroencephalogram (EEG)**

One of the most popular physiological data to collect in social interactions outside education within the past decade has been electrical oscillatory activity of the brain captured with EEG systems. Of the 83 total studies outside education, 25 studies (31.12%) measured brain activity with EEG. Brain activity synchrony was investigated in 14 articles, individual event related potential (ERP) in 8 studies, and other individual brain activity in 4 studies (Hernández-Gutiérrez et al., 2018 fit both latter categories).

A battery of statistical tests was utilized in studies of inter-brain synchrony with relatively minor overlap. The most popular analysis was the one-way repeated measures ANOVA (14.71%), followed by the two-way repeated measures ANOVA (11.76%), paired t-test (8.82%) and one-way ANOVA (8.82%). Less frequently, Pearson correlation, multiple linear regression, mixed repeated measures ANOVAs, two-way ANOVAs, and mixed model ANOVAs were used in 5.88% of the studies. The following analyses were only present in one study (2.94%): the multilevel model, linear mixed model, Spearman's correlation, Wilcoxon test, Pearson product moment correlation, linear trend analysis, general linear model, nonparametric bootstrap-based t-test, and weighted phase lag index.

Inter-brain synchrony (IBS, inter-neural synchrony, INS) has been explored in a wide variety of human social interaction contexts and is a measure of a diverse set of cognitive and emotional processes. For example, during structured and unstructured interactions, matching (both positive/negative) intra-dyadic

Delta-Beta coupling predicted peer gaze synchrony (Anaya et al., 2021). Zamm and colleagues (2021a) examined interneural synchrony during duet performances where pause duration was improvised and found shorter pauses were associated with higher synchrony and higher beta event related desynchronization (ERD, an indicator of sensorimotor prediction processes). In solo followed by duet performances, Zamm and colleagues (2021b) found interpersonal neural synchrony in the form of period coupling consistent with duet frequencies but are unsure of the contribution of movement artifacts. See Appendix E for more findings of EEG-based inter-brain synchrony.

Individual ERP measurements were analyzed by mixed repeated measures ANOVAs the most often (28.57%), then Wilcoxon tests (14.29%), with the remaining analyses only present in one study each (7.14%): two-way repeated measures ANOVA, one-way ANOVA, mixed model ANOVA, paired t-test, repeated measures GLM, LMM, Threshold-free cluster enhancement, and Spearman's correlation.

Apanovich and colleagues (2018) looked at speeded decision making for a discriminatory task in a cooperative and competitive context with holistic and analytic subjects and found that P300, which reflects this decision making, is positively associated with holistic subjects in cooperation and analytic subjects in competition. During a social game task (Prisoner's Dilemma, PD), Chen and colleagues (2017) found feedback-related negativity (FRN) was not sensitive to social distance (stranger vs. friend), but the P300 component of the ERP was during the late-stage evaluation of choice outcomes. During a joint four-alternative forced choice task with monetary rewards however, FRN differed depending on the outcome (rewards) or social situation (Czeszumski et al., 2019). See Appendix E for all findings related to ERPs.

Individual brain activity apart from ERPs was most frequently examined by mixed repeated measures ANOVAs (37.50%), and other analyses were used once (12.50%): ANCOVA, Pearson correlation, Wilcoxon test, two-way repeated measures ANOVA, and mixed model ANOVA.

In a timed recall task with cooperative and competitive modes, high performance in the cooperative condition for high behavioral activation system (BAS) subjects, left prefrontal cortex (PFC) activity was elevated (Balconi et al., 2017). During a naturalistic co-watching experience, Soto-Icaza and colleagues

(2019) found a potential neural mechanism that contributes to initiation of JA to explicit mentalization of increased beta band activity in the temporoparietal junction. See Appendix E for more findings on EEG-based individual brain activity during social interactions.

### **Magnetoencephalography (MEG)**

MEG is a noninvasive test akin to EEG except that it measures the electrical currents of the brain indirectly through measurement of the magnetic fields they produce. One study (1.20%) identified employed MEG. During a (turn-taking) speaking and listening task, alpha and gamma signals were synchronized between dyad members only during verbal turn-taking (Ahn et al., 2017). The researchers used weighted phase lag index (WPLI) of true vs false dyads to capture inter-brain synchronization.

### **Eye-tracking**

Eye-tracking was one of the most popular (23 studies, 27.71%) methods of using biodata to analyze human social interactions outside of the education field. For the utility of gaze data, 11 of 23 studies overlapped with at least one other study. The most common was to measure social attention (Barzy et al., 2020; De Lillo et al., 2021; Feeth et al., 2013; Hanley et al., 2015), followed by cognitive load (De Lillo et al., 2021; Ho et al., 2015), visual attention (Fu et al., 2019; Nelson & Mondloch, 2019), joint visual attention (Anaya et al., 2021; Caruana et al., 2020), and anxiety (Azriel et al., 2020; Hessels et al., 2018). Other measures of interest were only present in a single study (see Appendix E, Eye movements).

Analysis of gaze data was most frequently performed with mixed repeated measures ANOVAs (16.67%), followed by mixed model ANOVAs (11.43%) and Pearson correlations (11.43%). The remaining analyses were used twice (5.71%): repeated measures ANOVAs, one-way ANOVAs, two-way repeated measures ANOVAs, and linear mixed models, or once (2.86%): non-linear cross correlation (CC), linear regression model, intraclass correlation (ICC), multilevel model, chi-square likelihood ratio, independent sample t-test, one-tailed Pearson product-moment correlation, mixed effects logistic regression model, mixed effects repeated measures linear regression, Mann-Whitney U test, Cohen's d for effect size, and (qualitative) conversational analysis.

For the most popular area of investigation, a pattern of research focus is observed. For instance, investigations on social attention utilizing eye-tracking

data explored differences between individuals and dyads frequently in the context of autism spectrum disorder (ASD). In all, studies focused on differences among children vs. adults (De Lillo et al., 2021), children diagnosed with autism spectrum disorder (ASD) (Barzy et al., 2020; Hanley et al., 2015), and high vs low Autism quotient (AQ) score subjects (Freeth et al., 2013). If other research in this area overlapped in a general sense, such as gaze data used to assess cognitive load, the studies utilized vastly differing subjects and paradigms, and no commonalities could be identified (see De Lillo et al., 2021; Ho et al., 2015). Of the 23 eye-tracking studies, 12 associated gaze data with diverse and distinct variables or measures.

In terms of findings, investigations on social attention are briefly summarized; the remaining studies' conclusions are presented in Appendix E. In a naturalistic conversation task with ASD diagnosed participants, those with ASD spent more time gazing on the background and less time on the experimenter's face and talking about oneself increased gaze to the experimenter's face (Barzy et al., 2020). During a semi-structured social interaction, ASD participants gazed more at the mouth and less at the eye region, which resulted in missing non-verbal signs (Hanley et al., 2015). Freeth and colleagues (2013) found that AQ scores did not impact gaze behavior on face or background in real-life or video-based interactions. When comparing gaze of individuals of different developmental stages (adolescent, young adult, and older adult) during naturalistic FTF conversations, adolescents and older adults exhibited reduced social attention compared to young adults (De Lillo et al., 2021). Thus, the studies evidence that age and ASD can impact gaze behavior in social interactions, but ASD-associated traits may not.

### **Facial temperature**

Facial temperature (FT) served as a measure of physiological arousal; Ioannou and colleagues (2014) examined FT over six regions of interest (ROIs) to include most of the face over social proximity and gaze conditions and found strong positive associations between closeness and direct gaze with FT. The researchers used mixed repeated measures ANOVAs, one-way ANOVAs, and Pearson correlations to characterize the relation of different facial regions' temperature to each other and experimental conditions.

### **Embodied movements**



Three studies (3.61%) investigated a form of embodied movement synchrony during social interactions, including postural sway, finger movement, and hand movement. Each data analysis occurred only in one study (16.67% frequency per analysis): CC, CRQA, cross-fuzzy entropy (the results of these were compared in Strang et al., 2013), repeated measures ANOVA, multiple linear regression, and MdRQA. During the cooperative mode of a Tetris-like game, Strang and colleagues (2013) found elevated postural sway coupling synchrony (as an index to physio-behavioral coupling; PBC). When participants jointly improvised finger movements, EEG responses related to processing self and other movements increased in magnitude, especially for the leader (Varlet et al., 2020). Two-way repeated measures ANOVAs observed differences in self and other EEG measures over conditions, and one-way ANOVAs were used to test for differences between dyads. In a Lego car-building task, Wallot and colleagues (2016) used MdRQA to calculate hand movement synchrony, specifically interested in %Determinism, and determined significance through false pair analysis. Car size and aesthetic appeal were negatively correlated with hand movement synchrony in the hierarchical condition, but all objective and subjective measures were positively associated with hand movement synchrony in the free condition.

### **Functional magnetic resonance imaging (fMRI)**

In a social encounter with a third object of attention, Oberwelland and colleagues (2016) found that brain regions associated with emotions and motivation/reward processes were recruited in self-initiated JA, especially with a familiar partner. Specific areas of the brain showed lower activation in adolescents compared to children, evidencing a developmental effect. A mixed ANOVA model evaluated effects of conditions, group, random effects, and inter-subject factors on fMRI data. In a study by Rauchbauer and colleagues (2019), robot heads with retroprojected faces and humans were interactive partners. Mixed model ANOVAs assessed the effects of the interacting agent across relevant factors and observed activity in ROIs for mentalizing and social motivation showed marked elevation in human-human compared to human-robot interactions.

### **Functional near-infrared spectroscopy (fNIRS)**

Jiang et al. (2015) used fNIRS hyperscanning on a triad during a leaderless conversation to observe IBS patterns assessed by WTC across leader-follower

(LF) and follower-follower (FF) pairs and found that IBS was significantly higher in leader-initiated interactions. Xue and colleagues (2018) hyperscanned dyads during creativity tasks with various combinations of highly- and lowly-creative subjects to determine how IBS relates to performance and creativity, and after WTC analysis, t-map generation including all channels, and ANOVAs, low-low dyads had significant IBS and all dyads performed similarly. Sun and colleagues (2021) measured the effects of team-member social experience on IBS during a joint drawing task and found through WTC, a two-way ANOVA, and Pearson correlations that IBS was negatively correlated with task performance.

#### **Pre-ejection period (PEP), cardiac output (CO) and total peripheral resistance (TPR)**

Peters et al (2014) used ECG, ICG, and blood pressure measurements of PEP, CO, and TPR to investigate sympathetic arousal before and during dyadic interactions where one member purposely suppressed or expressed emotions. Mixed ANOVAs assessed differences of physiological variables compared to baseline during anticipation of conversation and in the conversation. Peters and colleagues found that suppression of affective signals, regardless of valence, caused threat-related physiological responses for both partners.

#### **Heart rate (HR)**

Shalom and colleagues (2015) investigated the relations of self-reported arousal, social anxiety, and control and success to HR during FTF and CMC (text) communication. HR, along with SC, discussed above, was a measure of anxiety levels; utilizing one-way within-subjects ANOVAs to observe differences between baseline and conditions and a 2 x 2 mixed model ANOVA to observe differences across anxiety levels and conditions, researchers found that although self-report measures indicated reduced anxiety in the CMC condition compared to FTF, HR measures did not differ between FTF and CMC. Another study employed competitive and cooperative tasks to investigate the task-related differences in HR through use of repeated-measures ANOVA with a general linear model within groups between periods (Sariñana-González et al., 2019). They found that participants that cooperated had the highest HR compared to the non-social task or competitive task, women had higher HRs than men, and men were more sensitive to their performance outcome in cooperation tasks.

#### **Heart rate variability (HRV)**

Sariñana-González and colleagues, during the same experiment, also investigated HRV in the same fashion, and through the same analyses as used for HR, found that participants that cooperated had lower HRV. A study by Vanderhasselt and colleagues (2018), measured HRV as an index of regulatory effort as participants received FTF social feedback of their photograph. With use of mixed repeated measures ANOVA with time and gender as factors, they found gender-specific results for all measures: HRV was higher for women, indicating emotion regulation or motivation differences, men had larger HRV increases across the board, the greatest during negative feedback. Birze et al. (2020) measured HRV during police communicators' daily work to determine the effect of persistent workplace stressors and posttraumatic stress symptoms. A multiple linear regression was calculated to predict HRV during acute stress and chronic subjective stressors. They found that gendered psychological stressors showed physiological stress responses- an increase in HRV.

Mønster et al. (2016) investigated HRV synchrony to unveil emotional team dynamics during an assembly-line task and used cross recurrence quantification analysis (CRQA) to estimate synchrony, and compared to false-pairs, synchrony was associated with negative group affect.

### **Inter beat interval (IBI)**

Strang and colleagues (2013) examined IBI (as well as postural sway, described above) synchrony as an index to physio-behavioral coupling (PBC) to examine relations between it and team performance and perceived team trust and cohesion across time with various time-series measures of CC, CRQA, and CFEn. They found that PBC was driven by team-task environment, and that PBC was negatively correlated to team performance and attributes. Chanel et al. (2012) during competitive or cooperative gameplay, used IBI as an index to physiological compliance (PC) to detect the effects of social presence. Repeated measures ANCOVAs observed differences between game mode and social presence condition, and correlations were run between PC and questionnaire items. They found that PC increased with self-reported social interaction involvement and higher PC during competitive mode. Linear mixed models were applied and found that IBI coherence was predicted by the negative feeling subscale (questionnaire).

### **Pupillary dilation**

Eye-contact perception (Honma et al. (2012) had participants sit 80 cm apart looking at each other and measured pupillary dilation during eye-tracking-confirmed eye contact. Welch's t-tests compared pupillary dilation between conditions and Pearson correlations checked for within-gender associations and ANCOVA was run to test for gender differences in pupillary dilation. They found that the perceivers were unable to correctly gauge gaze direction of partners, and pupils were significantly larger during perceived pupil fixation from the viewer, with females showing a higher gaze direction accuracy than men.

### **Electromyography (EMG)**

Hietanen et al. (2020), using live, video call, and video conditions, tested the effects of mutual eye contact on facial muscle activity associated with emotional processes. Within-subject ANOVAs compared EMG responses over conditions and gaze directions, and EMG was compared over time. The authors found that being seen, not physical presence, was required for autonomic arousal of the person being viewed. Mønster and colleagues (2016), by comparing EDA synchrony to EMG data that indicated emotional valence (zygomaticus major-smile and corrugator supercillii- frown) with Pearson correlations, found that EDA synchrony was associated with negative group affect. During a competitive or cooperative game mode at home or in the laboratory, Chanel et al. (2012) measured EMG data from the OO and CS facial muscles to determine emotional contagion and found that social climate could be inferred from EMG synchrony in this case. The PC indexed by EMG data showed an inverse relationship to PC indexed by IBI data, meaning in the cooperative game mode, participants exhibited greater EMG synchrony.

# 5 DISCUSSION

This chapter presents the interpretation and discussion of the results guided by the research questions. Special attention is given to the current benefits of using physiological data in educational studies, and the possible future directions in this field. This chapter also includes discussions about the limitations, shortcomings and the biases of data handling and analysis techniques that can hinder experimental designs based on biosensing methodologies. At the end of the chapter, recommendations for the future of the education field as well as considerations related to the literature outside education are provided. The aim of this discussion is to bring attention to current challenges and offer recommendations based on careful analysis that may spark improvements in methodologies employed in social learning sciences.

## *5.1 Current strengths and future directions of physiological data in education*

The discussion of the advantages of the current use of physiological data in education is based on the 26 studies identified, and specifically focuses on the context of collaborative learning. Analysis of bio-physiological signals during collaboration, an important twenty-first century skill, unveil underlying attentional, cognitive, and emotional processes that leads to deeper understanding.

Through the exploratory experimentation employed by the studies, findings have supported the notion that physiological response data and other biodata (termed biodata for simplicity) can act as indicators for at least one component of many important measures of interest related to collaborative learning. **The analysis of the findings revealed the plausible utility of biodata for research in education involving social interactions, particularly in learning analytics research.** Additionally, intervention in collaborative learning with

biodata-driven tools has been fruitful. As the collaborative activities unfold, participant biodata has shown statistically significant relations with metacognitive regulations, adaptive regulations, and socioemotional states, and predictive of task performance, learning gains, and components of collaboration quality. These findings are collated under the research-based viewpoint that biodata has a high potential for use in learning analytics, especially in group learning settings.

#### *Measuring prior knowledge*

For example, Both Molinari (2017) and Thepsoonthorn et al. (2016) have shown that intra-dyadic embodied movements (JVA and head nodding synchrony, respectively) indicate prior knowledge state. Prior knowledge state is worthy of consideration in CL contexts because learners are typically tasked with co-constructing an external representation that reflects shared understanding from different or similar starting points of knowledge. During the process of co-construction, they must externalize their internal representations, negotiate meaning, and re-internalize within the milieu of multiple external representations (Boshuizen & Tabachneck-Schijf, 1998). It is conceivable that biodata streams from eye- and head-tracking devices and software could be used in combination with other measures of prior knowledge (e.g., quizzes, surveys) to arrange students properly for group work. However, much work lies ahead to determine the impact of group prior knowledge compositions on CL measures of interest such as CL quality, task performance, and learning gains.

#### *Indicating cognitive challenge and difficulty*

Several studies have analyzed biodata in reference to meaningful measures of collaborative learning as interpreted through the theory of self-regulated learning (SRL). SRL understands that challenges create the opportunity to characterize and measure the strategic adaptations learners make (Hadwin et al., 2017). The use of biodata can unveil previously invisible processes. Indeed, metacognitive monitoring events were found to be reflected in EDA arousal (Malmberg et al., 2019a; 2019b), and groups that faced difficulties achieved higher levels of EDA synchrony (Malmberg et al., 2019a), and those that synchronized displayed more negative facial expressions (Malmberg et al., 2019b). Negatively valenced utterances coincide with synchronous EDA activity and may indicate cognitive challenges or that “things

are off-track” (Haataja et al., 2021). The takeaways are that EDA, facial expression recognition technology can be combined with other measures, such as linguistic analysis, to track the progress and challenge during CL. This can be complemented by measures of individual cognitive load by measuring pupil dilation (Schneider & Pea, 2013), and measurement of perceived workload of collaborators (part of common ground) through EDA synchrony analysis (Montague et al., 2014).

Meaningful CL involves challenging situations and activities. Sobocinski et al. (2020) showed that adaptive regulations that are crucial to collaborative success such as on-track, adaptive, and maladaptive sequences can be recognized, in part, by computing aggregated heart rate state of a CL group. Also, asynchronous EDA is an indicator that the collaboration is on-track (Haataja et al., 2021). These biodata streams show the potential to aid understanding the characteristics of the CL process via the SRL model, the challenges faced by the group, and how they deal with them. Important to note is that numerous findings evidence a statistically significant but weak association between biodata variables and measures of interest, so a combination of biodata streams may be advantageous. Dindar and colleagues (2019) found that the relationship between physiological synchrony and metacognitive monitoring are group- and task-dependent. Thus, more sophisticated theoretical models are needed to understand the processes involved in CL.

#### *Measuring individual and group level emotional components of CL*

The individual and group level emotional components of CL also influence and shape the CL process. They can be captured during CL with EDA measurements and facial expression recognition software. EDA measurements can provide group-level activation level information, which has shown associations with group facial emotion valence (Törmanen et al., 2021), and can contribute to understanding affective experiences as a component of CL. Though not yet investigated in terms of CL quality or performance metrics, arousal within triads has been shown to be contagious on the pair-level, and co-arousal varies temporally (Pijera-Diaz et al., 2019). Given these findings, further investigation is warranted into the changes of symmetrical (similarities in direction and level of arousal) and asymmetrical (the contagious nature of

arousal) characteristics of CL over time, the valence of verbal and non-verbal interactions, and how they relate.

#### *Joint attention as an indicator of learning and a tool*

In addition to using biodata to better understand the CL process through the theoretical lens of SRL, how biodata streams relate to metacognitive events over time, the meaning of synchrony and asynchrony, or the impact of socioemotional factors, a number of studies have focused on the use of JVA as a predictor of collaboration quality, task performance, and learning gains. A few of these studies also utilized a mutual gaze awareness tool that displays the partner's gaze on a screen (in a computer supported collaborative learning [CSCL] context). Their findings have significant implications to the use of embodied movement data and many potential applications.

The gaze awareness intervention studies found that the gaze cursor serves as a deictic gesture, which significantly reduces the need for verbal interactions and increases JVA, ultimately elevating shared cognition (Schneider & Pea, 2013). This tool shows the most promise for use in dyads in CSCL or virtual reality CL contexts, where gaze data can be displayed to partners.

Many of the studies discussed in this section have employed a mixed methods design because qualitative analysis contextualizes the quantitative biodata. To analyze monitoring and adaptive regulation, utterances are manually coded to determine the function or valence (Ahonen et al., 2018; Dindar et al., 2019; Malmberg et al., 2019a; 2019b). To determine emotional valence during synchronous EDA, facial expressions are coded (Törmanen et al., 2021). This data enriches interpretation, gives meaning to the quantitative data, and greatly benefits research as a result. However, the process is typically laborious and lengthy. Machine learning algorithms (MLA) trained on JVA network analysis data under shared gaze and no gaze conditions was able to reach accuracy of classification (above or below median split) between 85-100% on each of the 8 components of collaboration quality, a 100% accuracy on the total quality after 13 minutes, and a ~80% prediction accuracy after 10 minutes (Schneider & Pea, 2014). On the implementation level, MLAs can be improved over time, but even the results from the article suggest this tool and MLA could be applied to classroom settings. Within 10-13 minutes, teachers



could have an accurate assessment of the collaboration quality of a large number of dyads, and intervene with scaffolding if necessary (Pea, 2004). In addition to visual data, concurrent utterances are also open to automated quantitative analysis. In another of Schneider and Pea's work (2015), natural language processing (NLP) coupled with JVA measures during the gaze aware and no-gaze conditions showed that language coherence co-occurring with JVA meant that both members of the dyads were in a "spatially-locked" discussion, and marked a learning moment during CL. Notably, whether the dyad was on or off-topic was of large importance in determining learning moments, as off-topic conversations were also found to have high JVA and coherence co-occurrence. The researchers also showed that moments of high JVA did not necessarily mark these learning moments and are less associated with learning gains than when JVA and coherence co-occurred.

Another application of the MLA, when fed with n-grams (words and phrases), cosine similarity scores, coherence, and convergence data, attained a 75% classification accuracy of student learning gains. The training set was 36 students, and the validation set was 4 students, meaning it classified 3 out of 4 correctly; thus, the results should be taken with a healthy dose of skepticism. However, in principle, the MLA showed the potential and should be further explored with larger datasets. If MLA could combine visual, language, and other data streams, it may be able to account for a wider variety of CL situations and more sophisticated models of CL.

#### *Synchrony is context- and time-dependent*

Note that the automated quantitative approaches are not a replacement for qualitative methods; rather, automated quantitative methods are informed by or checked against qualitative methods to operationalize or validate them, respectively. As the findings of JVA and EDA synchrony have suggested, physiological synchronization does not equate to strong collaboration or superior learning gains. Though there are context and group-dependent situations where synchronization indicates these desired outcomes, the reality of CL remains complex. Schneider and colleagues (2018) found that a group with high JVA can have poor learning gains when a partner took a passive or

*free rider*<sup>10</sup> role, and a group with high JVA and learning gains challenged each other often. Schneider et al. (2020) illustrated that it wasn't synchrony that was important but the frequency of cycling in and out of it. Specifically, they found that EDA synchrony across the whole session was not different between high and low performing groups, but the high performing groups experienced a higher frequency of syncing and desyncing. This cycle matched up with the iterative cycling of individual work and group interactions that lead to a shared understanding of concepts. Further, the number of high/low synchronization cycles for each physiological coupling index was moderately correlated with many measures of collaboration quality.

In line with unravelling complexity, Ahonen et al. (2018) found that a positive feedback event came immediately after a depression of EDA, but a negative feedback event came after an EDA spike, showing anticipatory relief and reactive frustration, respectively. Perhaps more interestingly, they uncovered a role-dependent EDA dynamic, where drivers (code writers) were aroused before the feedback event (test code) and navigators (assisted the writers) were more aroused post-feedback, indicating that the latter may not have maintained a mental model of the goodness of code. In terms of a learning analytics application, the differences in task performance, collaboration quality, and learning gains might be investigated in relation to this role-dependent arousal dynamic. If patterns are identified and can be replicated, it is plausible that intervention with the learners' roles or tasks may improve various facets of CL in classrooms.

#### *Measuring cognitive and relational components of CL through neurophysiology*

Lastly, this section turns its attention to the findings from neurophysiological investigations in interactive learning contexts. All of these studies investigated teacher-student relations during a dyadic interactive learning session. fNIRS data from the prefrontal cortex showed that lower activity was correlated to ability to transfer knowledge, which indicates efficiency (Holper et al., 2013). Thus, adding this data stream in the wild could aid in identifying students that are learning less efficiently. Multiple studies found

---

<sup>10</sup> A free rider effect could explain passivity of a student whose partner took an active, leading role (Salomon & Globerson, 1989) in Schneider and colleagues' (2018) study, the design which allowed for performance of most able member to determine the performance of the group.

teacher prediction of student knowledge state through momentary synchronization of brain signals (IBS/INS), though across different tasks and different brain regions and frequency bands (Holper et al., 2013; Zheng et al., 2018).

Liu and colleagues (2019) found prefrontal teacher-student INS was associated with higher student performance in the prior-knowledge condition; INS occurred to a greater degree during face-to-face interactions compared to computer mediated communication. Likewise, Pan and colleagues (2018) found that higher IBS and learning performance was associated with an interactive teacher-student learning session. In practice, fNIRS hyperscanning (measuring from more than 1 person at a time) could be utilized in the classroom to measure the effect of teaching styles on learning performance at the neurophysiological level, beyond the simple paradigms seen to date. As well, fNIRS data streams of the prefrontal cortex may allow prediction of the receptiveness of a learner to novel information by estimating their prior knowledge state relative to the teaching content and allow for teacher adjustment.

EEG hyperscanning techniques have also revealed that student-teacher closeness is closely tied to total interdependence (brain oscillatory activity synchrony); also, brain-to-brain synchrony increases with shared attention or engagement with a stimulus, and even those who experienced a FTF interaction showed more synchrony (Dikker et al., 2017). Mutual closeness ratings for pairs of students that engaged in 2 min eye contact were positively associated with IBS. In light of these findings, social presence and eye contact act as mediators of synchronous brain activity, which could be a marker of closeness and shared intentionality. However, decomposition of constructs such as closeness or empathy into more essential psychological processes would facilitate connecting hypotheses to neurophysiological metrics. However, the replicability of these findings was called into question when a follow-up study was conducted by Bevilacqua et al. (2019) in a similar classroom setting. Neither student-student nor student-teacher IBS was associated with memory retention, but student-teacher closeness ratings were. Thus, a physiological measure that can properly index a predictive construct (or basic elements of it) such as this should be investigated. As well, since the theoretical grounds upon

which these EEG studies rest is IBS arises via joint attention, eye-trackers should be used to determine the relations between biodata streams and whether they can be triangulated to better measure or predict learning gains.

Taken together, the scope of research in education that uses physiological response data in CL contexts has moved towards a more sophisticated understanding of the trajectories, dynamics, task-dependence, group-dependence, and role-dependence revealed by the biodata. Future work is needed to construct sophisticated models that can account for the nuances of the results. Additionally, triangulation of multiple data streams, including multimodal biodata, should be leveraged to contextualize data and tease apart measures. Automated approaches such as machine learning algorithms show promise, and should be validated on larger sample sizes, a diverse set of CL contexts, and with diverse data streams. Importantly, the interdisciplinary challenges remain large and require an equally large interdisciplinary collaborative effort to produce high-quality research to employ cutting edge technologies into well-designed, theoretically grounded studies.

### 5.1.1 Qualitative approach to embodied movement data

One study in the field of education employed a qualitative-only approach to biodata analysis. Shvarts and Abrahamson (2019) simultaneously analyze the eye-tracking, video, and audio data streams with a micro-ethnographic approach. They aimed to identify patterns from these data with the theoretical view that joint sensory-motor behaviors may precede joint actions (Goodwin, 2017) or participatory sense-making (De Jaegher & Di Paolo, 2007). Use of the eye-tracking data allows insight into the otherwise imperceptible interaction of tutor gaze following the student gaze. The patterns of tutor monitoring of student actions provide key information for analyzing the materialization of intersubjective coupling.

In pairs, tutors supported students with verbal interaction in a guiding fashion and deliberately avoided direct instruction. The study identified four phases in the activity. The first phase is convergence, where the tutor's gaze follows and anticipates the student's actions, fitting in with the theory that the subjects are intersubjectively coupled perception-action systems. In the second

phase, tutor's gaze turns from following the real into showing the ideal, but is not decoupled, for the tutor's gaze returns to the student's gaze repetitively. In the second phase, the divergence of tutor from student gaze occurs from attentional anchors, or objects or areas that facilitate action coordination (Abrahamson & Sánchez-García, 2016). In the third phase, tutor guidance on motor coordination of the manipulated object facilitates students' development of novel attentional anchors that are separate from tutors' ideal ones. In the fourth phase, students' fluent motor performance of the task indicates the student is in the micro-zone of proximal development where the tutor can offer pivotal guidance. By asking key questions that reframed student attention at this moment, the student is guided into understanding the mathematical rule. Interestingly, the student's attentional anchors converge with the tutor's attention anchors after verbalization and apparent realization of this rule.

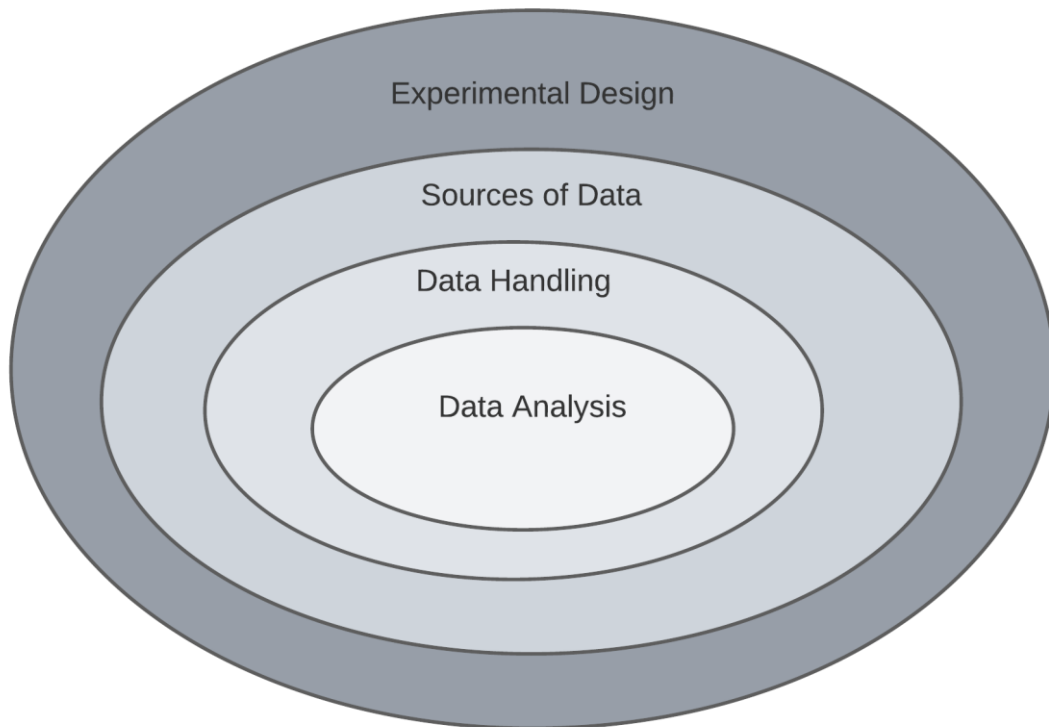
Visualizing embodied movement data streams allows analysis of micro-processes during CL. In addition to eye-tracking, qualitative analysis of posture, and body part movement (head, limb, finger, etc.) with deep theoretical application could provide new insights into CL dynamics and processes. This multichannel data is open for application and hence further development of various theoretical standpoints in education research. Ultimately, these theoretical innovations could broaden the horizons of educational practice.

## *5.2 Methodological challenges in biodata research*

Part of the limitations and biases brought about by the experimental design arise from the sources of data. Likewise, some of the limitations and biases of the data sources are from defining data meanings. Lastly, a portion of the limitations and biases of defining data meanings stems from data analysis methods. In addition, the larger categories impact their nested category: overall experimental design impacts the data collection sources, which impacts the way experimenters handle the data, which influences how they analyze it. The model's elements (see Figure 2) are instantiated in the following sections of the discussion. The model is a product of analysis, in which distinct categories were identified through systematic and iterative coding and referencing (non-exhaustively for literature outside of education). As such, the model is intended

as a preliminary aid to guide the thinking around the tacit features of paradigms in this thesis's domains of interest.

Figure 2 illustrates the identified limitations and biases as a nested model. Each category represents choices made by the experimenters that are nested sequentially. Meaning, the general, parent source of bias is the experimental design.



**FIGURE 2.** A nested model of issues related to the sources of limitations and biases in studying human social interactions with biodata in education

The discussion employs the terms *objective data* and *subjective data*. In the context of this SLR, objective data is gathered from biosensors (hardware or software), and subjective data is generated directly through experimenter interpretation of subjects' verbal or non-verbal actions. This section largely deals with how the use of 'objective' biodata is inherently biased due to the concepts, theories, and methods that define, situate, and utilize them. Thus, the term should be taken with due consideration.

A substantial subset of the studies in education identified employed an experimental design that can be analyzed with a 2 x 2 confusion matrix (see

Figure 3) (Ahonen et al., 2018; Dindar et al., 2019; Haataja et al., 2018; 2021; Malmberg et al., 2019a; 2019b; Pijeira-Díaz et al., 2019; Sobocinski et al., 2020), as put forth by Winne (2019). Rows indicate the presence or absence of a physiological datum, and the columns are presence or absence of a CAMMS event (cognitive, affective, metacognitive, motivational, and/or social events). The researchers of these studies hypothesized that changes in CAMMS events can be represented by changes in physiological data. Based on research, this is plausible, but researchers are at an early stage of unravelling the complex relations. A few studies interpret a CAMMS event to take place when a physiological marker is present (true positive) (Malmberg et al., 2019a; 2019b; Pijeira-Díaz et al., 2019). For example, Malmberg and colleagues (2019a; 2019b) and Pijeira-Díaz and colleagues (2019) identified EDA spikes and qualitatively analyzed the corresponding audio and video recording to determine the presence or absence of CAMMS events.

		CAMMS Event	
		present	absent
Physiological Datum	present	true positive	false positive
	absent	false negative	true negative

**FIGURE 3.** Confusion matrix depicting the relations of a physiological (or bio) datum with a CAMMS event

These experimental approaches, while able to identify true positives, do not account for the other three pairings in the confusion matrix. During these studies, there are many cases where the physiological marker is present, but a CAMMS event is absent, indicating a false positive result. For instance, does group-level EDA synchrony spike at the beginning of a collaborative exam due to stress? In these studies, false positive situations are mentioned but not examined; for example, in Malmberg and colleagues' study (2019), 37 out of 131 EDA joint arousal episodes contained a CAMMS event. Further quantitative and qualitative

analysis on these 94 episodes is feasible. In the same study, researchers used facial expression recognition and identified 126 instances of emotion valence out of 131 episodes; thus, this data in combination with exploratory qualitative analysis has the potential to shed light on other factors (e.g., see Schneider & Pea, 2015 for analysis of linguistic features). Other studies observed EDA synchrony around task phases (Ahonen et al., 2018; Haataja et al., 2021). If EDA spikes during code testing occurred, was it due to the event per se, or unrelated interaction between group members such as eye contact or discourse?

Some designs preclude investigation of false positives, namely, when a physiological marker is absent and the CAMMS event is present (e.g., Malmberg et al., 2019a; 2019b; Pijeira-Díaz et al., 2019). These studies also fall short of addressing false negatives, in which a physiological marker is absent, but the CAMMS event is present. One way to examine false negatives is to have a confederate make scripted verbalizations at particular times that according to theory trigger CAMMS events in group members (Winne, 2019).

In contrast to the designs of Malmberg et al. and Pijeira-Díaz et al., whole-session correlative approaches, such as in Haataja and colleagues' (2018) work, combine time series group-level EDA synchrony with CAMMS events. This allowed visualization and quantification of how often they coincide. Since they were weakly correlated, the ability to discriminate between true and false positives from physiological indicators is low. Yet, this approach again affords the opportunity to further analyze moments of high EDA synchrony. As well, the differences across dimensions of the CAMMS events can be assessed in relation to EDA synchrony levels. The discrepancies between the data streams can be mapped. These approaches would shed light on false negatives.

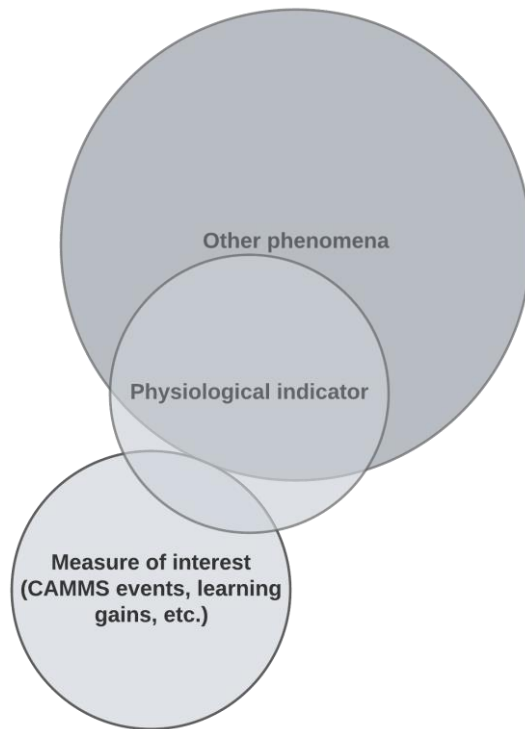
Dindar and colleagues (2019) computed MdrQA indices over 1-min windows with vs. without monitoring. They investigated whether monitoring durations correlated with MdrQA indices. This approach partially addresses some issues of the confusion matrix; namely, they are examining the relations between physiological synchrony (PS) and CAMMS events by comparing PS with and without CAMMS, which allows them to shed light on the predictive power of PS data to mark CAMMS events. If CAMMS events are closely tied to PS, false positive rates would be low. Because they compared MdrQA indices (PS) across monitoring and non-monitoring episodes, they're able to detect whether true



positives would be more frequent than false positives if you used the MdRQA indices to define a monitoring event.

Sobocinski and colleagues (2020) ran correlations between physiological state transitions (from aggregated HR) and sequence types (on-track, adaptive, and maladaptive), which could provide insight into the predictive power of physiological state transitions to these features of collaboration. However, the correlation found between physiological state transitions and adaptive sequences was weak, conveying a weak ability to distinguish between true and false positives from physiological indicators. Schneider et al., 2020 looked at cycles of synchrony through correlations of high and low synchrony periods with other measures of interest. These correlations enabled them to assess the how synchrony is related to learning gains, and features of collaboration quality. There were few significant findings, all with weak associations (e.g., learning gains,  $r(30) = 0.35$ ,  $p < 0.05$ ).

In sum, for studies that employed correlations on whole-session data, all showed significant but weak associations of physiological markers with CAMMS events or other measures. This reveals they are weak or partial predictors for each other (see Figure 4). It is crucial to use exploratory analyses because research has indicated these complex phenomena require complex explanations, and as such, physiological data may not provide a simple, reliable way to indicate CAMMS events, learning gains, etc.



**FIGURE 4.** Physiological data indicates measures of interest *and* other phenomena

The last consideration through the confusion matrix is whether studies can identify true negatives, that is, if a physiological marker is absent, is a CAMMS event also absent? Confirming the absence of a physiological indicator depends on the source of data and its handling, which are discussed in later sections. To ensure the absence of a CAMMS event, researchers must make explicit the answers to: what are the features of baselines that researchers utilize to signal a CAMMS event? What additional data support the absence of a CAMMS event?

Five other studies, instead of hypothesizing variance in CAMMS events, hypothesized variance of learning outcomes and task performance can be represented in variance in physiological data, JVA specifically (Schneider & Pea, 2013; 2014; 2015; Schneider et al., 2018; 2020). Instead of centering investigations of CAMMS events around physiological events, or exploring their relations across time, these study designs sought to observe correlations between conditions and outcomes and differences of outcomes based on JVA conditions. Another confusion matrix with learning outcomes, collaboration

quality, or task performance as the columns can be applied. Since these studies seek to detect potential physiological predictors of the measures of interest, some employed an in-depth analysis on a small number of groups to elucidate potential factors. Concretely, Schneider and colleagues (2018) analyzed two groups with high JVA, high task performance but substantially different high and low learning gains; gestures and proposal reactions were found to vary significantly between the groups. Thus, deeper analysis on false positive situations could facilitate identification of measurable factors that can be integrated to reduce false positives in the future.

In-depth analysis of select groups that follow and do not follow a trend serves only as an exploratory method, and thus results are not generalizable. Though studies such as Schneider and colleagues' (2018) sometimes acknowledge this, follow-up studies have yet to be conducted to validate preliminary findings.

### *5.3 Data analysis: the state of the art, bias, and methods for bias reduction*

#### 5.3.1 The state of the art and future potential

Qualitative analysis was performed in half of the studies in the education field, for all but one in a mixed methods design. The qualitative analyses focused on a range of features of interactions in CL. For an overview of the purposes of qualitative analysis in the education field studies, see Appendix D. This section focuses on the use of quantitative data due to the benefits brought about from recent innovations and its potential.

To deal with biodata streams, quantitative analyses were used in all but one study in the education field. Indeed, most use-cases of physiological response technologies leverage the richness of the data to extract meaningful measures, compare them, and draw conclusions. The most frequently used quantitative analyses are Pearson correlations and various forms of an analysis of variance (ANOVA). Pearson correlations are used to determine the statistical association between two continuous variables in terms of magnitude and direction. ANOVAs are used to determine the differences between means of

unrelated and related groups, across conditions, and across time and conditions with data that meet the assumptions of normal distribution and homogeneity of variance (and that measures are independent of each other when not repeated measures). While these bread-and-butter analyses and others often used in education are valuable, novel statistical techniques to the field offer ability to uncover new information about CL from biodata. The most promising data analyses used were matrix analysis, MdRQA, and MWE.

Matrix analysis can assess the impact of factors on connectivity measures. As Haataja et al. (2018) hypothesized in their discussion, there are possibly key variables that affect the coupling of monitoring and synchrony, such as empathy, the quality of monitoring, or shared understanding for example. Though the design and analysis they performed were not able to test in this manner, a matrix analysis such as that employed by Gillies et al. (2016) or Schneider and Pea (2014; 2015) has the means to detect the relation of factors on connectivity measures and is especially useful for groups of more than two. Taking for example the network analysis employed by Gillies and colleagues (2016), the approach allowed for quantification of *degree*, *participation index*, *assortativity*, and *clustering coefficient*<sup>11</sup>. These computations allow for exploration of how students are connected to each other with a variety of focuses, which can also be analyzed in terms of other factors, such as questionnaire results or experimental conditions.

The connections (assessed by biodata) between learners are dynamic, varying over time (Ahonen et al., 2018; Dindar et al., 2019; Haataja et al., 2021). A detrending moving-average cross-correlation (DMCA) has been used to examine monitoring and group averaged PS (Haataja et al., 2018), however, to move beyond the correlations of mean activity to an event, an MdRQA can provide a wealth of information to characterize the coordination pattern of several variables over time. Haataja and colleagues (2021) performed the MdRQA for real and false groups and compared the measures: percent

---

<sup>11</sup> In the network where students are nodes and edges (connections) are correlations between student biodata (EDA), *degree* is a measure of the number of connections a student makes with others, *participation index* measures the within-group to outside-group connections, *assortativity* measures the degree of connection a student has with like-connected students, and *clustering coefficient* quantifies the degree of mutual connectedness between clusters of connections (clusters are  $n > 2$ ) (Gillies et al., 2016).

recurrence (%REC), percent determinism (%DET), average diagonal line length (ADL). Whereas PCIs serve as a toolbox for evaluation of synchrony of *two* EDA signals, the utility of MdrQA lies in its ability to indicate synchrony of *three* time series data streams. Wallot et al. (2016) explains that %REC quantifies the sharedness of signals' individual components, %DET quantifies the repetition of these components in terms of the greater patterns of synchrony, and ADL measures the mean magnitude of the repeated synchrony patterns. Thus, the strength of MdrQA lies in its ability to observe relations a) between three signals, b) over time, and c) of a complex dynamical system, where factors on different levels interact with each other. Hilpert and Marchand (2018) have identified a current gap between educational theory and current methodologies employed to explore them. Thus, that data analysis techniques like MdrQA can be applied in complex systems research to adequately address these sophisticated theories.

An alternate method for focusing on the temporality of synchronous activity with high granularity is the use of the minimum width envelope (MWE). MWE can compare conditions on difference curves over time with 95% confidence without needing to meet the assumptions for bootstrapping (i.e., MWE obtained through permutation testing showed the same results as bootstrapping) (Ahonen et al., 2018). Importantly, MWE allows investigation with use of distribution information rather than central statistics, and Ahonen and colleagues (2018) used this to investigate the EDA difference curves over time with high temporal granularity, which revealed a time-lag in response between roles. These methods may be particularly valuable for investigating the relations between physiological data and events during CL.

### 5.3.2 Data analysis bias requires recognition

Comprising a class of experimental tools per se, data analysis techniques apply rules of data aggregation, explore for relations within data, and decompose relations within data. Though infrequently comprehensively unpacked, data analysis methodologies are interwoven with theories. When examining the results of an experiment with a given theory, there is bias present from that perspective, and an equal bias present with interpretation from an alternate

theoretical perspective. Are researchers investigating inter-individual differences, inter-group differences, the landscape of connections (matrix analysis), or exploring grand averaged data? Are they interested in temporal, spatial, or level-based dynamics? After testing, interpreting statistical significance must be accompanied by recognition of the host of inherent assumptions, such as the data distribution qualities, the scale of the data, and the importance and roles of random factors.

The origins and properties of data analysis methodologies must be carefully considered when questioning research reliability. Importantly, the suitability of data analysis is inherently contingent on data attributes such as normality, independence, scale, and granularity, which are impacted by the experimental design, sources of data collection, data handling, and theoretical perspectives.

### 5.3.3 Example: EDA data analysis in studies of emotion and future potential

Since EDA data often represents a diverse range of states and responses that are nearly impossible or highly challenging to disentangle, the use of multimodal data, especially biodata, provides another promising avenue. Indeed, the vast majority (10/13) of studies investigating EDA also tracked EEG and EMG signals, eye movements, heart rate, and heart rate variability. However, to address ambiguities associated with EDA biodata usage, researchers benefit from triangulating EDA and other separately collected biodata streams. Meaning, the use of multimodal biodata that are measures (at least in part) of the same cognitive or emotional process fortifies conclusions. To illustrate constraints of current data collection and analysis in education, an example of measuring emotion valence through EDA is elaborated upon.

Of the 26 studies identified in education, 8 investigated emotions, mostly as related to metacognitive monitoring events (Ahonen et al., 2018; Haataja et al., 2018; 2021; Dindar et al., 2019; Malmberg et al., 2019a; 2019b; Sobocinski et al., 2020; Törmänen et al., 2021). Besides the qualitative coding of emotions by Törmänen and colleagues (2021), the remaining studies used mostly EDA to signal emotion level or valence, save for Malmberg and colleagues (2019a) who

used facial expression recognition from video for valence data. Of the 3 studies that investigated valence related to EDA, 2 used EDA changes to mark transcripts for valence coding of utterances.

Emotions play an important role in collaborating, and valence, a key dimension to distinguish between emotions (Fontaine et al., 2007), is a measure of great interest in CL research (e.g., Ahonen et al., 2018; Malmberg et al., 2019a; Sobocinski et al., 2020). Traditional manual coding of facial expressions and/or utterances may not be able to uncover (nearly or completely) invisible emotional reactions, are subjective (and require reliability analysis) and time-consuming. Using biodata, such as EDA, shows potential to circumvent these issues by providing an objective, automatic, and fast method to assess group valence. In the first study to do this in education, Ahonen and colleagues (2018) provided a novel approach of valence detection from 10s time lagged grand averaged SCRs and their minimum width envelope (MWE) confidence bands from 2 rooms (N=14 and N=16). Researchers dispute the use of EDA signal for understanding emotional valence (Imai, 2010), and while Ahonen and colleagues (2018) seem to agree that individual subject data is too noisy, they argue group analyses that could extract team performance are valuable, stable, and interpretable (as they clearly reflect sympathetic activation). They found that this approach is superior to extracting emotional valence from HRV data because EDA responses occur quickly (1-3 s) after onset and as such, can be time-locked to events. Also, EDA showed larger effect sizes compared to HRV. Group-level SCRs provided evidence for anticipatory relief and reactive frustration around a run/test programming event, based on NASA-Task Load Index (TLX) post-session questionnaires (non-significantly different between groups). A crucial component of their methodology that allowed them to achieve apparent detection of valence from SCR was their novel (to the field) use of MWE, which allowed calculation of confidence bands in grand averaged, time series data with high temporal resolution, while controlling for the autocorrelation in time series data, and without reliance on bootstrapping and the assumptions that coincide with it. Taken together, Ahonen and colleagues (2018) combination of methods to investigate event-locked SCR has moved the education field closer to measuring collaboration in natural settings.

Yet, there are current limits. This study suggests group-level SCR synchrony could be used *post-dictively*, e.g., 10 s post-event, to understand group-level emotional responses to an event. However, variations among the group-level data indicate problems with its *post-dictive* utility, but as it was not the aim of this study, a small number of groups precluded analysis of between-group effects. Further, it should be noted that although the authors allude to practitioners using such an approach for real-time assessment of classroom collaboration events, there are many technical and computational hurdles to overcome before that is within sight. Still, the potential for biodata streams to be used in this fashion exists and is the aspiration stated in the learning analytics literature (e.g., Ahonen et al., 2018, Schneider & Pea, 2013; 2015; 2020). While further development of Ahonen and colleagues' (2018) set of methods is an avenue to improve group-level valence detection, it has not been explored in the three years since, nor can it address individual-level valence, which could provide valuable insights into emotion contagion (Järvelä et al., 2014) or co-regulation of emotions in a group (Hadwin et al., 2017).

Outside of the education field, this review identified several studies dealing with emotion level detection (Anaya et al., 2021, Balconi et al., 2020; Balconi & Fronda, 2020; Kraus et al., 2020), but only one dealing specifically with emotion valence classification, achieved through EMG data from the zygomaticus major (smile) and corrugator supercilii (frown) (Mønster et al., 2016).

Taking emotion valence as an example, through reading the texts as part of this systematic review, EEG was identified as a promising tool for classifying emotion valence, though it has not been used in the human social interaction context as far as the methods of this review can detect. Balconi and colleagues (2020), though not themselves utilizing these features of EEG signal, indicate that the hemispheric lateralization model of emotions provides clear evidence of alpha band modulation based on emotion valence (left: positive valence, right: negative valence) first found by Sutton and Davidson (1997). EEG data offers the same benefits as EDA of being objective, quickly responsive, and having the ability to detect the otherwise undetectable. In affective neuroscience, use of EEG to detect emotional valence is developed enough to use machine learning algorithms (MLA) for automatic classification (see Suhaimi et al., 2020 for a



review), and has even been combined with facial expression recognition software for real-time application (Hassouneh et al., 2020).

The practicalities of implementation should be briefly addressed. Notably, the same low-cost, wireless EEG devices used in education were used in most affective neuroscience studies from 2016-2019, and this equipment combined with MLAs achieved high recognition accuracies (up to >90%) (Suhaimi et al., 2020). In all, these features of EEG methods provide a plausible application to collaborative learning or other social interaction contexts in the education field. At least, EEG data can objectively indicate individual-level valence, be compared with other biodata streams such as EDA or EMG, potentially triangulate with these additional channels, and be compared or triangulated with subjective data such as qualitative coding of valence. At best, EEG combined with MLA could provide real-time data on valence in collaborative activities as part of a dashboard for practitioners.

#### *5.4 Data handling bias and methods for reduction*

The studies in and outside education lacked standardized methodologies for defining and processing physiological data. Though many studies that examine a type of physiological data and other biodata often share aims, they frequently employ differing methods. To establish baseline EEG signal for example, Pérez et al. (2019) correlated all combinations of electrode signals, Santamaria et al. (2020) main-task data was z-normalized with baseline task data, Soto-Icaza et al. (2019) first filtered the data then defined baseline as -500 ms to 0 ms before stimulus onset time-frequency analysis, but -300 ms to 0 s for ERP analysis baseline, and Wagner-Altendorf et al. (2020) defined baseline as -100 ms to 0 ms before stimulus onset. Though all studies subsequently performed a form of baseline subtraction to observe signal associated with a stimulus or conditions, methodology was diverse. Of all the studies identified in this review, only few studies from the same first author used the same data handling approaches (see figure 5).

However, there are justifications for why this might be the case. First, the apparatuses used are often dissimilar across studies, which output differing raw data in terms of quality (noise levels) and sampling rate. Second, the

environments in which the studies are conducted differ in terms of sound and light, both of which can affect EEG and ECG signal quality; only two studies ensured rooms were dimly-lit and quiet (Hoehl et al., 2014; Vanderhasselt et al., 2018). As well, participant movement and speech varied greatly across studies, which can result in artifacts for EEG, fNIRS, EDA, and ECG data.

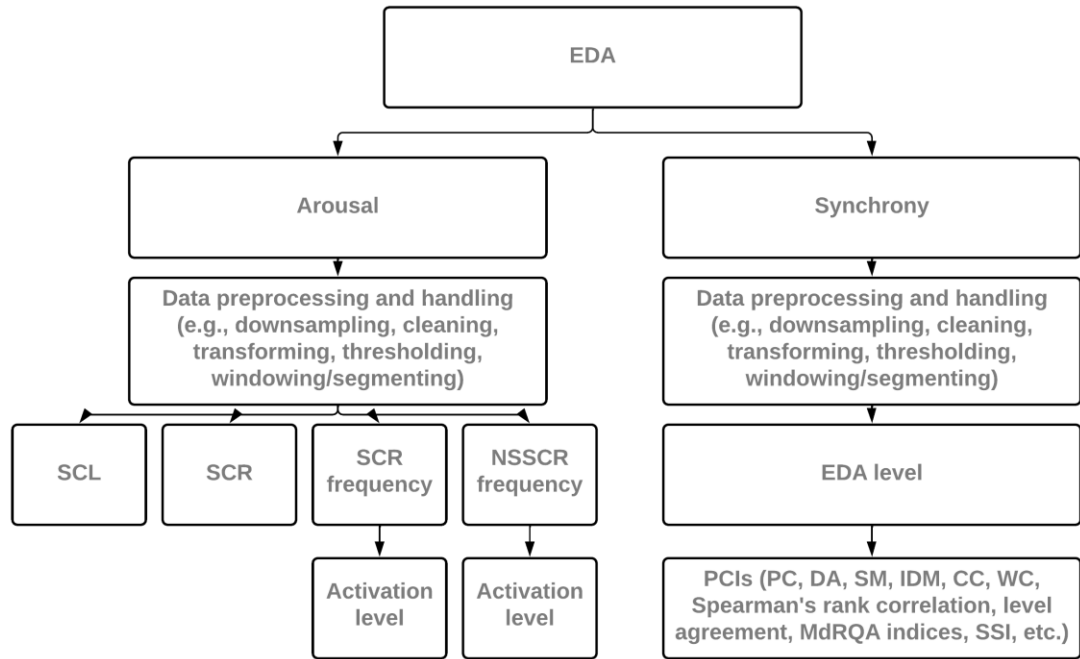
Preprocessing of data typically included downsampling to match frequencies of eye-trackers to video framerates (e.g., Cañigüeral et al., 2020), or to make data handling more manageable (e.g., raw EEG data at 5 kHz downsampled to 250 Hz) (Kraus et al., 2020).

While some decisions are justifiable, others are seemingly idiosyncratic, and future studies should clarify the reasoning for data handling decisions. Back in education, Malmberg and colleagues (2019a) transformed EDA data, set thresholds, and set a seemingly arbitrary 1-min window for segmenting time series data. To analyze skin conductance (SC) synchrony, Haataja et al. (2018) compared moment-by-moment average SC slopes of students for synchrony, used moving 5-s windows, ran correlations between students in 15-s windows, computed the ratio of sum of positive correlations over absolute value negative correlations, then natural log transformed these values for concordance over a given period for single session index (SSI). To investigate temporal changes in EDA synchrony and see possible co-occurrence with students monitoring, they looked at single session index (SSI) in 120-s moving windows, where all 3 possible pairs' synchrony was averaged. Averaging here is beneficial for assessing overall synchrony but ignores potential intragroup differences in synchronous activity; what if two members were synchronized but another was a free rider? This method cannot detect it. Further, a free rider effect would dramatically reduce the mean synchrony value and affect findings.

Processing procedures are often heavily impacted by the data analysis choice, and this was the justification most often given. For instance, studies that decided *a priori* to use MdRQA analysis for example must downsample to a low frequency (4 Hz or 10 Hz), standardize data, and decompose data with adaptive smoothing (Haataja et al., 2021; Wallot et al., 2016). Although these choices are not erroneous, they are researcher-dependent. Even with the two studies cited above, Haataja et al. referenced Wallot et al., yet chose to downsample differently. What are the effects of the idiosyncrasies across all the reviewed

studies in terms of results? Hypothetically, researchers choosing different processing procedures (e.g., see figure 5) open the possibility for varying results. Future research might pre-register parameter decisions and rationales should be a standard inclusion in high-quality peer-reviewed articles. In addition, varying parameters such as thresholds, window duration, step duration, etc. should be integrated into analyses to indicate the variance of results based on variance of methodologies, providing an “error bar” effect. Or standardized tasks for gathering baseline data (e.g., Santamaria et al., 2020) for reference to define within and between-subject normal states. Making explicit these measurements and procedures could improve investigators’ abilities to improve the signal to noise ratio, draw fair comparisons between studies, and better mutually-support findings to home in on the ability of physiological data to indicate events or measure outcomes.

In addition to the processing components of handling, crucially, the features of the raw data chosen to be analyzed has a large impact on analysis approaches, results, and conclusions. EDA has various components of interest for analysis, and varying ways to process them before analysis (see figure 5). Taking EDA outside education for example, some studies chose to examine SCL as a measure of acute stress response (Potts et al., 2019), while another used SCR to assess arousal during reciprocated gaze as an indicator of attachment and avoidance characteristics (Prinsen et al., 2019), but Kraus and colleagues (2020) observed NSSCR changes as a measure of emotional arousal and apprehension, noting that SCL measures other phenomena such as cognitive load.



**FIGURE 5.** The divergences of EDA data handling

In education, EDA arousal and synchrony definitions are equally diverse. Haataja et al. (2021) used EDA slope to indicate arousal and %DET (an MdRQA index) to mark physiological synchrony, whereas years prior, Haataja et al. (2018) compared EDA slope moment-by-moment comparisons to measure synchrony (SSIs computed as described above). Pijeira-Díaza et al. (2019) used SCR peak frequency to assess activation level, and medium and high levels were combined, but Törmänen et al. (2021) used NSSCR peak frequency, where low, medium, and high levels were established. Two studies took a shotgun approach to observing differences in physiological synchrony from EDA slope physiological concordance indices; Schneider et al. (2020) chose Pearson correlation, DA, SM, and IDM, where Montague et al. (2014) chose DA, SM, IDM, CC, and WC, which are all varying ways to determine relations between two slopes. The benefits to this approach are that something will stick and that the indices may ; for instance, Schneider et al. found DA to be best correlated to dimensions of collaboration quality and Pearson correlation best related to learning gains. The disadvantage is that there is virtually nonexistent theoretical understanding of their meaning, and by extension no rationale for choosing one over another. As seen in table 5,

EDA data for example can be associated with several measures, from emotional to cognitive. Though the choices to use various components of EDA for example are not “wrong,” the specifics of each method are flimsy. The vast literature allows researchers to choose methodologies from a diverse assortment, and especially when justifications are not given, leave readers in a position of ignorance about the why and doubt about the findings. In short, unjustified, inconsistent uses and interpretations of the data limit the strength of the research.

Like data processing, perhaps decisions regarding biodata component selection should be pre-registered and rationales explained. Additionally, researchers could also compute physiological data according to differing methods found in the literature to reveal the impact these differences could have on results. Clear indications of the methods and their rationale might afford researchers the ability to compare and critique them more openly, potentially providing a route for improvement upon them or selection of standards. As well, studies with similar methodologies may allow for appropriate comparison of findings and ultimately close in on understanding the roles biodata is apt to take when investigating outcomes or cognitive, emotional, and attentional processes during social interactions.

# 6 CONCLUSION

This systematic review described the integration of physiological and biodata to study human social interactions in the education field, of which roughly a third investigated EDA and a third eye-tracking to gain insights into the relations between biodata and cognitive and emotional processes during CL. Education research that utilizes biodata are pushing the field forward in respect to understanding the physiological and embodied underpinnings of CL elements and processes such as metacognitive monitoring, adaptive regulations, and socioemotional state.

Examination of findings revealed a high potential for biodata use in predicting these features of CL, as well as its quality, task performance, and learning gains. Some studies employed powerful data analysis techniques that enable testing of necessarily complex hypotheses based on educational theories. Machine learning algorithms have also shown ability to accurately classify participant characteristics from biodata and have strong prospects for development.

This systematic review also identified current gaps related to paradigmatic ambiguities from experimental design, data collection, data handling and data analysis. Researchers should consider the confusion matrix to account for all 4 pairs of crossovers through strategic experimental design, sources of data collection, and data analysis. Multimodal data especially offers the ability to remove potentially confounding factors and triangulate data for verification. Standardization of protocol for signal selection, thresholding, and data processing, and statistical test usage for application of educational theories would help reduce current biases, hence improving the validity and reliability of studies in the education field.

In view of the strengths and weaknesses of the state of the art and the resultant suggestions for the future, this thesis makes a strong argument for increasing interdisciplinary collaboration in this area. Expertise in physiology,

neurophysiology, computer science, and data science are required to apply paradigms that can live up to the educational theories they work under (Hilpert & Marchand, 2018). With such collaborative effort, the utility of physiological data in education will expand and drive the field toward a deeper apprehension of learning in social contexts.

# 7 FINAL CONSIDERATIONS

## *7.1 Limitations*

This systematic review describes research contained in the databases Education Collection (ProQuest), Education Resources Complete (ProQuest), Teacher Reference Center (ProQuest), PsycINFO (Ovid), PubMed (NCBI), and ScienceDirect (Elsevier), as well as from the American Journal of Physiology – Advances in physiology education, and Google Scholar to collect any missing studies. The articles gathered are from the last decade and are peer-reviewed. This approach intended to gather all the high-quality works involving physiological and biodata in human social interactions but is not a representation of all the works published on this topic. The judgment is made that peer-reviewed journal articles are “high-quality” and hence “valid”, in line with the view of the academic community, though researchers have found evidence for journals with low-standards in the peer-review process (Eddie & Conklin, 2019). Several conference papers were identified with intriguing results, most of which were addressed in the introduction of the thesis but were not subject to further analysis.

When critically evaluating the quality of research, validity and reliability are the two main features to reflect on. A study with high validity has examined what it intended. A study with high reliability is trustworthy through its repeatability. But the starting point for repeating an SLR would have a large impact on results. If a researcher followed the protocol described in this thesis, the reliability would likely be far higher than if they were merely provided with the same research questions and databases. SLRs conducted by novice researchers may have low reliability if provided with the same research questions and databases (Kitchenham et al., 2011). An author with different key words for the same phenomena could have collected a different data set, though this effect would be larger on the literature outside education. In



consideration of the limitations of this study and increase its reliability and validity, effort was put forth to maximize transparency, objectivity, and repeatability through detailing every step of the process and following PRISMA guidelines and diagrammatic representation (Liberati et al., 2009). In this thesis, the analysis of every study in the education field was provided with the background of study summaries and descriptions in the appendices and results section. Interpretation of the study's qualities and findings was transparent, allowing evaluation of legitimacy.

## *7.2 Reflections on ethics*

The research of this thesis was conducted with scientific and defined methods and reported findings openly and honestly (Kuula, 2011, p. 26). The literature analyzed does not contain sensitive or personally identifiable information, and thus avoids the ethical issues related to subject involvement such as anonymity or consent. The process of this SLR had to be handled responsibly to ensure that the data is accurately represented and analyzed. All the articles are available to those with database access, and full data extracts are available upon request, as all data was saved at every stage.

Since systematic reviews are often read and cited in documents that may influence future research, this review paid careful attention to representing the literature accurately and made explicit the limitations of the study in terms of methods and results (Suri, 2018). SLR analysis is partially subjective by virtue of author interpretation. Importantly, effort was made to bring to light the strengths and weaknesses of studies and the state of the education field through systematic reflection. This thesis purposely informed the reader of the selective inclusivity of the SLR with justification. The review's findings were communicated with audience-appropriate transparency, while keeping within the requirements of the faculty for academic writing.

# REFERENCES

- Abrahamson, D., & Sánchez-García, R. (2016). Learning Is Moving in New Ways: The Ecological Dynamics of Mathematics Education. *Journal of the Learning Sciences*, 25(2), 203–239. <https://doi.org/10.1080/10508406.2016.1143370>
- Agrafioti, F., Hatzinakos, D., & Anderson, A. K. (2012). ECG Pattern Analysis for Emotion Detection. *IEEE Transactions on Affective Computing*, 3(1), 102–115. <https://doi.org/10.1109/T-AFFC.2011.28>
- Ahn, S., Cho, H., Kwon, M., Kim, K., Kwon, H., Kim, B. S., Chang, W. S., Chang, J. W., & Jun, S. C. (2017). Interbrain phase synchronization during turn-taking verbal interaction—A hyperscanning study using simultaneous EEG/MEG. *Human Brain Mapping*, 39(1), 171–188. <https://doi.org/10.1002/hbm.23834>
- Ahonen, L., Cowley, B., Torniainen, J., Ukkonen, A., Vihavainen, A., & Puolamäki, K. (2016). Cognitive Collaboration Found in Cardiac Physiology: Study in Classroom Environment. *PLoS One*, 11(7), e0159178. <http://dx.doi.org/10.1371/journal.pone.0159178>
- Ahonen, L., Cowley, B. U., Hellas, A., & Puolamäki, K. (2018). Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment. *Scientific Reports*, 8(1), 3138. <https://doi.org/10.1038/s41598-018-21518-3>
- Allen, C. (2017). On (not) defining cognition. *Synthese*, 194(11), 4233–4249. <https://doi.org/10.1007/s11229-017-1454-4>

- Anaya, B., Vallorani, A., & Pérez-Edgar, K. (2021). Dyadic behavioral synchrony between behaviorally inhibited and non-inhibited peers is associated with concordance in EEG frontal Alpha asymmetry and Delta-Beta coupling. *Biological Psychology*, 159, 108018. <https://doi.org/10.1016/j.biopsycho.2021.108018>
- Anderson, J. R. (2005). *Cognitive Psychology and Its Implications*. Macmillan.
- Apanovich, V. V., Bezdenezhnykh, B. N., Sams, M., Jääskeläinen, I. P., & Alexandrov, Y. (2018). Event-related potentials during individual, cooperative, and competitive task performance differ in subjects with analytic vs. Holistic thinking. *International Journal of Psychophysiology*, 123, 136–142. <https://doi.org/10.1016/j.ijpsycho.2017.10.001>
- Arroyo, I., Cooper, D., Burlison, W., Woolf, B., Muldner, K., & Christopherson, R. (2009). *Emotion Sensors Go To School*. 200. <https://doi.org/10.3233/978-1-60750-028-5-17>
- Arslan Aydin, Ü., Kalkan, S., & Acartürk, C. (2021). Speech Driven Gaze in a Face-to-Face Interaction. *Frontiers in Neurobotics*, 15. <https://www.frontiersin.org/article/10.3389/fnbot.2021.598895>
- Artino, A. R., Holmboe, E. S., & Durning, S. J. (2012). Can achievement emotions be used to better understand motivation, learning, and performance in medical education? *Medical Teacher*, 34(3), 240–244. <https://doi.org/10.3109/0142159X.2012.643265>
- Azevedo, R. (2015). Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues. *Educational Psychologist*, 50(1), 84–94. <https://doi.org/10.1080/00461520.2015.1004069>

- Azevedo, R., & Gašević, D. (2019). Analyzing Multimodal Multichannel Data about Self-Regulated Learning with Advanced Learning Technologies: Issues and Challenges. *Computers in Human Behavior*, *96*, 207–210.  
<https://doi.org/10.1016/j.chb.2019.03.025>
- Azriel, O., Lazarov, A., Segal, A., & Bar-Haim, Y. (2020). Visual attention patterns during online video-mediated interaction in socially anxious individuals. *Journal of Behavior Therapy and Experimental Psychiatry*, *69*, 101595.  
<https://doi.org/10.1016/j.jbtep.2020.101595>
- Baker, M. (2003). Computer-Mediated Argumentative Interactions for the Co-Elaboration of Scientific Notions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (pp. 47–78). Springer Netherlands.  
[https://doi.org/10.1007/978-94-017-0781-7\\_3](https://doi.org/10.1007/978-94-017-0781-7_3)
- Baker, M., Lyon (f, U. L., Hansen, T., Joiner, R., & Traum, D. (1999). *Chapter 3 THE ROLE OF GROUNDING IN COLLABORATIVE LEARNING TASKS*.
- Baker, R., & Siemens, G. (2014). *Educational data mining and learning analytics* (pp. 253–272). <https://doi.org/10.1017/CBO9781139519526.016>
- Balconi, M., Crivelli, D., & Vanutelli, M. E. (2017). Why to cooperate is better than to compete: Brain and personality components. *BMC Neuroscience*, *18*(Babiloni, F., Cincotti, F., Mattia, D., De Vico Fallani, F., Tocci, A., Bianchi, L., et al. (2007). High resolution EEG hyperscanning during a card game. In: 29th Annual international conference on engineering, medicine and biology society, 2007. EMBS 2007. IEEE. 2007;4957-60.). <https://doi.org/10.1186/s12868-017-0386-8>

- Balconi, M., & Fronda, G. (2021). Inter-brain EEG connectivity in hyperscanning for Italian and French gestures: The culture-related nonverbal language. *Culture and Brain*. <https://doi.org/10.1007/s40167-021-00103-y>
- Balters, S., & Steinert, M. (2017). Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices. *Journal of Intelligent Manufacturing*, 28(7), 1585–1607. <https://doi.org/10.1007/s10845-015-1145-2>
- Barraza, P., Pérez, A., & Rodríguez, E. (2020). Brain-to-Brain Coupling in the Gamma-Band as a Marker of Shared Intentionality. *Frontiers in Human Neuroscience*, 14. <https://www.frontiersin.org/article/10.3389/fnhum.2020.00295>
- Bartholow, B. D. (2010). Event-Related Brain Potentials and Social Cognition: On Using Physiological Information to Constrain Social Cognitive Theories. *Social Cognition*, 28(6), 723–747. <https://doi.org/10.1521/soco.2010.28.6.723>
- Barzy, M., Ferguson, H. J., & Williams, D. M. (2020). Perspective influences eye movements during real-life conversation: Mentalising about self versus others in autism. *Autism: The International Journal of Research and Practice*, 24(8), 2153–2165. <https://doi.org/10.1177/1362361320936820>
- Bates, E., Thal, D., Whitesell, K., Fenson, L., & et al. (1989). Integrating language and gesture in infancy. *Developmental Psychology*, 25(6), 1004–1019. <https://doi.org/10.1037/0012-1649.25.6.1004>
- Behrens, F., Snijdwint, J. A., Moulder, R. G., Prochazkova, E., Sjak-Shie, E. E., Boker, S. M., & Kret, M. E. (2020). Physiological synchrony is associated with cooperative success in real-life interactions. *Scientific Reports*, 10(1), 19609. <https://doi.org/10.1038/s41598-020-76539-8>

- Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., & Dikker, S. (2019). Brain-to-Brain Synchrony and Learning Outcomes Vary by Student–Teacher Dynamics: Evidence from a Real-world Classroom Electroencephalography Study. *Journal of Cognitive Neuroscience*, 31(3), 401–411. [https://doi.org/10.1162/jocn\\_a\\_01274](https://doi.org/10.1162/jocn_a_01274)
- Birze, A., LeBlanc, V., Regehr, C., Paradis, E., & Einstein, G. (2020). The “Managed” or Damaged Heart? Emotional Labor, Gender, and Posttraumatic Stressors Predict Workplace Event-Related Acute Changes in Cortisol, Oxytocin, and Heart Rate Variability. *Frontiers in Psychology*, 11, 604. <https://doi.org/10.3389/fpsyg.2020.00604>
- Boland, A., Cherry, G., & Dickson, R. (Eds.). (2017). Doing a systematic review: a student’s guide (2nd ed.). SAGE Publications. [https://uk.sagepub.com/sites/default/files/upm-assets/87631\\_book\\_item\\_87631.pdf](https://uk.sagepub.com/sites/default/files/upm-assets/87631_book_item_87631.pdf)
- Boshuizen, H. P. A., & Tabachneck-Schijf, J. M. (1998). Problem solving with multiple representations by multiple and single agents: an analysis of the issues involved. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen, & T. de Jong (Eds.), *Learning with multiple representations* (pp. 137–151). Elsevier.
- Boucsein, W. (2012). *Electrodermal Activity*. Springer Science & Business Media.
- Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3), 1465–1477. <https://doi.org/10.1016/j.cognition.2007.05.012>
- Brennan, S., & Hanna, J. (2009). Partner-Specific Adaptation in Dialog. *Top. Cogn. Sci.* <https://doi.org/10.1111/j.1756-8765.2009.01019.x>

- Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*, 10(3), 169–175.  
<https://doi.org/10.1002/mrdd.20036>
- Bulu, S. T. (2012). Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education*, 58(1), 154–161.  
<https://doi.org/10.1016/j.compedu.2011.08.024>
- Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (2007). *Handbook of Psychophysiology*. Cambridge University Press.
- Cañigueral, R., & Hamilton, A. F. de C. (2019). Being watched: Effects of an audience on eye gaze and prosocial behaviour. *Acta Psychologica*, 195, 50–63.  
<https://doi.org/10.1016/j.actpsy.2019.02.002>
- Cartaud, A., Ruggiero, G., Ott, L., Iachini, T., & Coello, Y. (2018). Physiological Response to Facial Expressions in Peripersonal Space Determines Interpersonal Distance in a Social Interaction Context. *Frontiers in Psychology*, 9. <https://www.frontiersin.org/article/10.3389/fpsyg.2018.00657>
- Caruana, N., Alhasan, A., Wagner, K., Kaplan, D. M., Woolgar, A., & McArthur, G. (2020). The effect of non-communicative eye movements on joint attention. *Quarterly Journal of Experimental Psychology*, 73(12), 2389–2402.  
<https://doi.org/10.1177/1747021820945604>
- Chanel, G., Betrancourt, M., Pun, T., Cereghetti, D., & Molinari, G. (2013). Assessment of Computer-Supported Collaborative Processes Using Interpersonal Physiological and Eye-Movement Coupling. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 116–122. <https://doi.org/10.1109/ACII.2013.26>

- Chanel, G., Kivikangas, J. M., & Ravaja, N. (2012). Physiological compliance for social gaming analysis: Cooperative versus competitive play. *Interacting with Computers*, 24(4), 306–316. <https://doi.org/10.1016/j.intcom.2012.04.012>
- Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, 15(4), 481–498. [https://doi.org/10.1016/S0885-2014\(01\)00037-5](https://doi.org/10.1016/S0885-2014(01)00037-5)
- Chen, Y., Lu, J., Wang, Y., Feng, Z., & Yuan, B. (2017). Social distance influences the outcome evaluation of cooperation and conflict: Evidence from event-related potentials. *Neuroscience Letters*, 647, 78–84. <https://doi.org/10.1016/j.neulet.2017.03.018>
- Cherubini, M., Nüssli, M.-A., & Dillenbourg, P. (2008). Deixis and gaze in collaborative work at a distance (over a shared map): A computational model to detect misunderstandings. *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, 173–180. <https://doi.org/10.1145/1344471.1344515>
- Chi, M. T. H. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science*, 1(1), 73–105. <https://doi.org/10.1111/j.1756-8765.2008.01005.x>
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. <https://doi.org/10.1037/10096-006>
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39. [https://doi.org/10.1016/0010-0277\(86\)90010-7](https://doi.org/10.1016/0010-0277(86)90010-7)



- Cohen, E. G. (1994). Restructuring the Classroom: Conditions for Productive Small Groups. *Review of Educational Research*, 64(1), 1–35.  
<https://doi.org/10.3102/00346543064001001>
- Cooper, M. A. (1999). Classroom choices from a cognitive perspective on peer learning. In *Cognitive perspectives on peer learning* (pp. 215–233). Lawrence Erlbaum Associates Publishers.
- Critchley, H. D. (2002). Review: Electrodermal Responses: What Happens in the Brain. *The Neuroscientist*, 8(2), 132–142.  
<https://doi.org/10.1177/107385840200800209>
- Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current Opinion in Psychology*, 17, 7–14. <https://doi.org/10.1016/j.copsyc.2017.04.020>
- Czeszumski, A., Ehinger, B. V., Wahn, B., & Konig, P. (2019). The social situation affects how we process feedback about our actions. *Frontiers in Psychology*, 10(Social Psychology [3000]). <https://doi.org/10.3389/fpsyg.2019.00361>
- Damon, W. (1984). Peer education: The untapped potential. *Journal of Applied Developmental Psychology*, 5(4), 331–343. [https://doi.org/10.1016/0193-3973\(84\)90006-6](https://doi.org/10.1016/0193-3973(84)90006-6)
- Damon, W., & Phelps, E. (1989). Critical distinctions among three approaches to peer education. *International Journal of Educational Research*, 13(1), 9–19.  
[https://doi.org/10.1016/0883-0355\(89\)90013-X](https://doi.org/10.1016/0883-0355(89)90013-X)
- Dawson, G., Munson, J., Estes, A., Osterling, J., McPartland, J., Toth, K., Carver, L., & Abbott, R. (2002). Neurocognitive Function and Joint Attention Ability in Young Children with Autism Spectrum Disorder Versus Developmental Delay. *Child Development*, 73(2), 345–358. <https://doi.org/10.1111/1467-8624.00411>

- De Backer, L., Van Keer, H., & Valcke, M. (2015). Exploring evolutions in reciprocal peer tutoring groups' socially shared metacognitive regulation and identifying its metacognitive correlates. *Learning and Instruction, 38*, 63–78.  
<https://doi.org/10.1016/j.learninstruc.2015.04.001>
- de Bruijn, E. R. A., Miedl, S. F., & Bekkering, H. (2011). How a co-actor's task affects monitoring of own errors: Evidence from a social event-related potential study. *Experimental Brain Research, 211*(3), 397.  
<https://doi.org/10.1007/s00221-011-2615-1>
- de Bruin, A. B. H., Thiede, K. W., Camp, G., & Redford, J. (2011). Generating keywords improves metacomprehension and self-regulation in elementary and middle school children. *Journal of Experimental Child Psychology, 109*(3), 294–310. <https://doi.org/10.1016/j.jecp.2011.02.005>
- De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences, 6*(4), 485–507. <https://doi.org/10.1007/s11097-007-9076-9>
- De Lillo, M., Foley, R., Fysh, M. C., Stimson, A., Bradford, E. E. F., Woodrow-Hill, C., & Ferguson, H. J. (2021). Tracking developmental differences in real-world social attention across adolescence, young adulthood and older adulthood. *Nature Human Behaviour, 5*(10), 1381–1390. <https://doi.org/10.1038/s41562-021-01113-9>
- Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., Ding, M., & Poeppel, D. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology, 27*(9), 1375–1380.  
<https://doi.org/10.1016/j.cub.2017.04.002>

- Dindar, M., Alikhani, I., Malmberg, J., Järvelä, S., & Seppänen, T. (2019). Examining shared monitoring in collaborative learning: A case of a recurrence quantification analysis approach. *Computers in Human Behavior, 100*, 335–344. <https://doi.org/10.1016/j.chb.2019.03.004>
- D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, Analytic, Automated (AAA) Measurement of Engagement During Learning. *Educational Psychologist, 52*, 1–20. <https://doi.org/10.1080/00461520.2017.1281747>
- D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153–170. <https://doi.org/10.1016/j.learninstruc.2012.05.003>
- Duffy, M. C., Azevedo, R., Sun, N.-Z., Griscom, S. E., Stead, V., Crelinsten, L., Wiseman, J., Maniatis, T., & Lachapelle, K. (2015). Team regulation in a simulated medical emergency: An in-depth analysis of cognitive, metacognitive, and affective processes. *Instructional Science, 43*(3), 401–426. <https://doi.org/10.1007/s11251-014-9333-6>
- Edie, A. H., & Conklin, J. L. (2019). Avoiding predatory journals: Quick peer review processes too good to be true. *Nursing Forum, 54*(3), 336–339. <https://doi.org/10.1111/nuf.12333>
- Ellamil, M., Berson, J., & Margulies, D. S. (2016). Influences on and Measures of Unintentional Group Synchrony. *Frontiers in Psychology, 7*. <https://www.frontiersin.org/article/10.3389/fpsyg.2016.01744>
- Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utrainen, K., & Kyngäs, H. (2014). Qualitative Content Analysis: A Focus on Trustworthiness. *SAGE Open, 4*(1), 2158244014522633. <https://doi.org/10.1177/2158244014522633>

- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. <https://doi.org/10.1111/j.1365-2648.2007.04569.x>
- Eriksson, M., & Lindström, B. (2006). Antonovsky's sense of coherence scale and the relation with health: A systematic review. *Journal of Epidemiology and Community Health*, 60(5), 376–381. <https://doi.org/10.1136/jech.2005.041616>
- Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, 56(2), 171–184. <https://doi.org/10.1016/j.ijpsycho.2004.11.003>
- Ferguson, H. J., & Breheny, R. (2011). Eye movements reveal the time-course of anticipating behaviour based on complex, conflicting desires. *Cognition*, 119(2), 179–196. <https://doi.org/10.1016/j.cognition.2011.01.005>
- Fishman, I., & Ng, R. (2013). Error-related brain activity in extraverts: Evidence for altered response monitoring in social context. *Biological Psychology*, 93(1), 225–230. <https://doi.org/10.1016/j.biopsycho.2013.02.010>
- Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The World of Emotions is not Two-Dimensional. *Psychological Science*, 18(12), 1050–1057. <https://doi.org/10.1111/j.1467-9280.2007.02024.x>
- Freeth, M., Foulsham, T., & Kingstone, A. (2013). What Affects Social Attention? Social Presence, Eye Contact and Autistic Traits. *PLOS ONE*, 8(1), e53286. <https://doi.org/10.1371/journal.pone.0053286>
- Fritz, T., Begel, A., Müller, S. C., Yigit-Elliott, S., & Züger, M. (2014). Using psychophysiological measures to assess task difficulty in software development.

*Proceedings of the 36th International Conference on Software Engineering*,  
402–413. <https://doi.org/10.1145/2568225.2568266>

Fu, X., Nelson, E. E., Borge, M., Buss, K. A., & Pérez-Edgar, K. (2019). Stationary and ambulatory attention patterns are differentially associated with early temperamental risk for socioemotional problems: Preliminary evidence from a multimodal eye-tracking investigation. *Development and Psychopathology*, 31(3), 971–988. <https://doi.org/10.1017/S0954579419000427>

Fujiki, M., Brinton, B., & Clarke, D. (2002). Emotion Regulation in Children With Specific Language Impairment. *Language, Speech, and Hearing Services in Schools*, 33(2), 102–111. [https://doi.org/10.1044/0161-1461\(2002/008\)](https://doi.org/10.1044/0161-1461(2002/008))

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. <https://doi.org/10.1007/s11528-014-0822-x>

Gillies, R. M., Carroll, A., Cunnington, R., Rafter, M., Palghat, K., Bednark, J., & Bourgeois, A. (2016). Multimodal representations during an inquiry problem-solving activity in a Year 6 science class: A case study investigating cooperation, physiological arousal and belief states. *Australian Journal of Education*, 60(2), 111–127. <http://dx.doi.org/10.1177/0004944116650701>

Goodwin, C. (Ed.). (2017). What Is Co-Operative Action, and Why Is It Important? In *Co-Operative Action* (pp. 1–20). Cambridge University Press.  
<https://doi.org/10.1017/9781139016735.001>

Haataja, E., Malmberg, J., Dindar, M., & Järvelä, S. (2021). The pivotal role of monitoring for collaborative problem solving seen in interaction, performance, and interpersonal physiology. *Metacognition and Learning*.  
<https://doi.org/10.1007/s11409-021-09279-3>

- Haataja, E., Malmberg, J., & Järvelä, S. (2018). Monitoring in collaborative learning: Co-occurrence of observed behavior and physiological synchrony explored. *Computers in Human Behavior*, *87*, 337–347. <https://doi.org/10.1016/j.chb.2018.06.007>
- Hadwin, A., Järvelä, S., & Miller, M. (2017). Self-Regulation, Co-Regulation, and Shared Regulation in Collaborative Learning Environments. In *Handbook of Self-Regulation of Learning and Performance* (2nd ed.). Routledge.
- Hadwin, A., & Oshige, M. (2011). *Self-Regulation, Coregulation, and Socially Shared Regulation: Exploring Perspectives of Social in Self-Regulated Learning Theory*. <https://doi.org/10.1177/0161468111111300204>
- Haensel, J. X., Danvers, M., Ishikawa, M., Itakura, S., Tucciarelli, R., Smith, T. J., & Senju, A. (2020). Culture modulates face scanning during dyadic social interactions. *Scientific Reports*, *10*(1), 1958. <https://doi.org/10.1038/s41598-020-58802-0>
- Hajcak, G., McDonald, N., & Simons, R. F. (2003). Anxiety and error-related brain activity. *Biological Psychology*, *64*(1–2), 77–90. [https://doi.org/10.1016/s0301-0511\(03\)00103-0](https://doi.org/10.1016/s0301-0511(03)00103-0)
- Hanley, M., Riby, D. M., Carty, C., Melaugh McAteer, A., Kennedy, A., & McPhillips, M. (2015). The use of eye-tracking to explore social difficulties in cognitively able students with autism spectrum disorder: A pilot investigation. *Autism: The International Journal of Research and Practice*, *19*(7), 868–873. <https://doi.org/10.1177/1362361315580767>
- Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent

multi-agent system. *Computers in Human Behavior*, 48, 615–625.

<https://doi.org/10.1016/j.chb.2015.02.013>

Harley, J. M., Jarrell, A., & Lajoie, S. P. (2019). Emotion regulation tendencies, achievement emotions, and physiological arousal in a medical diagnostic reasoning simulation. *Instructional Science*, 47(2), 151–180.

<https://doi.org/10.1007/s11251-018-09480-z>

Hassouneh, A., Mutawa, A. M., & Murugappan, M. (2020). Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked*, 20, 100372. <https://doi.org/10.1016/j.imu.2020.100372>

Heaphy, E. D., & Dutton, J. E. (2008). Positive Social Interactions and the Human Body at Work: Linking Organizations and Physiology. *Academy of Management Review*, 33(1), 137–162. <https://doi.org/10.5465/amr.2008.27749365>

Hernández-Gutiérrez, D., Abdel Rahman, R., Martín-Loeches, M., Muñoz, F., Schacht, A., & Sommer, W. (2018). Does dynamic information about the speaker's face contribute to semantic speech processing? ERP evidence. *Cortex*, 104, 12–25. <https://doi.org/10.1016/j.cortex.2018.03.031>

Hessels, R. S., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2017). Gaze behavior to faces during dyadic interaction. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, 71(3), 226–242. <https://doi.org/10.1037/cep0000113>

Hessels, R. S., Holleman, G. A., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2018). Eye contact takes two – autistic and social anxiety traits predict gaze behavior in dyadic interaction. *Journal of Experimental Psychopathology*, 9(2), jep.062917. <https://doi.org/10.5127/jep.062917>

- Hietanen, J. O., Peltola, M. J., & Hietanen, J. K. (2020). Psychophysiological responses to eye contact in a live interaction and in video call. *Psychophysiology*, 57(6), e13587. <https://doi.org/10.1111/psyp.13587>
- Hilpert, J. C., & Marchand, G. C. (2018). Complex Systems Research in Educational Psychology: Aligning Theory and Method. *Educational Psychologist*, 53(3), 185–202. <https://doi.org/10.1080/00461520.2018.1469411>
- Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and Listening with the Eyes: Gaze Signaling during Dyadic Interactions. *PLOS ONE*, 10(8), e0136905. <https://doi.org/10.1371/journal.pone.0136905>
- Hoehl, S., Michel, C., Reid, V. M., Parise, E., & Striano, T. (2014). Eye contact during live social interaction modulates infants' oscillatory brain activity. *Social Neuroscience*, 9(3), 300–308. <https://doi.org/10.1080/17470919.2014.884982>
- Holkamp, Y. H., & Schavemaker, J. G. M. (2014). *A Comparison of Human and Machine Learning-based Accuracy for Valence Classification of Subjects in Video Fragments*. 4.
- Holleman, G. A., Hessels, R. S., Kemner, C., & Hooge, I. T. C. (2020). Implying social interaction and its influence on gaze behavior to the eyes. *PLOS ONE*, 15(2), e0229203. <https://doi.org/10.1371/journal.pone.0229203>
- Holper, L., Goldin, A. P., Shalóm, D. E., Battro, A. M., Wolf, M., & Sigman, M. (2013). The teaching and the learning brain: A cortical hemodynamic marker of teacher–student interactions in the Socratic dialog. *International Journal of Educational Research*, 59, 1–10. <https://doi.org/10.1016/j.ijer.2013.02.002>
- Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O’Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018



- for information professionals and researchers. *Education for Information*, 34(4), 285–291. <https://doi.org/10.3233/EFI-180221>
- Honma, M., Tanaka, Y., Osada, Y., & Kuriyama, K. (2012). Perceptual and not physical eye contact elicits pupillary dilation. *Biological Psychology*, 89(1), 112–116. <https://doi.org/10.1016/j.biopsycho.2011.09.015>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Hu, Y., Pan, Y., Shi, X., Cai, Q., Li, X., & Cheng, X. (2018). Inter-brain synchrony and cooperation context in interactive decision making. *Biological Psychology*, 133, 54–62. <https://doi.org/10.1016/j.biopsycho.2017.12.005>
- Huang, X., Kortelainen, J., Zhao, G., Li, X., Moilanen, A., Seppänen, T., & Pietikäinen, M. (2016). Multi-modal emotion analysis from facial expressions and electroencephalogram. *Computer Vision and Image Understanding*, 147, 114–124. <https://doi.org/10.1016/j.cviu.2015.09.015>
- Iiskala, T., Vauras, M., Lehtinen, E., & Salonen, P. (2011). Iiskala, T., Vauras, M., Lehtinen, E. & Salonen, P. (2011). Socially Shared Metacognition within Primary School Pupil Dyads' Collaborative Processes. *Learning and Instruction*, 21(3), 379–393. *Learning and Instruction*, 21, 379–393.
- Imai, Y. (2010). Emotions in SLA: New Insights From Collaborative Learning for an EFL Classroom. *The Modern Language Journal*, 94(2), 278–292. <https://doi.org/10.1111/j.1540-4781.2010.01021.x>
- Ioannou, S., Morris, P., Mercer, H., Baker, M., Gallese, V., & Reddy, V. (2014). Proximity and gaze influences facial temperature: A thermal infrared imaging study. *Frontiers in Psychology*, 5, 845. <https://doi.org/10.3389/fpsyg.2014.00845>

- Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (2009). Influence of group member familiarity on online collaborative learning. *Computers in Human Behavior*, 25(1), 161–170. <https://doi.org/10.1016/j.chb.2008.08.010>
- Järvelä, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational Psychologist*, 48(1), 25–39. <https://doi.org/10.1080/00461520.2012.748006>
- Järvelä, S., Kirschner, P. A., Hadwin, A., Järvenoja, H., Malmberg, J., Miller, M., & Laru, J. (2016). Socially shared regulation of learning in CSCL: Understanding and prompting individual- and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning*, 11(3), 263–280. <https://doi.org/10.1007/s11412-016-9238-2>
- Järvelä, S., Kivikangas, J. M., Kätsyri, J., & Ravaja, N. (2014). Physiological Linkage of Dyadic Gaming Experience. *Simulation & Gaming*, 45(1), 24–40. <https://doi.org/10.1177/1046878113513080>
- Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, 72, 101203. <https://doi.org/10.1016/j.learninstruc.2019.04.004>
- Järvenoja, H., & Järvelä, S. (2009a). Emotion control in collaborative learning situations: Do students regulate emotions evoked by social challenges/. *British Journal of Educational Psychology*, 79(3), 463–481. <https://doi.org/10.1348/000709909X402811>
- Järvenoja, H., & Järvelä, S. (2009b). Emotion control in collaborative learning situations: Do students regulate emotions evoked by social challenges/. *British*

*Journal of Educational Psychology*, 79(3), 463–481.

<https://doi.org/10.1348/000709909X402811>

Jermann, P., Mullins, D., Nüssli, M.-A., & Dillenbourg, P. (2011). Collaborative Gaze Footprints: Correlates of Interaction Quality. *CSCL*.

Jiang, J., Chen, C., Dai, B., Shi, G., Ding, G., Liu, L., & Lu, C. (2015). Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences of the United States of America*, 112(14), 4274–4279. <https://doi.org/10.1073/pnas.1422930112>

Johnson, S. D., Suriya, C., Won Yoon, S., Berrett, J. V., & La Fleur, J. (2002). Team development and group processes of virtual learning teams. *Computers & Education*, 39(4), 379–393. [https://doi.org/10.1016/S0360-1315\(02\)00074-X](https://doi.org/10.1016/S0360-1315(02)00074-X)

Jongerius, C., Twisk, J. W. R., Romijn, J. A., Callemein, T., Goedemé, T., Smets, E. M. A., & Hillen, M. A. (2021). The Influence of Face Gaze by Physicians on Patient Trust: An Observational Study. *Journal of General Internal Medicine*. <https://doi.org/10.1007/s11606-021-06906-2>

Kajopoulos, J., Cheng, G., Kise, K., Müller, H. J., & Wykowska, A. (2021). Focusing on the face or getting distracted by social signals? The effect of distracting gestures on attentional focus in natural interaction. *Psychological Research*, 85(2), 491–502. <https://doi.org/10.1007/s00426-020-01383-4>

Kempler, T. M., & Linnenbrink, E. A. (2006). Helping behaviors in collaborative groups in math: A descriptive analysis. In S. Karabenick, & R. Newman (Eds.), *Help seeking in academic settings: goals, groups, and context* (pp. 89-115). Mahwah, NJ: Lawrence Erlbaum.

- Kendrick, K. H., & Holler, J. (2017). Gaze direction signals response preference in conversation. *Research on Language and Social Interaction*, 50(1), 12–32.  
<https://doi.org/10.1080/08351813.2017.1262120>
- Keskinarkaus, A., Huttunen, S., Siipo, A., Holappa, J., Laszlo, M., Juuso, I., Väyrynen, E., Heikkilä, J., Lehtihalmes, M., Seppänen, T., & Laukka, S. (2016). MORE – a multimodal observation and analysis system for social interaction research. *Multimedia Tools and Applications*, 75(11), 6321–6345.  
<https://doi.org/10.1007/s11042-015-2574-9>
- Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, 42, 419–427. <https://doi.org/10.1007/BF02344719>
- Kitchenham, B., Brereton, P., Li, Z., Budgen, D., & Burn, A. (2011). Repeatability of systematic literature reviews. *15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011)*, 46–55.  
<https://doi.org/10.1049/ic.2011.0006>
- Knight, A. P., Kennedy, D. M., & McComb, S. A. (2016). Using recurrence analysis to examine group dynamics. *Group Dynamics: Theory, Research, and Practice*, 20(3), 223–241. <https://doi.org/10.1037/gdn0000046>
- Kraus, J., Roman, R., Lacinová, L., Lamoš, M., Brázdil, M., & Fredrikson, M. (2020). Imagery-induced negative affect, social touch and frontal EEG power band activity. *Scandinavian Journal of Psychology*, 61(6), 731–739.  
<https://doi.org/10.1111/sjop.12661>
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394–421.  
<https://doi.org/10.1016/j.biopsycho.2010.03.010>

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, *19*(3), 335–353.

[https://doi.org/10.1016/S0747-5632\(02\)00057-2](https://doi.org/10.1016/S0747-5632(02)00057-2)

Kreijns, K., Kirschner, P. A., Jochems, W., & van Buuren, H. (2004). Determining Sociability, Social Space, and Social Presence in (A)synchronous Collaborative Groups. *CyberPsychology & Behavior*, *7*(2), 155–172.

<https://doi.org/10.1089/109493104323024429>

Kuula, A. (2011). Tutkimusetiikka: aineistojen hankinta, käyttö ja säilytys (2nd ed.). Tampere: Vastapaino

Kwon, K., Hong, R.-Y., & Laffey, J. M. (2013). The educational impact of metacognitive group coordination in computer-supported collaborative learning. *Computers in Human Behavior*, *29*(4), 1271–1281.

<https://doi.org/10.1016/j.chb.2013.01.003>

Kwon, K., Liu, Y.-H., & Johnson, L. P. (2014). Group regulation and social-emotional interactions observed in computer supported collaborative learning: Comparison between good vs. poor collaborators. *Computers & Education*, *78*, 185–200.

<https://doi.org/10.1016/j.compedu.2014.06.004>

Lenoble, Q., & El Haj, M. (2021). “Look at me”—Eye movements during autobiographical retrieval in face-to-face interactions. *Journal of Psychophysiology*, Bartoli, E., Smorti, A. (2018). Facing the language-memory problem in the study of autobiographical memory. *Integrative Psychological Behavioral Sciences*, *53*(3), 374-396. 10.1007/s12124-018-9434-x2018-23566-001, No-Specified. <https://doi.org/10.1027/0269-8803/a000276>

- 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. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLOS Medicine*, 6(7), e1000100. <https://doi.org/10.1371/journal.pmed.1000100>
- Linnenbrink-Garcia, L., & Pekrun, R. (2011). Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology*, 36(1), 1–3. <https://doi.org/10.1016/j.cedpsych.2010.11.004>
- Liu, J., Zhang, R., Geng, B., Zhang, T., Yuan, D., Otani, S., & Li, X. (2019). Interplay between prior knowledge and communication mode on teaching effectiveness: Interpersonal neural synchronization as a neural marker. *NeuroImage*, 193, 93–102. <https://doi.org/10.1016/j.neuroimage.2019.03.004>
- Liu, Y., Wang, T., Wang, K., & Zhang, Y. (2021). Collaborative Learning Quality Classification Through Physiological Synchrony Recorded by Wearable Biosensors. *Frontiers in Psychology*, 12. <https://www.frontiersin.org/article/10.3389/fpsyg.2021.674369>
- Lopes, P. N., Salovey, P., Côté, S., & Beers, M. (2005). Emotion regulation abilities and the quality of social interaction. *Emotion (Washington, D.C.)*, 5(1), 113–118. <https://doi.org/10.1037/1528-3542.5.1.113>
- Ludwig, K. (2015). Is Distributed Cognition Group Level Cognition? *Journal of Social Ontology*, 1(2), 189–224. <https://doi.org/10.1515/jso-2015-0001>
- Malmberg, J., Haataja, E., Seppänen, T., & Järvelä, S. (2019a). Are we together or not? The temporal interplay of monitoring, physiological arousal and physiological synchrony during a collaborative exam. *International Journal of*

*Computer-Supported Collaborative Learning*, 14(4), 467–490.

<https://doi.org/10.1007/s11412-019-09311-4>

Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X., & Siipo, A. (2019b).

Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior*, 96, 235–245. <https://doi.org/10.1016/j.chb.2018.06.030>

Malmberg, J., Järvelä, S., & Järvenoja, H. (2017). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. *Contemporary Educational Psychology*, 49, 160–174.

<https://doi.org/10.1016/j.cedpsych.2017.01.009>

Matthews, G., Middleton, W., Gilmartin, B., & Bullimore, M. A. (1991). Pupillary diameter and cognitive load. *Journal of Psychophysiology*, 5(3), 265–271.

McConnell, M. M., & Eva, K. W. (2012). The Role of Emotion in the Learning and Transfer of Clinical Skills and Knowledge. *Academic Medicine*, 87(10), 1316–1322. <https://doi.org/10.1097/ACM.0b013e3182675af2>

McMillan, D. W. (1996). Sense of community. *Journal of Community Psychology*, 24(4), 315–325. [https://doi.org/10.1002/\(SICI\)1520-6629\(199610\)24:4<315::AID-JCOP2>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1520-6629(199610)24:4<315::AID-JCOP2>3.0.CO;2-T)

McMillan, D. W., & Chavis, D. M. (1986). Sense of community: A definition and theory. *Journal of Community Psychology*, 14(1), 6–23.

[https://doi.org/10.1002/1520-6629\(198601\)14:1<6::AID-JCOP2290140103>3.0.CO;2-I](https://doi.org/10.1002/1520-6629(198601)14:1<6::AID-JCOP2290140103>3.0.CO;2-I)

Meijer, J., Veenman, M. V. J., & van Hout-Wolters, B. H. A. M. (2006). Metacognitive activities in text-studying and problem-solving: Development of a taxonomy.

Educational Research and Evaluation, 12(3), 209–237.

<https://doi.org/10.1080/13803610500479991>

Mitkidis, P., McGraw, J. J., Roepstorff, A., & Wallot, S. (2015). Building trust: Heart rate synchrony and arousal during joint action increased by public goods game. *Physiology & Behavior*, 149, 101–106.

<https://doi.org/10.1016/j.physbeh.2015.05.033>

Molenaar, I., & Chiu, M. M. (2014). Dissecting sequences of regulation and cognition: Statistical discourse analysis of primary school children's collaborative learning. *Metacognition and Learning*, 9(2), 137–160.

<http://dx.doi.org/10.1007/s11409-013-9105-8>

Molinari, G. (2017). From learners' concept maps of their similar or complementary prior knowledge to collaborative concept map: Dual eye-tracking and concept map analyses. *Psychologie Française*, 62(3), 293–311.

<https://doi.org/10.1016/j.psfr.2015.11.001>

Mønster, D., Håkonsson, D. D., Eskildsen, J. K., & Wallot, S. (2016). Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology & Behavior*, 156, 24–34.

<https://doi.org/10.1016/j.physbeh.2016.01.004>

Montague, E., Xu, J., & Chiou, E. (2014). Shared Experiences of Technology and Trust: An Experimental Study of Physiological Compliance Between Active and Passive Users in Technology-Mediated Collaborative Encounters. *IEEE Transactions on Human-Machine Systems*, 44(5), 614–624.

<https://doi.org/10.1109/THMS.2014.2325859>

Müller, R., Helmert, J. R., Pannasch, S., & Velichkovsky, B. M. (2013). Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending



to?: *Quarterly Journal of Experimental Psychology*.

<http://journals.sagepub.com/doi/full/10.1080/17470218.2012.737813>

Mundy, P., & Newell, L. (2007). *Attention, Joint Attention, and Social Cognition*.

<https://journals-sagepub-com.libproxy.tuni.fi/doi/full/10.1111/j.1467-8721.2007.00518.x>

Nelson, N. L., & Mondloch, C. J. (2019). Children's perception of emotions in the context of live interactions: Eye movements and emotion judgements.

*Behavioural Processes*, 164, 193–200.

<https://doi.org/10.1016/j.beproc.2019.05.006>

Noy, L., Levit-Binun, N., & Golland, Y. (2015). Being in the zone: Physiological markers of togetherness in joint improvisation. *Frontiers in Human*

*Neuroscience*, 9. <https://www.frontiersin.org/article/10.3389/fnhum.2015.00187>

Nüssli, M.-A., Jermann, P., Sangin, M., & Dillenbourg, P. (2009). Collaboration and abstract representations: Towards predictive models based on raw speech and eye-tracking data. *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning - Volume 1*, 78–82.

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., Herpertz-Dahlmann, B., Konrad, K., & Schulte-Rüther, M. (2016). Look into my eyes: Investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *NeuroImage*, 130, 248–260.

<https://doi.org/10.1016/j.neuroimage.2016.02.026>

Okon-Singer, H., Hendler, T., Pessoa, L., & Shackman, A. J. (2015). The neurobiology of emotion–cognition interactions: Fundamental questions and strategies for future research. *Frontiers in Human Neuroscience*, 9, 58.

<https://doi.org/10.3389/fnhum.2015.00058>

- Oxman, A. D. (1994). Systematic Reviews: Checklists for review articles. *BMJ*, 309(6955), 648–651. <https://doi.org/10.1136/bmj.309.6955.648>
- Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., & Goodwin, M. S. (2017). Interpersonal Autonomic Physiology: A Systematic Review of the Literature. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 21(2), 99–141. <https://doi.org/10.1177/1088868316628405>
- Pan, Y., Novembre, G., Song, B., Li, X., & Hu, Y. (2018). Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *NeuroImage*, 183, 280–290. <https://doi.org/10.1016/j.neuroimage.2018.08.005>
- Pérez, A., Dumas, G., Karadag, M., & Duñabeitia, J. A. (2019). Differential brain-to-brain entrainment while speaking and listening in native and foreign languages. *Cortex*, 111, 303–315. <https://doi.org/10.1016/j.cortex.2018.11.026>
- Perry, A., Saunders, S. N., Stiso, J., Dewar, C., Lubell, J., Meling, T. R., Solbakk, A.-K., Endestad, T., & Knight, R. T. (2017). Effects of prefrontal cortex damage on emotion understanding: EEG and behavioural evidence. *Brain*, 140(4), 1086–1099. <https://doi.org/10.1093/brain/awx031>
- Peters, B. J., Overall, N. C., & Jamieson, J. P. (2014). Physiological and cognitive consequences of suppressing and expressing emotion in dyadic interactions. *International Journal of Psychophysiology*, 94(1), 100–107. <https://doi.org/10.1016/j.ijpsycho.2014.07.015>
- Pfeiffer, U. J., Schilbach, L., Jording, M., Timmermans, B., Bente, G., & Vogeley, K. (2012). Eyes on the Mind: Investigating the Influence of Gaze Dynamics on the

- Perception of Others in Real-Time Social Interaction. *Frontiers in Psychology*, 3, 537. <https://doi.org/10.3389/fpsyg.2012.00537>
- Pfeiffer-Lessmann, N., Pfeiffer, T., & Wachsmuth, I. (n.d.). *An Operational Model of Joint Attention—Timing of Gaze Patterns in Interactions between Humans and a Virtual Human*. 6.
- Piaget, J. (2001). *The language and thought of the child* (M. Gabain & R. Gabain, Trans.). Routledge.
- Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10), 1175–1191. <https://doi.org/10.1109/34.954607>
- Pijera-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2016). Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 64–73. <https://doi.org/10.1145/2883851.2883897>
- Pijera-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2019). Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members? *Computers in Human Behavior*, 92, 188–197. <https://doi.org/10.1016/j.chb.2018.11.008>
- Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotzinger, J. K. (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. *Research Synthesis Methods*, 10(3), 330–342. <https://doi.org/10.1002/jrsm.1354>

- Potts, S. R., McCuddy, W. T., Jayan, D., & Porcelli, A. J. (2019). To trust, or not to trust? Individual differences in physiological reactivity predict trust under acute stress. *Psychoneuroendocrinology*, *100*, 75–84.  
<https://doi.org/10.1016/j.psyneuen.2018.09.019>
- Prinsen, J., Deschepper, A., Maes, E., & Alaerts, K. (2019). Attachment styles have a modulatory impact on psychophysiological arousal evoked by reciprocated and unreciprocated gaze. *Biological Psychology*, *148*, 107773.  
<https://doi.org/10.1016/j.biopsycho.2019.107773>
- Rahal, R.-M., Fiedler, S., & De Dreu, C. K. W. (2020). Prosocial Preferences Condition Decision Effort and Ingroup Biased Generosity in Intergroup Decision-Making. *Scientific Reports*, *10*(1), 10132. <https://doi.org/10.1038/s41598-020-64592-2>
- Ramirez, R., & Vamvakousis, Z. (2012). Detecting Emotion from EEG Signals Using the Emotive Epoc Device. In F. M. Zanzotto, S. Tsumoto, N. Taatgen, & Y. Yao (Eds.), *Brain Informatics* (pp. 175–184). Springer. [https://doi.org/10.1007/978-3-642-35139-6\\_17](https://doi.org/10.1007/978-3-642-35139-6_17)
- Rauchbauer, B., Nazarian, B., Bourhis, M., Ochs, M., Prévot, L., & Chaminade, T. (2019). Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1771), 20180033.  
<https://doi.org/10.1098/rstb.2018.0033>
- Reimann, P., Markauskaite, L., & Bannert, M. (2014). e-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, *45*(3), 528–540.  
<https://doi.org/10.1111/bjet.12146>

- Richardson, D. C., & Dale, R. (2005). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science*, 29(6), 1045–1060.  
[https://doi.org/10.1207/s15516709cog0000\\_29](https://doi.org/10.1207/s15516709cog0000_29)
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The Art of Conversation Is Coordination. *Psychological Science*, 18(5), 407–413.  
<https://doi.org/10.1111/j.1467-9280.2007.01914.x>
- Richardson, J. C., & Swan, K. (2003). Examining Social Presence in Online Courses in Relation to Students' Perceived Learning and Satisfaction. *Online Learning*, 7(1), Article 1. <https://doi.org/10.24059/olj.v7i1.1864>
- Rogat, T. K., & Linnenbrink-Garcia, L. (2011). Socially Shared Regulation in Collaborative Groups: An Analysis of the Interplay Between Quality of Social Regulation and Group Processes. *Cognition and Instruction*, 29(4), 375–415.  
<https://doi.org/10.1080/07370008.2011.607930>
- Roos, A.-L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2021). Test Anxiety and Physiological Arousal: A Systematic Review and Meta-Analysis. *Educational Psychology Review*, 33. <https://doi.org/10.1007/s10648-020-09543-z>
- Roschelle, J. (1992). Learning by Collaborating: Convergent Conceptual Change. *Journal of the Learning Sciences*, 2(3), 235–276.  
[https://doi.org/10.1207/s15327809jls0203\\_1](https://doi.org/10.1207/s15327809jls0203_1)
- Rösler, L., Göhring, S., Strunz, M., & Gamer, M. (2021). Social anxiety is associated with heart rate but not gaze behavior in a real social interaction. *Journal of Behavior Therapy and Experimental Psychiatry*, 70, 101600.  
<https://doi.org/10.1016/j.jbtep.2020.101600>

- Ruissen, M. I., & de Bruijn, E. R. A. (2015). Is it me or is it you? Behavioral and electrophysiological effects of oxytocin administration on self-other integration during joint task performance. *Special Issue: Neuro-Cognitive Mechanisms of Social Interaction*, 70, 146–154. <https://doi.org/10.1016/j.cortex.2015.04.017>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. <https://doi.org/10.1037/h0077714>
- Saab, N. (2012). Team regulation, regulation of social activities or co-regulation: Different labels for effective regulation of learning in CSCL. *Metacognition and Learning*, 7(1), 1–6. <https://doi.org/10.1007/s11409-011-9085-5>
- Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research*, 13(1), 89–99. [https://doi.org/10.1016/0883-0355\(89\)90018-9](https://doi.org/10.1016/0883-0355(89)90018-9)
- Santamaria, L., Noreika, V., Georgieva, S., Clackson, K., Wass, S., & Leong, V. (2020). Emotional valence modulates the topology of the parent-infant inter-brain network. *NeuroImage*, 207, 116341. <https://doi.org/10.1016/j.neuroimage.2019.116341>
- Sariñana-González, P., Romero-Martínez, Á., & Moya-Albiol, L. (2019). Cooperation Between Strangers in Face-to-Face Dyads Produces More Cardiovascular Activation Than Competition or Working Alone. *Journal of Psychophysiology*, 33(2), 65–75. <https://doi.org/10.1027/0269-8803/a000210>
- Scherer, K., & Moors, A. (2018). The Emotion Process: Event Appraisal and Component Differentiation. *Annual Review of Psychology*, 70. <https://doi.org/10.1146/annurev-psych-122216-011854>
- Schneider, B., Dich, Y., & Radu, I. (2020). Unpacking the relationship between existing and new measures of physiological synchrony and collaborative

learning: A mixed methods study. *International Journal of Computer-Supported Collaborative Learning*, 15(1), 89–113. <https://doi.org/10.1007/s11412-020-09318-2>

Schneider, B., Link to external site, this link will open in a new window, Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning*, 13(3), 241–261. <http://dx.doi.org/10.1007/s11412-018-9281-2>

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, 8(4), 375–397. <http://dx.doi.org/10.1007/s11412-013-9181-4>

Schneider, B., & Pea, R. (2014). Toward collaboration sensing. *International Journal of Computer-Supported Collaborative Learning*, 9(4), 371–395. <http://dx.doi.org/10.1007/s11412-014-9202-y>

Schneider, B., & Pea, R. (2015). Does Seeing One Another's Gaze Affect Group Dialogue? A Computational Approach. *Journal of Learning Analytics*, 2(2), 107–133.

Schutz, P. A., & Pekrun, R. (Eds.). (2007). *Emotion in education* (pp. xiv, 348). Elsevier Academic Press.

Shalom, J. G., Israeli, H., Markovitzky, O., & Lipsitz, J. D. (2015). Social anxiety and physiological arousal during computer mediated vs. Face to face communication. *Computers in Human Behavior*, 44, 202–208. <https://doi.org/10.1016/j.chb.2014.11.056>

- Sharafi, Z., Guéhéneuc, Y.-G., & Soh, Z. (2015). A Systematic Literature Review on the Usage of Eye-tracking in Software Engineering. *Elsevier Journal of Software and Information Technology (IST)*. <https://doi.org/10.1016/j.infsof.2015.06.008>
- Sharot, T., & Phelps, E. A. (2004). How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience*, 4(3), 294–306. <https://doi.org/10.3758/CABN.4.3.294>
- Shvarts, A., & Abrahamson, D. (2019). Dual-eye-tracking Vygotsky: A microgenetic account of a teaching/learning collaboration in an embodied-interaction technological tutorial for mathematics. *Learning, Culture and Social Interaction*, 22, 100316. <https://doi.org/10.1016/j.lcsi.2019.05.003>
- So, H.-J., & Brush, T. A. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education*, 51(1), 318–336. <https://doi.org/10.1016/j.compedu.2007.05.009>
- Sobocinski, M., Järvelä, S., Malmberg, J., Dindar, M., Isosalo, A., & Noponen, K. (2020). How does monitoring set the stage for adaptive regulation or maladaptive behavior in collaborative learning? *Metacognition and Learning*, 15(2), 99–127. <https://doi.org/10.1007/s11409-020-09224-w>
- Solhjo, S., Haigney, M. C., McBee, E., van Merriënboer, J. J. G., Schuwirth, L., Artino, A. R., Battista, A., Ratcliffe, T. A., Lee, H. D., & Durning, S. J. (2019). Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load. *Scientific Reports*, 9(1), 14668. <https://doi.org/10.1038/s41598-019-50280-3>
- Sonnenberg, C., & Bannert, M. (2016). Evaluating the Impact of Instructional Support Using Data Mining and Process Mining: A Micro-Level Analysis of the



Effectiveness of Metacognitive Prompts. *Journal of Educational Data Mining*, 8(2), 51–83.

Soto-Icaza, P., Vargas, L., Aboitiz, F., & Billeke, P. (2019). Beta oscillations precede joint attention and correlate with mentalization in typical development and autism. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 113, 210–228. <https://doi.org/10.1016/j.cortex.2018.12.018>

St. John, A. M., Kao, K., Choksi, M., Liederman, J., Grieve, P. G., & Tarullo, A. R. (2016). Variation in infant EEG power across social and nonsocial contexts. *Journal of Experimental Child Psychology*, 152, 106–122. <https://doi.org/10.1016/j.jecp.2016.04.007>

Stahl, G. (2007). Meaning making in CSCL: Conditions and preconditions for cognitive processes by groups. *Proceedings of the 8th International Conference on Computer Supported Collaborative Learning - CSCL'07*, 652–661. <https://doi.org/10.3115/1599600.1599723>

Stevanovic, M., Henttonen, P., Koskinen, E., Perakyla, A., von-Wendt, T. N., Sihvola, E., Tani, P., Ravaja, N., & Sams, M. (2019). Physiological responses to affiliation during conversation: Comparing neurotypical males and males with Asperger syndrome. *PLoS ONE*, 14(9). <https://doi.org/10.1371/journal.pone.0222084>

Strang, A., Funke, G., Russell, S., Dukes, A., & Middendorf, M. (2013). Physio-Behavioral Coupling in a Cooperative Team Task: Contributors and Relations. *Journal of Experimental Psychology. Human Perception and Performance*, 40. <https://doi.org/10.1037/a0033125>

Su, Y., Li, Y., Hu, H., & Rosé, C. P. (2018). Exploring college English language learners' self and social regulation of learning during wiki-supported

collaborative reading activities. *International Journal of Computer-Supported Collaborative Learning*, 13(1), 35–60. <https://doi.org/10.1007/s11412-018-9269->

[y](#)

Suhaimi, N. S., Mountstephens, J., & Teo, J. (2020). EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities. *Computational Intelligence and Neuroscience*, 2020, e8875426.

<https://doi.org/10.1155/2020/8875426>

Sun, B., Xiao, W., Lin, S., Shao, Y., Li, W., & Zhang, W. (2021). Cooperation with partners of differing social experience: An fNIRS-based hyperscanning study. *Brain and Cognition*, 154, 105803. <https://doi.org/10.1016/j.bandc.2021.105803>

Suri, H. (2020). *Ethical Considerations of Conducting Systematic Reviews in Educational Research* (pp. 41–54). <https://doi.org/10.1007/978-3-658-27602->

[7\\_3](#)

Sutton, S. K., & Davidson, R. J. (1997). Prefrontal Brain Asymmetry: A Biological Substrate of the Behavioral Approach and Inhibition Systems. *Psychological Science*, 8(3), 204–210. <https://doi.org/10.1111/j.1467-9280.1997.tb00413.x>

Szanto, T. (2014). How to share a mind: Reconsidering the group mind thesis. *Phenomenology and the Cognitive Sciences*, 13(1), 99–120.

<https://doi.org/10.1007/s11097-013-9323-1>

Thepsoonthorn, C., Yokozuka, T., Miura, S., Ogawa, K., & Miyake, Y. (2016). Prior Knowledge Facilitates Mutual Gaze Convergence and Head Nodding Synchrony in Face-to-face Communication. *Scientific Reports*, 6(1), 38261.

<https://doi.org/10.1038/srep38261>

Tomasello, M. (1995). Joint attention as social cognition. In *Joint attention: Its origins and role in development* (pp. 103–130). Lawrence Erlbaum Associates, Inc.

- Tomasello, M., & Farrar, M. J. (1986). Joint Attention and Early Language. *Child Development*, 57(6), 1454–1463. <https://doi.org/10.2307/1130423>
- Törmänen, T., Järvenoja, H., & Mänty, K. (2021). Exploring groups' affective states during collaborative learning: What triggers activating affect on a group level? *Educational Technology Research and Development*, 69(5), 2523–2545. <https://doi.org/10.1007/s11423-021-10037-0>
- Tseng, H. W., & Yeh, H.-T. (2013). Team members' perceptions of online teamwork learning experiences and building teamwork trust: A qualitative study. *Computers & Education*, 63, 1–9. <https://doi.org/10.1016/j.compedu.2012.11.013>
- Tsuji, S., Fiévét, A.-C., & Cristia, A. (2021). Toddler word learning from contingent screens with and without human presence. *Infant Behavior and Development*, 63, 101553. <https://doi.org/10.1016/j.infbeh.2021.101553>
- Vanderhasselt, M.-A., De Raedt, R., Nasso, S., Puttevils, L., & Mueller, S. C. (2018). Don't judge me: Psychophysiological evidence of gender differences to social evaluative feedback. *Biological Psychology*, 135, 29–35. <https://doi.org/10.1016/j.biopsycho.2018.02.017>
- Vanderhasselt, M.-A., Remue, J., Ng, K. K., Mueller, S. C., & De Raedt, R. (2015). The regulation of positive and negative social feedback: A psychophysiological study. *Cognitive, Affective, & Behavioral Neuroscience*, 15(3), 553–563. <https://doi.org/10.3758/s13415-015-0345-8>
- Vanutelli, M. E., Gatti, L., Angioletti, L., & Balconi, M. (2017). Affective Synchrony and Autonomic Coupling during Cooperation: A Hyperscanning Study. *BioMed Research International*, 2017, 1–9. <https://doi.org/10.1155/2017/3104564>

- Varlet, M., Nozaradan, S., Nijhuis, P., & Keller, P. E. (2020). Neural tracking and integration of “self” and “other” in improvised interpersonal coordination. *NeuroImage*, 206, 116303. <https://doi.org/10.1016/j.neuroimage.2019.116303>
- Visschers-Pleijers, A. J. S. F., Dolmans, D. H. J. M., de Leng, B. A., Wolfhagen, I. H. A. P., & van der Vleuten, C. P. M. (2006). Analysis of verbal interactions in tutorial groups: A process study. *Medical Education*, 40(2), 129–137. <https://doi.org/10.1111/j.1365-2929.2005.02368.x>
- Volet, S., Vauras, M., Salo, A.-E., & Khosa, D. (2017). Individual contributions in student-led collaborative learning: Insights from two analytical approaches to explain the quality of group outcome. *Learning and Individual Differences*, 53, 79–92. <https://doi.org/10.1016/j.lindif.2016.11.006>
- Volet, S., Vauras, M., & Salonen, P. (2009). Self- and Social Regulation in Learning Contexts: An Integrative Perspective. *Educational Psychologist*, 44(4), 215–226. <https://doi.org/10.1080/00461520903213584>
- Vranjes, J., Bot, H., Feyaerts, K., & Brone, G. (2019). Affiliation in interpreter-mediated therapeutic talk: On the relationship between gaze and head nods. *Interpreting*, 21(2), 220–244. <https://doi.org/10.1075/intp.00028.vra>
- Vygotsky, L. S., & Cole, M. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wagner-Altendorf, T. A., van der Lugt, A. H., Banfield, J. F., Meyer, C., Rohrbach, C., Heldmann, M., & Münte, T. F. (2020). The Electrocortical Signature of Successful and Unsuccessful Deception in a Face-to-Face Social Interaction. *Frontiers in Human Neuroscience*, 14, 277. <https://doi.org/10.3389/fnhum.2020.00277>

- Wallot, S., Mitkidis, P., McGraw †, J. J., & Roepstorff, A. (2016). Beyond Synchrony: Joint Action in a Complex Production Task Reveals Beneficial Effects of Decreased Interpersonal Synchrony. *PLoS One*, 11(12), e0168306.  
<http://dx.doi.org/10.1371/journal.pone.0168306>
- Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In *Handbook of educational psychology* (pp. 841–873). Prentice Hall International.
- Webb, N., Troper, J., & Fall, R. (1995). Constructive Activity and Learning in Collaborative Small Groups. *Journal of Educational Psychology*, 87, 406–423.  
<https://doi.org/10.1037/0022-0663.87.3.406>
- Wei, Y., Wu, Y., & Tudor, J. (2017). A real-time wearable emotion detection headband based on EEG measurement. *Sensors and Actuators A: Physical*, 263, 614–621. <https://doi.org/10.1016/j.sna.2017.07.012>
- Wilder, S. (2014). Effects of parental involvement on academic achievement: A meta-synthesis. *Educational Review*, 66(3), 377–397.  
<https://doi.org/10.1080/00131911.2013.780009>
- Winne, P. H. (2019). Paradigmatic Dimensions of Instrumentation and Analytic Methods in Research on Self-Regulated Learning. *Computers in Human Behavior*, 96, 285–289. <https://doi.org/10.1016/j.chb.2019.03.026>
- Winne, P., Hadwin, A., & Perry, N. (2013). Metacognition and computer-supported collaborative learning. *International Handbook of Collaborative Learning*, 462–479.
- Wise, A. F., & Shaffer, D. W. (2015). Why Theory Matters More than Ever in the Age of Big Data. *Journal of Learning Analytics*, 2(2), 5–13.  
<https://doi.org/10.18608/jla.2015.22.2>

- Xolocotzin Eligio, U., Ainsworth, S. E., & Crook, C. K. (2012). Emotion understanding and performance during computer-supported collaboration. *Computers in Human Behavior*, 28(6), 2046–2054.  
<https://doi.org/10.1016/j.chb.2012.06.001>
- Xu, J., Mitra, S., Van Hoof, C., Yazicioglu, R. F., & Makinwa, K. A. A. (2017). Active Electrodes for Wearable EEG Acquisition: Review and Electronics Design Methodology. *IEEE Reviews in Biomedical Engineering*, 10, 187–198.  
<https://doi.org/10.1109/RBME.2017.2656388>
- Xue, H., Lu, K., & Hao, N. (2018). Cooperation makes two less-creative individuals turn into a highly-creative pair. *NeuroImage*, 172, 527–537.  
<https://doi.org/10.1016/j.neuroimage.2018.02.007>
- Yin, Z., Wang, Y., Liu, L., Zhang, W., & Zhang, J. (2017). Cross-Subject EEG Feature Selection for Emotion Recognition Using Transfer Recursive Feature Elimination. *Frontiers in Neurorobotics*, 11, 19.  
<https://doi.org/10.3389/fnbot.2017.00019>
- Yin, Z., Zhao, M., Wang, Y., Yang, J., & Zhang, J. (2017). Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer Methods and Programs in Biomedicine*, 140, 93–110.  
<https://doi.org/10.1016/j.cmpb.2016.12.005>
- Yu, H., Duan, Y., & Zhou, X. (2017). Guilt in the eyes: Eye movement and physiological evidence for guilt-induced social avoidance. *Journal of Experimental Social Psychology*, 71, 128–137.  
<https://doi.org/10.1016/j.jesp.2017.03.007>
- Zamm, A., Debener, S., Konvalinka, I., Sebanz, N., & Knoblich, G. (2021a). The sound of silence: An EEG study of how musicians time pauses in individual and

joint music performance. *Social Cognitive and Affective Neuroscience*, 16(1–2), 31–42. <https://doi.org/10.1093/scan/nsaa096>

Zamm, A., Palmer, C., Bauer, A.-K. R., Bleichner, M. G., Demos, A. P., & Debener, S. (2021b). Behavioral and neural dynamics of interpersonal synchrony between performing musicians: A wireless EEG hyperscanning study. *Frontiers in Human Neuroscience*, 15(Astolfi, L., Toppi, J., Borghini, G., Vecchiato, G., Isabella, R., De Vico Fallani, F., et al (2011). Study of the functional hyperconnectivity between couples of pilots during flight simulation: an EEG hyperscanning study. <http://dx.doi.org/10.1109/IEMBS.2011.6090654>). <https://doi.org/10.3389/fnhum.2021.717810>

Zawacki-Richter, O., Kerres, M., Bedenlier, S., Bond, M., & Buntins, K. (2020). Systematic Reviews in Educational Research. <https://doi.org/10.1007/978-3-658-27602-7>

Zheng, L., Chen, C., Liu, W., Long, Y., Zhao, H., Bai, X., Zhang, Z., Han, Z., Liu, L., Guo, T., Chen, B., Ding, G., & Lu, C. (2018). Enhancement of teaching outcome through neural prediction of the students' knowledge state. *Human Brain Mapping*, 39(7), 3046–3057. <https://doi.org/10.1002/hbm.24059>

# APPENDICES

*Appendix A: The title and abstract screening tool adapted from an example tool in Polanin et al. (2019).*

## **Citation, Title, and Abstract Screening (for pool stage 1)**

1. Does the **citation** indicate publication on or after 2011?
  - a. Yes: continue screening
  - b. No: stop screening
2. Does the **title or abstract** indicate that this is a peer-reviewed journal article?
  - a. Yes: continue screening
  - b. No: stop screening

## **Decision: Should this article be included in literature pool stage 1?**

- a. **Yes**, all screening questions answered Yes or Unclear
- b. **No**, at least one answers definitely “No”

**Abstract Screening for qualitative studies in all fields (pool stage 2a), quantitative or mixed methods studies in fields outside of education (pool stage 2) and quantitative or mixed methods studies in education (pool stage 3)**

3. Does the **abstract** indicate that physiological responses or states were studied?
  - a. Yes or Unsure/Unclear: continue screening



-Key words: “eye tracking”, “gaze tracking”, “gaze convergence”  
“visual attention”, “joint attention”, “joint visual”, “pupillary diameter”, “face emotion”, “facial emotion”, “facial expression”, biosensor, biosensing, biodata, “physiological response”, “physiological measures”, “physiological markers”, “physiological data”, “physiological linkage”, “physiological state”, “physiological activation”, “interpersonal physiology”, “physiological arousal”, “arousal contagion”, “physiological synchrony”, “physiological coupling”, “physiological compliance” ECG, EDA, GSR, EEG, IBS, INS, HR, HRV, IBI, “heart rate”, “inter-beat interval”, “inter-brain synchrony”, “inter-neural synchrony”, “neural synchronization”, fNIRS, “functional near-infrared spectroscopy”, hyperscanning, “electrodermal activity”, “electroencephalography”, “skin conductance”, “trace gesture”, “gesture detection”, “gesture-based”, “multimodal data”

b. No: stop screening

4. Does the **abstract** indicate that human subjects were studied?

a. Yes or Unsure/Unclear: continue screening

-Key words to *manually check for* exclusion: mice, mouse, rat, monkey, primate, macaque, ape, “model organism”, worm, dog, cat, animal, machine, AI, artificial

b. No: stop screening

5. Does the **abstract** indicate that social interactions were studied?

a. Yes or Unsure/Unclear: continue screening

-Key words: “social interaction”, “social group”, intersubjectivity, “sensorimotor communication”, socialization, socialisation, “social behaviour”, “social behavior”, “social relation”, “social group”, “social communication”, social embodiment, “non-verbal communication”, collaborator, “shared representation”, “group process”, “social process”, “social information” “student-teacher”, “teacher-student”, “student-student”, “group interactions”, “social factor”, co-regulation, face-to-face, “shared attention”, “collaborative learning”, “shared monitoring”, “joint monitoring”, collaboration, “group regulation”, dyad, triad, pair, “interpersonal synchrony”, “group work”, “joint action”, “joint learning”, “joint effort”, CSCL, “computer-supported collaborative learning”, “collaborative process”, “dyadic process”, “group member”, “cognitive interaction”, “emotional interaction”

b. No: stop screening

6. Does the **abstract** indicate that the study uses a quantitative or mixed methods design?

a. Yes or Unsure/Unclear: continue screening

b. No: transfer to pool stage 2a for separate analysis on state of qualitative research in this field

-Key words to check if qualitative only: qualitative, ethnography, “action research”, “social observation”, “focus group”, “case study”, “content analysis”, “discourse analysis”, ethnographic, ethnography, “grounded theory”, narrative, observational, phenomenological, phenomenology, genetic

7. Does the **abstract** NOT indicate that the study was conducted in the field of education?

a. Yes or Unsure/Unclear: continue screening

b. No: transfer to literature pool stage 3

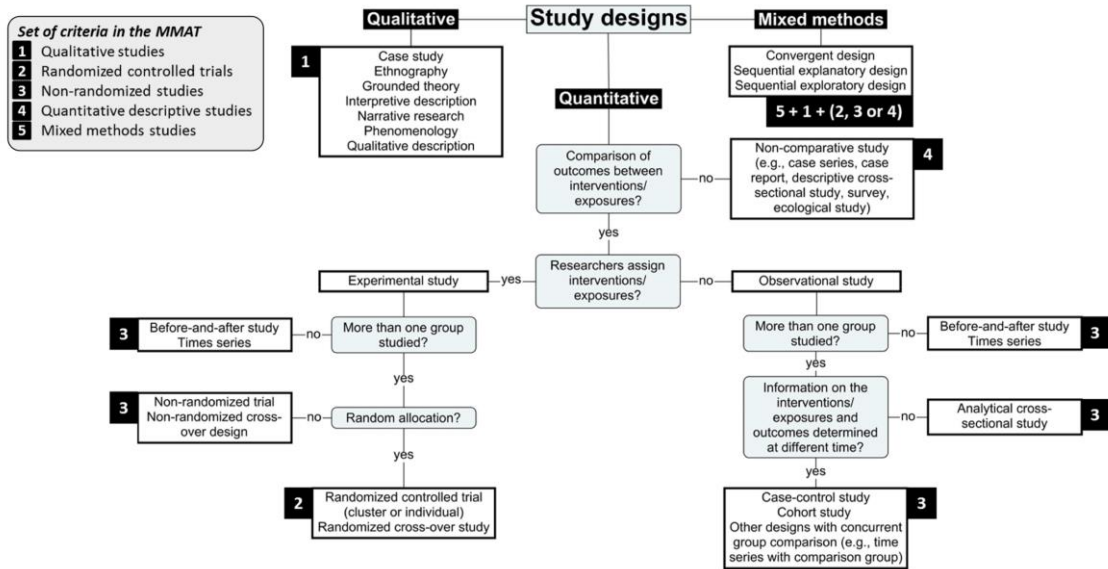
- Key words to check if in education field: education, educate, student, teacher, teach, learn, collaboration, collaborate, collaborative, academic, class, classroom, "knowledge construction", "construction of knowledge", curriculum, pedagogy, pedagogical, self-assessment, self-regulation, study, tutor, lecture

**Decision: Should this article be included in literature pool stage 2?**

a. **Yes**, all screening questions answered Yes or Unclear

b. **No**, at least one answers definitely "No"

*Appendix B: The decision-tree/algorithm from the Mixed Methods Appraisal Tool (MMAT) to categorize studies) from Hong et al. (2018)*



*Appendix C: The unedited Mixed Methods Appraisal Tool (MMAT) from Hong et al. (2018)*

Category of study designs	Methodological quality criteria	Responses			
		Yes	No	Can't tell	Comments
"Screening questions (for all types)" (Hong et al., 2018, p.2)	S1. "Are there clear research questions?" (Hong et al., 2018, p.2)				
	S2. "Do the collected data allow to address the research questions?" (Hong et al., 2018, p.2)				
	<i>"Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions."</i> (Hong et al., 2018, p.2)				
1. Qualitative	1.1. "Is the qualitative approach appropriate to answer the research question?" (Hong et al., 2018, p.2)				
	1.2. "Are the qualitative data collection methods adequate to address the research question?" (Hong et al., 2018, p.2)				
	1.3. "Are the findings adequately derived from the data?" (Hong et al., 2018, p.2)				
	1.4. "Is the interpretation of results sufficiently substantiated by data?" (Hong et al., 2018, p.2)				
	1.5. "Is the coherence between qualitative data sources, collection, analysis, and interpretation?" (Hong et al., 2018, p.2)				
2. "Quantitative randomized controlled trials" (Hong et al., 2018, p.2)	2.1. "Is randomization appropriately performed?" (Hong et al., 2018, p.2)				
	2.2. "Are the groups comparable at baseline?" (Hong et al., 2018, p.2)				
	2.3. "Are there complete outcome data?" (Hong et al., 2018, p.2)				
	2.4. "Are outcome assessors blinded to the intervention provided?" (Hong et al., 2018, p.2)				
	2.5. "Did the participants adhere to the assigned intervention?" (Hong et al., 2018, p.2)				
3. "Quantitative non-randomized" (Hong et al., 2018, p.2)	3.1. "Are the participants representative of the target population?" (Hong et al., 2018, p.2)				
	3.2. "Are measurements appropriate regarding both the outcome and intervention (or exposure)?" (Hong et al., 2018, p.2)				
	3.3. "Are there complete outcome data?" (Hong et al., 2018, p.2)				
	3.4. "Are the confounders accounted for in the design and analysis?" (Hong et al., 2018, p.2)				
	3.5. "During the study period, is the intervention administered (or exposure				

	occurred) as intended?" (Hong et al., 2018, p.2)				
4. Quantitative descriptive	4.1. "Is the sampling strategy relevant to address the research question?" (Hong et al., 2018, p.2)				
	4.2. "Is the sample representative of the target population?" (Hong et al., 2018, p.2)				
	4.3. "Are the measurements appropriate?" (Hong et al., 2018, p.2)				
	4.4. "Is the risk of nonresponse bias low?" (Hong et al., 2018, p.2)				
	4.5. "Is the statistical analysis appropriate to answer the research question?" (Hong et al., 2018, p.2)				
5. Mixed methods	5.1. "Is there an adequate rationale for using a mixed methods design to address the research question?" (Hong et al., 2018, p.2)				
	5.2. "Are the different components of the study effectively integrated to answer the research question?" (Hong et al., 2018, p.2)				
	5.3. "Are the outputs of the integration of qualitative and quantitative components adequately interpreted?" (Hong et al., 2018, p.2)				
	5.4. "Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?" (Hong et al., 2018, p.2)				
	5.5. "Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?" (Hong et al., 2018, p.2)				

*Appendix D: Summaries for studies using physiological and biodata in human social interactions in the education field in the past decade*

Study: Holper et al., 2013

Study design: Quantitative

PRT/biosensor: fNIRS

Biodata: Hbo and Hb concentration ( $\mu\text{mol/l}$ )

Utility: Hbo and Hb levels as prefrontal cortex activity

PMOIs: Hbo and Hb concentration

Other MOIs: Percent agreement with Socratic dialog, transfer ability (yes/no)

Participants: 34 participants, 24 females, mean age: 24 years, 10 participants were not included in analysis due to missing data for one subject

Task: Session: 4 periods performed by dyads, teacher/student role, sitting across each other in a quiet room while following the Socratic dialog, 2 min rest before and after, and a 10 min dual reading of a Meletos dialog for control at last. Teacher roles were trained how to ask the 50 questions to lead the students through the dialog.

Quantitative data analysis: ANOVA with fixed factors 'concentration' (2, Hbo or Hb) and 'condition' (3, phases teaching), with post hoc comparisons; ANOVA run on 'transfer' (yes or no) and 'condition', with a post hoc analysis to test differences between Hbo activity between students who did and did not transfer.

Study: Schneider & Pea, 2013

Study design: Mixed methods

PRT/biosensor: Eye-tracking glasses

Biodata: Eye movement, pupillary dilation

Utility: Gaze direction and movement, pupillary dilation: arousal

PMOIs: Joint visual attention (JVA/JA), cognitive load

Other MOIs: Learning gain, GPA, quality of collaboration, amount of speech production, pupil dilation, fixations, saccades; qualitative: student coordination, convention creation, hypothesis building, and theory sharing

Participants: 42 college students, 28 females, mean age: 23 years

Task: Collaborative explaining of diagram-based contrasting cases with 'visible gaze' or 'no-gaze' deictic conditions for 12 min, then read (same topic) for 12 min.

Quantitative data analysis: ANOVAs run to test the effects of joint attention on learning gains, quality of collaboration, and cognitive load (estimated by pupil dilation). ANOVAs: gaze condition on JVA, individual fixations & saccades on learning outcomes, gaze condition on amount speech production, Pearson correlation between speech production and JVA. Model for potential mediators of student learning tested: collaboration, JVA, cognitive load, GPA as covariate. Qualitative data analysis: Qualitative observation of gazes and utterances: Two random groups' videos at 0.5x speed were analyzed: gaze patterns because of mutual gaze awareness intervention in terms of student coordination ability, convention creation, hypothesis building, and theory sharing. Vignette aim: explanation for the mechanisms of gaze-awareness effects on collaboration; Reliability not calculated.

Study: Montague et al., 2014

Study design: Quantitative

PRT/biosensor: EDA sensors and ECG

Biodata: EDA and HR

Utility: EDA synchrony and IBI synchrony

PMOIs: PC through 5 PCIs

Other MOIs: Task demand and technology reliability conditions, group performance, passive user rating of active user's workload, and shared perception of technology trustworthiness

Participants: 48 participants, 31.3% female, mean age: 21.6 years

Task: A modified Multi-attribute Task Battery (MATB) program; monitoring, tracking, and resource management tasks simultaneously

Quantitative data analysis: PC indicators were signal matching (SM), instantaneous derivative matching (IDM), directional agreement (DA), cross correlation (CC), and weighted coherence (WC). PCIs were compared to each other with correlation coefficients. PCIs were compared with baseline with a linear mixed effects model (LME). The relations between PCIs and other MOIs were tested with LMEs.



Study: Schneider & Pea, 2014

Study design: Mixed methods

PRT/biosensor: Eye-tracking glasses

Biodata: Eye movement

Utility: Gaze direction & movement

PMOIs: JVA

Other MOIs: Quality of collaboration, learning gain, fixations, saccades, 8 social collaboration dimensions (see qualitative)

Participants: 42 participants, 28 females, mean age: 23 years

Task: Collaborative explaining of diagram-based contrasting cases with 'visible gaze' or 'no-gaze' deictic conditions for 12 min, then read (same topic) for 12 min.

Quantitative data analysis: Novel network graphs were constructed per subject where fixations are nodes and saccades are edges; dyad-level network graphs where nodes are screen areas and edges are saccades. ANOVAs computed network metrics (node quantity, node size, edge quantity, reciprocated edge quantity) based on gaze condition. Pearson correlations were run on betweenness centrality, JVA, and matrix metrics and qualitative sub-dimensions of collaboration quality and learning outcomes.

Qualitative data analysis: Social collaboration quality rating: Collaboration quality by rating: sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, reciprocal interaction, and individual task orientation; 20% coding overlap, Krippendorff's alpha: .81

Study: Schneider & Pea, 2015

Study design: Quantitative

PRT/biosensor: same as Schneider & Pea, 2013

Biodata: same as Schneider & Pea, 2013

Utility: same as Schneider & Pea, 2013

PMOIs: same as Schneider & Pea, 2013

Other MOIs: Collaboration quality, learning gains, linguistic features:  
convergence and coherence

Participants: same as Schneider & Pea, 2013

Task: Collaborative explaining of diagram-based contrasting cases with 'visible gaze' or 'no-gaze' deictic conditions for 12 min, then read (same topic) for 12 min.

Quantitative data analysis: Unigram, bigram, and trigram probabilities computed with Pearson correlations between n-grams and measures of interest; categories of common unigrams were built manually: jargon, diagram, location, conceptual discussion, uncertainty, anaphora (person), and anaphora (thing). Convergence (mimicking grammatical structure of interlocutor) for numerous grammatical features compared between visible gaze and no-gaze conditions (ANOVAs). Coherence (word repetition between dyad members): cosine similarity over 5-exchange window; predictor strength of coherence for collaboration quality and learning outcome by cosine similarity matrix. Similarity over time: correlations of 5 vs 5 utterances, 1 exchange sliding window

Study: Ahonen et al., 2016

Study design: Quantitative

PRT/biosensor: ECG

Biodata: HR

Utility: HRV-RMSSD synchrony

PMOIs: Social physiological compliance (SPC)

Other MOIs: Collaborative task type and task load

Participants: 38 participants, 16 females, 28 18-23 years old, 9 24-29 years old, and one over 34

Task: Participants seated and watch a baseline video together, then pair-programming collaborative task, dyads switch roles every 7 min, 90 min total, 2 separate, counterbalanced assignments

Quantitative data analysis: Pearson's product-moment correlation coefficients of the feature vectors within dyads for HR, SDNN, or rMSSD (per minute), then the average correlation of the dyads is an estimate of SPC. P-values were calculated through comparison to a distribution of 10000 random correlations. A

linear regression model was fit to study the association between SPC and self-report items.

Study: Gillies et al., 2016

Study design: Mixed methods

PRT/biosensor: EDA sensors

Biodata: EDA

Utility: EDA synchrony

PMOIs: Synchronous EDA interpreted as synchronous engagement

Other MOIs: EDA data in network analysis: degree, participation index, assortativity, and clustering coefficient. Student Attitudes and Beliefs Questionnaire (SABQ) scores, Children's self-efficacy scale scores, and MSLQ scores. For students: the PAT in Science scores (post experimentation). For qualitative measures, see qualitative data analysis section.

Participants: 20 year six students 10-12 years old (other data not provided)

Task: 1-hr lesson, Year 6 teacher and 20 students; students diagnosed case-study diseases with justifications and reasons for exclusions; students could access multimedia resources on the topic; students collaborated to make visual, auditory, and kinesthetic representations of learning progress

Quantitative data analysis: A between student correlation matrix was computed using whole-class and cooperative small group time-bin averaged EDA amplitude values, Pearson correlations for every possible pair across both conditions to create connectivity networks for each condition, network analysis computed facets of connectivity between students

Qualitative data analysis: Student behavior states: on-task behavior (task-oriented group, cooperative behavior); off task (noncompliance with the group); and independent behavior (on-task but working independently), Student behavior states were coded according to a scheme developed earlier at the individual level over 10-s non-overlapping intervals, then scores were aggregated to the group and reported as percentages of total group behavior.

Types of student language coded: social language, basic statement, basic, moderate, and advanced use of scientific language; Teacher measures: coding of multimodal representation use in class and teacher language: basic statement, asks an open question, asks a closed question, mediates student's

learning, and encourages student's ongoing engagement. Inter-rater reliability ensured >85%.

Study: Thepsoonthorn et al., 2016

Study design: Quantitative

PRT/biosensor: Glasses camera, web camera, and accelerometer combined

Biodata: Eye movement

Utility: Tracking of dyadic non-verbal interactional behaviors

PMOIs: Mutual gaze convergence and head nodding synchrony

Other MOIs: Prior knowledge

Participants: 30 participants, 6 females, age range: 21-47 years old.

Task: Face-to-face interaction of "lecturer and student" for 5 min, divided into two equal parts: part 1: content reviewed for prior knowledge activated participants; part 2: new information lecture, low relation with part 1

Quantitative data analysis: Mutual gaze convergence detected at 1 Hz if had 'straight gaze at each other'; Nodding synchrony determined in 1.8 s windows.

T-tests observed differences in average percentage of mutual gaze convergence between students with and without prior knowledge, in part 1 and 2; same for head nodding synchrony; Differences in measures of interest also assessed with t-tests when part 1 and 2 were combined.

Study: Dikker et al., 2017

Study design: Quantitative

PRT/biosensor: EEG system

Biodata: Electrical brain activity

Utility: Synchrony

PMOIs: Total interdependence of group, student-to-group, and student-to-student

Other MOIs: student (day-by-day and post-semester) ratings of 4 different teaching styles: reading aloud, video, lecture, and discussion; other individual variables: focus, group affinity, empathetic disposition, student-teacher closeness rating, and student-student closeness rating; conditions: FTF baseline, eye-contact with peer for 2 min + adjacent with peer, no FTF +

adjacent, and non-adjacent

Participants: 12 high school students, 9 females, age 17-18

Task: 11 recording days (50 min classes) over 3 months, EEG recorded during video, lecture, reading, and discussion teaching styles on all days. Randomized 3 types of 2 min baseline activities at beginning and end of each class: 2 min face wall, group, or peer (sit still, no talking, focus)

Quantitative data analysis: Total interdependence was defined as the spectral coherence of two signals and was calculated based on the Welch method.

Repeated-measures two-way ANOVAs followed by post-hoc tests were used to test variances of student ratings or brain synchrony across teaching styles.

Multilevel models were created to find the relationship between student-to-group total interdependence and questionnaire metrics. Namely, main effect of stimulus and ratings: group vs. individual and day-by-day vs. semester (2-way repeated measures ANOVAs); independent effects of stimulus vs. individual differences on brain synchrony: 2 state variables and 2 trait variables (repeated measures multilevel regression); effects of co-presence: student closeness and teacher likeability (1-way ANOVA and correlations).

Study: Molinari, 2017

Study design: Quantitative

PRT/biosensor: Computers equipped with remote eye-trackers

Biodata: Eye movements

Utility: Gaze behavior

PMOIs: Gaze fixations

Other MOIs: Number of elements in concept maps (CMs), number of shared elements between both CMs, number of elements in shared CM, number of new elements in shared CM, number of matching elements transferred to shared CM

Participants: 60 participants, 11 females, mean age: 20.5 years; divided randomly into 30 dyads. Data was analyzed for 28 participants due to technical difficulties

Task: Dyad in separate rooms on computers with eye-trackers for 80 min with 5 phases: prior knowledge assessment, individual learning (reading), individual concept mapping, collaborative concept mapping, outcome assessment

(individual learning performance and knowledge modeling accuracy), shared knowledge (SK): same texts; different knowledge (DK): complementary texts; Computer-based CM divided screen into CM, partner map, and own map; communication was done via audio

Quantitative data analysis: Analysis on 7 dyads within and between knowledge conditions (shared/different): Correlation of eye movement and CM measures with individual learning performance. Individual-level analysis: ratio of fixation time on CM, shifts of fixation between CMs, and individual CM measures; group-level analysis: shared CM measures; Descriptive statistics on transitions between AOIs across knowledge conditions and divided into first and second half of session

Study: Ahonen et al., 2018

Study design: Quantitative

PRT/biosensor: EDA sensors and ECG

Biodata: EDA and HRV-SDNN

Utility: EDA synchrony and HRV-SDNN synchrony

PMOIs: SPC

Other MOIs: Valence and engagement

Participants: 38 participants in final dataset, of which 18 females, mean age: 23 years

Task: Pair-programming task design: dyad members took multiple turns (self-paced) between driver and navigator for multiple, linked assignments at dyad's own pace, which contain "test" and "run" events (which evaluate code)

Quantitative data analysis: Average correlations of SCR and SCL between dyads compared to randomly simulated dyads. Minimum width envelope (MWE) method to visualize SCR over time across conditions (feedback and role) for grand averaged data, applied to event-based analysis (for valence and engagement); Computed average correlation of SDNN (and mean HR) in 60 and 300-s windows between collaborating individuals and compared to all pairwise SDNN correlations (with and without task-change periods).

Study: Bevilacqua et al., 2019

Study design: Quantitative

PRT/biosensor: EEG system

Biodata: Electrical brain activity

Utility: Synchrony

PMOIs: Total interdependence

Other MOIs: Teaching style (lecture or instructional video); student retention; pre and poststudy teacher closeness and content likeability; pre- and post-session likability of lesson and experiment

Participants: 12 high school students, 7 females, aged 16-18 years

Task: Six classroom sessions of 20 min per lesson of interleaved lecture and video teaching conditions (5 min each), students asked not to make sudden movements or talk

Quantitative data analysis: Total interdependence calculated same as Dikker et al. (2017) for every 1-s epoch for all student-student and student-teacher pairs. Student-group total interdependence was calculated by averaging all other students' total interdependence values compared to the remaining student's total interdependence values. Multilevel models with days nested within students were used to address RQs, namely repeated-measures multilevel regression analyses: independent variable x factor on dependent variable (e.g., teaching style X quiz scores on student-group TI)

Study: Haataja et al., 2018

Study design: Mixed methods

PRT/biosensor: EDA sensors

Biodata: EDA

Utility: EDA synchrony

PMOIs: Physiological concordance (PC) as an index of Physiological synchrony (PS)

Other MOIs: Metacognitive events, see qualitative

Participants: 48 participants, 27 females, mean age: 17.4

Task: 16 groups of 3 Finnish high school students collaborated to design a healthy breakfast with necessary information and a 5-phase script on the page (computer): task instruction, planning, knowledge acquisition, evaluate and discuss, and check your answer

Quantitative data analysis: Physiological concordance: temporal variation of students' average SC slopes within a moving 5-s window, Pearson correlations over 15-s windows, significance found by Monte Carlo shuffling. EDA and video data: 120-s moving window analyzed temporal changes in monitoring. Group monitoring vs group PC: detrending moving-average cross-correlation (DMCA) coefficient

Qualitative data analysis: Coding of monitoring instances based on utterances: Monitoring of cognition, behavior, emotion, and motivation (latter two combined due to low frequency) (Haataja et al., 2018); Cohen's kappa = .76

Study: Malmberg et al., 2019a

Study design: Mixed methods

PRT/biosensor: EDA sensors and Facial expression recognition software (used on video)

Biodata: EDA and LBP-TOP features

Utility: Arousal, synchrony, and Valence expressions

PMOIs: Arousal, PS, and Positive, negative, or neutral

Other MOIs: Metacognitive events, see qualitative

Participants: 48 participants, 27 females, mean age: 17.4 years

Task: 6 mixed-gender groups of 3 or 4 over 7 weeks collaboratively composed a midterm plan 3-5x; complex and open-ended task

Quantitative data analysis: NSSCRs analyzed on 1-min segments (not time-locked), descriptive analysis; Extraction of LBP-TOP features from video, valence model estimates valence after training, achieved 96.26% recognition rate on Cohn-Kanade database, frequency of valence expressions/1-min window, descriptive statistics

Qualitative data analysis: Qualitative content analysis: Only in high-arousal episodes of  $\geq 2$  students: identified work phases and interaction type for quality of collaboration: high-level interaction, low-level interaction, and confusion; Cohen's kappa = 0.65.

Study: Pan et al., 2018

Study design: Quantitative



PRT/biosensor: fNIRS

Biodata: Hbo and Hb

Utility: Synchrony

PMOIs: (interpersonal brain synchronization) IBS dynamics

Other MOIs: Part-learning and Whole-learning states

Participants: 24 female participants (mean age: 20.58) and 1 female music instructor (22 years old)

Task: Two Chinese songs were selected that have simple melodies (and general musical structure) and lyrics, convey similar emotions, and were unfamiliar; each song: 4 musical phrases, 6 s per phrase; 6 dyads per song; task: 3 phases: rest (3 min sitting FTF with eyes closed), learning (~9 min) and solo (2 min); Learning phase: instructor sang song 2x (~1 min), then for 8 min, PL: sang and imitated repeatedly phrase by phrase, WL: sang and imitated repeatedly whole song

Quantitative data analysis: IBS (averaged across all channels in each dyad) during task estimated by WTC, and compared to baseline with paired sample *t*-tests for each frequency band; from frequencies of interest (FOIs), IBS averaged, and *t*-tests compared vs zero and across groups; learning modes were compared with segments of IBS with independent sample *t*-tests; coupling directionality was computed with granger causality analysis.

Study: Schneider et al., 2018

Study design: Mixed methods

PRT/biosensor: Eye-tracking glasses

Biodata: Eye movements

Utility: Gaze direction and movements

PMOIs: JVA

Other MOIs: Learning gain, (analysis and construction) task performance, quality of collaboration

Participants: 54 participants, 7 females, mean age: 18 years

Task: Construction task with 2 subtasks: optimize warehouse space, minimize the avg. distance between shelves and docks

Quantitative data analysis: Correlations were run between joint attention and learning gains and performance during tasks, also divided by year of student

(1st, 2nd, or 3rd year). Correlations were run between quality of collaboration, its constitutive components, and percentage of joint attention. On 2 dyads that both showed high JVA, high task performance, but high and low learning gains: augmented (with spatial and speech duration information) cross-recurrence graphs visualized the synchronization of JVA. A measure of imbalance of 'visual leadership' was computed and plotted against learning gains per task type. Qualitative data analysis: For 2 dyads (see quantitative): multimodal analysis of students' interactions: gestures and speech: Gestures and proposal reactions through exploratory discovery using a simultaneous viewing of video, gaze indicator, and cross-recurrence graph. Collaboration quality by rating: sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, reciprocal interaction, and individual task orientation; 20% coding overlap, Krippendorff's alpha: .83

Study: Zheng et al., 2018

Study design: Quantitative

PRT/biosensor: fNIRS

Biodata: Hbo and Hb

Utility: Synchrony

PMOIs: INS as related to teaching style and outcome

Other MOIs: Numerical reasoning pre- and post-test scores

Participants: 4 participants, 2 females, mean age: 25 years, served as teachers and 60 participants, 30 females, mean age: 23 years, were students

Task: Teachers taught numerical reasoning to students (dyads); 8 training examples selected from CCSAPAT (Chinese numerical reasoning exam) training section; teachers flexibly taught from scripts. 4 teachers taught same content to 3 students in 3 styles. 10 min rest phase (still, relaxed, closed eyes); lecture: teacher explained each example (no questions allowed); interactive: teacher presented, student thought for ~20 s, then teacher guides student (Q&A format); video: student watched lecture recording; lecturing and interactive: teacher and student next to each other in silent room; video style: student alone in front of computer; both teaching periods were flexible, 13-26 min.

Quantitative data analysis: WTC conducted for all channel combinations between dyad members, averaged across the whole session for all 3 teaching

modes, and INS increase calculated. One-way ANCOVAs for all channel combinations for all frequencies to compare INS across teaching styles. One-way ANCOVA run for INS increase from teaching style and outcome across time. One-way ANCOVA conducted between INS increases with 2-14 s time-lags before and after (step = 2 s)

Study: Dindar et al., 2019

Study design: Mixed methods

PRT/biosensor: EDA sensors

Biodata: EDA

Utility: Synchrony

PMOIs: PS

Other MOIs: Metacognitive events, see qualitative

Participants: two male and one female participant in a triad, aged 15 to 16 years old

Task: 12 groups of 3 or 4 students in a CL session: listened to short lecture, then wrote a group essay using computers and internet; essay included experimental design and integration of historical research-based knowledge.

The second collaborative task dealt with experimentation and report writing.

Quantitative data analysis: Multidimensional Recurrence Quantification Analysis (MdrQA) was applied to quantitatively describe EDA signals for the triad members in recurrence plots. MdrQA was calculated for 1' windows. Every minute was visualized in recurrence plots. Correlations were calculated between monitoring durations and MdrQA indices. T-tests were conducted to reveal potential differences in PS between CL conditions with and without monitoring.

Qualitative data analysis: Coding of monitoring instances based on utterances: Monitoring cognition, behavior, emotion, and motivation of the group; 20% coding overlap, Cohen's kappa = .73

Study: Liu et al., 2019

Study design: Quantitative

PRT/biosensor: fNIRS

Biodata: Hbo and Hb

Utility: Synchrony

PMOIs: Prefrontal cortex synchrony related to prior knowledge and teaching mode

Other MOIs: perceived teacher-student interaction (via post-session questionnaire), familiarity with teaching materials/prior knowledge (via post-session questionnaire), students' post-test scores (learning outcome)

Participants: 84 participants, 64 female, mean age: 21.0 years

Task: mixed 2 x 2 experiment: communication mode (FTF/CMC) x prior knowledge state (with/without); teacher remotely controlled computers for presentation during lectures; FTF: teacher used expressions and gestures, students could nod or mutual gaze; CMC: sat back-to-back, no non-verbal cues, only 2 synced computers; teachers were trained and assessed beforehand

Quantitative data analysis: INS assessed with WTC analysis of Hbo signal.

One-sample *t*-tests for all channels' task-related INS. A *t*-map of INS was generated, then a mixed 2 x 2 ANOVA was run on channels with significant INS for each condition

Study: Malmberg et al., 2019b

Study design: Mixed methods

PRT/biosensor: EDA sensors

Biodata: EDA

Utility: EDA arousal and synchrony

PMOIs: Arousal (activation) and PS

Other MOIs: Metacognitive events, see qualitative

Participants: 31 high school students, 8 females, 15-16 years old

Task: 4 heterogeneous groups of 3 or 4 based on previous grades took a collaborative exam to design and report a physics experiment, average completion time: roughly 29 min, all groups successful

Quantitative data analysis: Moving 5-s windows of average EDA slope, Pearson correlations (PCs) on 15-s windows, single session index (SSI) computed;

Monte Carlo shuffling for significance of PCs; temporal variation with 120-s windows in 1-s steps

Qualitative data analysis: Qualitative content analysis: Utterances coded into monitoring of behavior, cognition, and motivation and emotions per subject.

“Behavior: monitoring task-related behavior, such as the resources needed for the task, monitoring task progression; cognition: monitoring task understanding and prior knowledge. Monitoring procedural knowledge and whether the study product is correct/in the normal range. Monitoring content understanding; motivation and emotions: monitoring current trends in motivation, monitoring volition and efficacy, monitoring emotional state” (Malmberg et al., 2019b). Student reactions to monitoring were coded: silent nodding, agreeing, or reacting visibly Cohen's kappa = .74

Study: Pijeira-Díaz et al., 2019

Study design: Quantitative

PRT/biosensor: EDA sensors

Biodata: EDA

Utility: EDA synchrony

PMOIs: Within-triad arousal

Other MOIs: PCIs: within-triad arousal level, directional agreement, and contagion during collaboration

Participants: 24 high school students 16-17 years old, 25% female

Task: In an elective advanced physics course, 4 triads of high-performing high school students (evenly distributed according to MSLQ categories) listened to teacher lectures and collaborated with hands-on experiments over 18 lessons (75 min/lesson)

Quantitative data analysis: EDA signal analyzed with 1-min moving window with 250-ms moving steps. SCR frequencies 1-3ppm (peaks/min) were categorized as low arousal, 20ppm and above high, and between values medium.

Descriptive statistical analysis for directional agreement: compared slopes of the trend lines in each 1s window; level agreement: 1min windows. High arousal intervals were analyzed to determine if other members of the group also had concurrent arousal, and with what latency to explore contagion. Other group members were also assessed as potential influencers to each high arousal event.

Study: Shvarts & Abrahamson, 2019

Study design: Qualitative

PRT/biosensor: Eye-tracking glasses

Biodata: Eye movement

Utility: Gaze direction and movement

PMOIs: Scanpaths

Other MOIs: See qualitative data analysis

Participants: 4 17–21-year-old participants in 2 pairs (other demographic data not provided)

Task: Interactive learning activity on computer screen, as student manipulates the vertex of a blue triangle to make it green, a tutor sitting next to the student helped without explicit solution giving

Qualitative data analysis: Micro-ethnography: searching for patterns across student actions (e.g., hand gestures), student/tutor gaze parameters, and dyad's multimodal utterance; coordination and discoordination between student and tutor attention and student action and tutor attention; joint attention coded through non-verbal and verbal context (i.e., to establish gaze overlap was not by chance); reliability not calculated.

Study: Schneider et al., 2020

Study design: Mixed methods

PRT/biosensor: EDA sensors

Biodata: EDA

Utility: EDA synchrony

PMOIs: Physiological synchrony (through 4 physiological coupling indicators (PCIs))

Other MOIs: task performance, learning gains, and collaboration quality

Participants: 84 participants, 60% female, 62% students, 19-51 years old

Task: 30 min sessions; participants in dyads learned pre-recording and then during recording, code robot to solve increasingly difficult mazes. 2 x 2 design: two interventions were used: visualization (proportion of verbal utterances of dyad members over the past 30 s) and explanation (on the benefits of collaboration)

Quantitative data analysis: PCIs: PC, DA, SM, and IDM. To test the relations between physiological data and dependent measures, correlations were run on learning gains, collaboration quality, and task performance. Correlations were

also run between the PCIs to test if they could potentially capture different aspects of physiological synchrony. 2-s moving windows for 2 groups' PS data to smooth noise and visualize for further qualitative analysis. Pearson correlation as an index for PS aggregated in 30-s time windows, then computed cycles of high/low synchrony and ran correlations between these and all measures of interest.

Qualitative data analysis: Learning tests pre, post, and gains scores: qualitative evaluation of free-response answers; Collaboration quality: nine scales from Meier et al. (2007); Student code quality: 4 dimensions rated on custom rubric 0-4 by two researchers; Highest group for PC score, learning gains, and collab quality compared to lowest group, video analysis: qualitative analysis of utterances: Collaboration quality by rating 9 dimensions: sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation; Student code quality: sensor thresholds, conditional statements, looping, nesting, and generalizability; Video analysis: events and behavior during peaks, valleys, and oscillations; Learning test and gains codings: 20% overlap, inter-rater reliability: .89; Collaboration quality: 20% overlap, Cohen's kappa: .65; Student code quality: discussion, 100% agreement; video coding: not calculated

Study: Sobocinski et al, 2020

Study design: Mixed methods

PRT/biosensor: ECG

Biodata: HR

Utility: Synchrony

PMOIs: Aggregated HR state (physiological state transitions)

Other MOIs: Metacognitive events, see qualitative data analysis

Participants: 31 high school students, 8 females, 15-16 years old

Task: students in an advanced physics course in groups of 3 or 4 (mixed evenly based on MSLQ categories) collaborated on an exam (a report: experimental setup, formulas, calculations, final result, evaluation of result), all groups scored 5.0-5.5/6.0

Quantitative data analysis: Descriptive statistics on group-level frequency and distribution of physiological state transitions related to qualitative results.

Spearman's rank order correlation was run between physiological state transitions and types of sequences (on-track, adaptive, maladaptive).

Qualitative data analysis: Coding of monitoring instances based on utterances: Monitoring coded into target (cognition, behavior, and motivation and emotion), valence (positive, negative, no valence), and phase (task understanding, goals and planning, task enactment, reflection, and adaptation), and reaction (present or not); 20% comparison, Cohen's kappa = .74, .69, .68 and .67

Study: Haataja et al., 2021

Study design: Mixed methods

PRT/biosensor: EDA sensors

Biodata: EDA

Utility: EDA synchrony

PMOIs: PS

Other MOIs: Monitoring valence, see qualitative data analysis

Participants: 57 university students, 32 females, mean age: 27.29 years

Task: 19 groups of 3 randomly assigned university students participated in a collaborative business simulation to run a company, with the goal to maximally increase the value of the company, dependent on 24 adjustable variables over 6 periods, with a 1-minute transition between each where the new value of the company was revealed, and problem-solving collaboration begun

Quantitative data analysis: MdRQA quantified the synchrony between two signals, %DET, %REC, and ADL were compared, GEEs modeled arousal and synchrony as predictors for valence, equality of participation, and task performance

Qualitative data analysis: Coding of monitoring instances based on utterances: Monitoring valence (positive, negative, and neutral); Cohen's kappa = 0.64

Study: Törmänen et al., 2021

Study design: Mixed methods

PRT/biosensor: EDA sensors

Biodata: EDA



Utility: EDA synchrony

PMOIs: Group activation level (activating or deactivating)

Other MOIs: Group affective state, task- and socially-related factors (see qualitative data analysis)

Participants: 41 6th grade primary school students, 23 female, 12-13 years old

Task: Collaborative (3-4 students/group) task to design a heat-efficient house, where each member had different useful knowledge provided

Quantitative data analysis: 30-s segments of student EDA data were examined for number of NSSCR peaks, 0-1 peaks/segment low, 2-9 medium,  $\geq 10$  high arousal. Segments of  $\geq 2$  students medium/high were 'activating', less were 'deactivating'; Chi-square tests of independence for exploring affective state and physiological activation relations; Chi-square test of independence observed relations between valence of interactions and task- and socially-related factors.

Qualitative data analysis: Qualitative content analysis: Affective state: 30-s segments of video were coded as: positive, negative, mixed, or neutral based on group emotional expressions based on clear verbal or bodily interactions.

Activating affect: segments of activation and preceding ones were coded; triggers (emotional verbal expressions) coded as 'task-' or 'socially-related';

Affective state: 40% overlap coded, Cohen's kappa: .72; activating affect: 30% overlap, Cohen's kappa: .87

*Appendix E: Summaries for studies using physiological and biodata in human social interactions outside of education 2011-2021*

<b>Study</b>	<b>Biodata</b>	<b>Data analysis</b>	<b>Conclusions</b>
Ahn et al., 2017	EEG and MEG	Between-dyad phase synchronizations were calculated and averaged across trials. Weighted phase lag index (WPLI) was computed at each bin and averaged for spectral bands: theta, alpha, beta, and gamma. Surrogate data was constructed and analyzed to determine statistical significance for functional connectivity. Mean WPLI over subjects was computed and plotted as a topography that describes functional connections between individuals, from 0 to 1. MEG data, after cleaning, was analyzed the same as the EEG data.	Statistically significant oscillations in functional connectivity in terms of phase synchronization within-dyad compared to control. MEG data showed alpha and gamma oscillations and phase synchronization in verbal turn-by-turn interactions, EEG showed alpha band synchronization only.

Anaya et al., 2021	EEG and eye-tracking	Multilevel models assessed synchrony through separate analysis of peer gaze/energy level (qualitatively coded) and neural measures (Delta-Beta coupling and frontal EEG Alpha asymmetry). Intradynamic synchrony during unstructured and structured interactions was computed through three-level multilevel models. Follow-up models were used to probe for differences in intradyadic synchrony between structured task types, then 200-ressampling bootstrapping was applied.	Energy levels and peer gaze were elevated in the unstructured task compared to the cooperation tasks. No gaze differences were found between BI and control groups of children. Significant positive dyadic synchrony for peer gaze and energy was found in both task types; peer gaze synchrony was relatively greater in the structured task. Left frontal EEG Alpha asymmetry (not right) was correlated to reduced behavioral synchrony during the unstructured task. "Peer gaze was asynchronous when BI children exhibited negative Delta-Beta coupling and their partner exhibited positive coupling" (Anaya et al., 2021).
Apanovich et al., 2018	EEG	EEG epochs were averaged individually for each condition and were time-locked to the stimuli presentation. Each ERP was based on a max of 30 trials. Holistic and analytic groups were analyzed across three conditions; pairwise comparisons were run with the Wilcoxon test.	Variability of EEG and EOG measures was higher in holistic subjects. P300 amplitude was highest in the individual condition and lower in the group conditions. P300 latency and amplitude suggests decision making was easier for holistic subjects in cooperation, and easier for analytic subjects in competition.

Arslan Aydin  
et al., 2021

Eye-  
tracking

Quantitative: Frequencies of gaze aversion and contact (determined by fixation data, not raw gaze) were compared between roles with paired-sample t-tests. Penalized quasi-likelihood (PQL) models were fit with varying interactions of role, gender, partner gender and random effects, and models were compared with ANOVAs."A comparative analysis was conducted by Convolutional Neural Network (CNN) models that employed specific architectures, namely, VGGNet and ResNet" (Arslan Aydin et al., 2021). Qualitative: Segmentation and annotation were conducted speech recording extracts (including sub-words and pauses). Dialogue act analysis was carried out using text to speech transcription and then coding based on speaker's intention.

Frequency and duration of gaze direction varies significantly based on interview role. ResNet predicted gaze direction with 70% accuracy. Analysis methods using raw gaze data did not detect the differences that fixation data could.

Astolfi et al.,  
2020

EEG

Trial duration, max ball height, and the score of correct trials for each trial were dependent variables in a one-way ANOVA, where task type (Joint action, PC, and Solo) was a within-groups factor. Graph theory analysis was conducted on connectivity networks. One-way repeated measures ANOVAs were conducted with the main factor as experimental condition, separately for each frequency band and for indices extracted from single- and multi-brain networks. It was tested if multiple-brain indices could characterize a task: a classification study used graph indices as features of a classifier to distinguish between Joint-Solo, Joint-PC, and PC-Solo. Correlation of inter-brain indices and average behavioral values (i.e., avg duration of keeping ball on bar, score of correct trials, and max ball height) were run across the four frequency bands.

Group analysis revealed inter-brain links were stronger in the joint condition, an indication of shared representations. Modulation of inter-brain data occurred with level of cooperation or successful interaction between participants. Multiple- and single-subject analysis could both uncover the difference of the joint condition from the other conditions, but multiple-subject analysis resulted in higher classification accuracies. Inter-brain networks primarily involved lower frequencies (theta and alpha).

Azriel et al., 2020	Eye-tracking	To evaluate gaze pattern differences between groups, separate mixed-model ANOVAs were performed for each eye-tracking measure, with Anxiety as a between-subjects factor and AOI and task as within-subjects factors (Azriel et al., 2020). After finding significant three-way interaction, group x task ANOVAs were run for each AOI. Main and two-way interaction effects, if significant, were followed up with Bonferroni corrected post-hoc comparisons. Power analysis was conducted.	Socially anxious individuals (SA) experienced anxiety during VMC interview and presentation tasks, but non-anxious individuals did not. SA did not dwell longer on self, which might be due to context, i.e., lacking evaluative feedback in this study. SA dwelled longer on the confederate during the interview relative to the presentation and longer on non-face during the presentation compared to the interview. Number of fixations per AOI revealed attentional maintenance differences. Non-SA were not sensitive to the contexts.
Balconi & Fronda, 2020	EEG	One-way ANOVAs were run on individual level to test for effects of independent measures on each frequency band. 2. Frequency band IBS was analyzed by use of the partial correlation coefficient. 3. another ANOVA to observe IBS as a function of experimental conditions. For inter-brain analyses, correlational indices were calculated to compute synchronization, which were then used as dependent variables in mixed-model ANOVAs, with the same repeated factors as the single-brain analyses (Balconi & Fronda, 2020).	Gesture type and valence modulate cortical activity according to the encoder/decoder role and French/Italian culture of belonging. Single- and inter-brain analyses revealed type and valence of gestures have individual differences in modulation and coupling according to culture. Gestures with higher representation for the culture increased intra-brain connections and implicit coupling and synchronization with interactors.

Balconi et al., 2017 fNIRS

The mean concentrations in the time series were used to calculate the effect size per condition per channel within a subject. The 8 channels' effect sizes were averaged and compared across subjects and channels. Effects of reaction times and error rates between low- and high-BAS subjects were revealed with ANOVAs. ANCOVAs probed at interaction effects between frequency bands and O2Hb modulation separately as functions of BAS. Pearson correlation coefficients were calculated to find relations among error rates, reaction times, self-perception, band, and EEG modulation.

Select trait components of BAS and a cooperative condition were positively associated with positive self-representation and superior task performance/improved cognitive outcomes, along with significantly increased left PFC activity. High-BAS individuals could be more focused on rewards and proactive behaviors particularly in a cooperative setting.

Balconi et al., 2020 EEG

Inter-brain connectivity/inter-subjective coherence was obtained via calculation of the partial correlation coefficient for each pair of channels and dyad per frequency band. Mixed measure ANOVAs were run to observe the effects of conditions, electrode site, and role on the neurophysiological dependent variables (frequency band signals). ANOVAs were then conducted in a similar manner with EEG coherence indices as dependent variables.

Reviews without numerical ratings were associated with positive feelings, elevated dyadic engagement, and inter-brain coherence.

Barraza et al.,  
2020 EEG

Essentially, the phase difference between electrode pairs in a time window was computed, followed by an analysis of the stability of difference. IBS matrices were averaged across electrode pairs and trials to result in a grand average time-frequency matrix for each dyad, which were then compared within-task with paired t-tests to reveal effects of time-frequency. Two-way repeated measures ANOVAs observed differences in time-frequency windows with the within-subjects factor of condition (competition, cooperation, solo).

Evidence points towards an idiosyncratic neurodynamic of IBS for social interactions of a competitive and cooperative nature. Theta band activity was associated with inter-subject motor action, and gamma band with cognitive processes underlying cooperation (e.g., shared intentionality). Interestingly, as the reaction times of the subjects decreased (i.e., in the competitive condition), coupled theta band activity increased, and required both subjects to move.

Barzy et al.,  
2020 Eye-tracking

Fixation duration on each AOI per condition and group was calculated for each participant and question and analyzed as a proportion of all fixation durations on all AOIs. Linear mixed models and lmer were used to analyze the data using a distinct model per AOI, with Perspective and Group as fixed effects and Items and Participants as random effects. To compare two levels of Group and three levels of Perspective, the effects were contrast coded.

Autistic individuals spent more time looking at the background and less time looking at the experimenter's face while talking over all perspective conditions. This study had 50 participants with a statistical power of 87.5%, while previous studies had less of both. Social attention's development in ASD and its potential modulation by the effects of social dynamics between interlocutors are not well-understood. Talking about oneself increased gaze towards the experimenter's eyes.



Behrens et al., 2020	Eye-tracking and HR	<p>A lagged windowed cross-correlation analysis quantified physiological synchrony of skin conductance level and heart rate measures separately. To analyze the dynamics of dyadic interactions, both time series were chunked into 8-s windows, which overlapped by 6 s. Maximum cross-correlations among 4 s windows are averaged over all window segments for each Face condition to allow for lag in response. Synchrony measures were used in multilevel linear regression analyses with predictor, outcome variable, main effects, and two-way interaction effects defined depending on the hypothesis tested. A sensitivity analysis was conducted post-hoc to show the minimum true effect detectable is 0.70 with 80% power.</p>	<p>This study makes no claims on the direction of observed effects, social processes may confound models such as the Perception Action Model, which suggests that synchrony drives social perception. Skin conductance but not heart rate synchrony affected but did not predict cooperative success in the modified prisoner's dilemma game.</p>
Birze et al., 2020	HR	<p>A multiple linear regression was calculated to predict RMSSD during an acutely stressful event and were calculated with chronic subjective stressors. Correlation coefficients were obtained for the relations of chronic subjective stressors and RMSSD.</p>	<p>Gendered psychological stressors showed physiological stress responses in the workplace setting, which may impact their long-term health. Variables such as time of day, birth control use, age, and sex mediate HRV. As well, the past experiences of the communicators and the type and duration of stressor may have impacted results.</p>

Cañigueral & Hamilton, 2019

Eye-tracking

Two-way repeated measures ANOVAs were used to analyze gaze to ROIs, using Audience condition and Time window as within-subject factors per task. A mixed ANOVA was applied to test whether gaze patterns post-answering was affected by the answer or belief of being watched (bootstrapped with 10,000 permutations due to lack of data normality).

The less prosocial choices were, the more the participants gazed at their confederate. When participants believe they were being watched, they gazed less on the video of the confederate, perhaps to signal preference for no interaction or for social norm conformity. The type of communicative interaction may determine gaze behavior, if indeed significantly impacted by social norms. A video context may weaken the audience effect compared to real-life.

Cañigueral et al., 2020

Eye-tracking

The authors performed aggregated and time-course analyses of three conditions: two on eye gaze (proportion of time looking to eye region and mouth region) and one on facial motion (number of facial action units). Aggregated analyses used a 2-way repeated measures ANOVA for all time-points and trials for each Condition as the within-subject factor, mean proportion time looking at each ROI as the dependent variable, and Group as the between-subject factor. Time-course analyses were conducted in a similar manner over 10-15 second intervals (5 per question and answer) with Time-window as a within-subject factor.

The belief of being watched and speaking reduced face orientation to the partner in both autistic and typical groups, and both groups used facial displays as a social signal.

Caruana et al., 2020	Eye-tracking	Analyses of linear and logistic mixed random effects were conducted for accuracy and SRT data to probe for stimulus interaction and context effects. Mixed random-effects models account for random effects (item and subject-level variance) when estimating interactions and fixed effects, so they were used instead of ANOVAs. The models were compared with chi-square likelihood ratios to compute the variance explained by the fixed effects and interactions, and the degree to which parameters improved the fit of the model (Caruana et al., 2020).	JA responsivity was impacted by the social partner's non-communicative eye movements' presence and predictability before a bid for JA. Differences in gaze-following expertise were found, opening an avenue to explore its development.
Chanel et al., 2012	EDA, EMG, and HR	Two indices of PC: the correlation coefficient $R_{sig}$ provided an index of compliance per signal in the time domain, the weighted coherence $C_{sig}$ was used as a frequency domain compliance measure (Chanel et al., 2012). For each signal, two tailed $t$ -tests found whether the computed PC indices $Z_{sig}$ were different from 0. Each PC index was a dependent variable in separate analyses of covariance matrices with repeated variables of game mode (competitive/cooperative) and location (lab/home) to discern differences. Correlations of PC with questionnaire items were performed.	PC increased with player self-reported involvement in the social interaction. Higher PC was observed when participants played bomberman competitively vs cooperatively, and participants' PC reactions had higher similarity in competitive play.

Chen et al.,  
2017

EEG

ERPs were constructed by averaging the trials per condition. The mean amplitude of peaks from 200 to 300 ms were calculated for Feedback-related negativity (FRN) peaks, and P300 amplitudes were identified. To disentangle FRN with positive ERP components, opponent Cooperation ERP responses were subtracted from Aggression's for each participant participant choice condition. Fz and Pz electrodes were selected for the FRN and P300 amplitudes. FRN and P300 data were analyzed with a 2 x 2 x 2 repeated measures ANOVA (social distance, participant choice, and opponent choice). The differenceFRN was analyzed with a repeated measures ANOVA 2 x 2 (Social distance x Participant's choice).

FRN and P300 were affected by opponent's choice; P300 was affected by social distance, but FRN was not. Opponent cooperation was associated with smaller FRN and larger P300. Playing with strangers elicited a larger P300 than with friends. Results suggest social distance largely affected the late-stage evaluation of choice outcomes, as in early-stage there were no differences between P300 signals between social distance conditions.

Czeszumski  
et al., 2019

EEG

Exploratory analyses used 62 electrodes. EEG data in the form of ERP analysis: FRNs were defined as the mean amplitudes over six electrodes between 200 and 300 ms after feedback per trial. The FRN was modeled with a single-trial based LMM analysis using outcomes (win/lose) and social situations (cooperative/competitive) as fixed effects with an interaction between them (Czeszumski et al., 2019). Random effects were modeled with random intercepts and slopes. To test for differences in effects of social situation on FRN according to Perspective Taking Scores, Spearman's Rho was calculated. After visual inspection of the ERPs' grand average, a peak-to-peak amplitude analysis was applied using the same frontal electrodes as the mean amplitude analysis. Then, a LMM analysis was conducted. Threshold-free cluster enhancement (TFCE) assessed activations of each electrode and time as an exploratory analysis, with corrections for multiple comparisons. EEG data was analyzed with two-way repeated measures 2 (outcome) x 2 (social situation) x 62 (electrodes) x 600 (time points 0-600 ms) ANOVAs, enhanced with the TFCE method and used 5,000 permutations with random data point assignment to participants. F-values above the 95th percentile in the max F-value distribution were considered significant.

Outcome (positive, negative, or neutral rewards) and social situation (cooperative or competitive) influenced FRN. Feedback processing differs between social situations, with cooperative conditions having more positive amplitudes. Those that scored higher on a perspective taking test showed more striking differences in amplitudes between conditions.

de Bruijn et al., 2011	EEG	Trials were averaged to ERPs per condition and per subject relative to baseline. ERP analyses: Ne/ERN amplitudes were plugged into a 2 x 2 repeated measures GLM with the within-subject factors of correctness and compatibility (de Bruijn et al., 2011).	Error-monitoring behavior was adjusted by subjects in accordance with other's tasks. Shared representations affect the error monitoring process. Increased Ne/ERNs following errors to compatible no-go stimuli, which evidences that when co-performing a task, subjects incorporate the partner's task into one's own error-monitoring process (de Bruijn et al., 2011)
De Lillo et al., 2021	Eye-tracking	The study had a mixed design, crossing the between-subjects factor of Age Group with the within-subjects factors of AOI and Condition (Speaking or Listening). Analyses were done on the proportion of AOIs on each region. Mixed model repeated measures ANOVAs were used. AOI coding reliability was assessed with 10% of the data. Data quality across age group was ensured by running one-way ANOVAs fixation numbers and total duration of fixations.	Adolescents and older adults exhibited reduced social attention compared to young adults in face-to-face conversations while speaking and listening, attributed to cognitive load.
Ferguson & Breheny, 2011	Eye-tracking	Log-ratio measures of probability of fixation on reality or alternative boxes were calculated and then analyzed according to 5 word-regions defined for each trial (e.g., "the" "[object]" "is in" "box" "A/[B]"). Repeated-measures ANOVAs used Movement (move and no-move) and Knowledge (shared and privledged) as factors.	This look-and-listen study evidences, through visual bias, spontaneous use of theory of mind (ToM) during communication. After initial ignorance, a pull of reality (PoR) effect occurs when hearing the later words that overrides the speaker's perspective due to knowledge that the speaker did not see the last half of the video.

Fishman &  
Ng, 2013

EEG

ERN was averaged across individuals in each extraversion group (low, mid, and high) separately for correct and error trials in both feedback conditions. A 2 x 2 within-subjects repeated measures ANOVA to reveal the relation of ERN amplitude to the social context of feedback with both correct and error feedback. To test if ERN amplitude was a function of feedback condition and extraversion, a 2 x 3 repeated-measures ANOVA was conducted (feedback context and extraversion group, respectively). Follow-up one-way ANOVAs were run to explore, per social feedback condition, the differences in ERN with extraversion groups as a between-subjects factor. Paired-samples t-tests were used to compare the ERN amplitudes across social contexts separately for each extraversion group.

When feedback was non-social, ERNs were similar across subjects of differing levels of extraversion. In the social feedback condition, extraversion was positively associated with absolute ERN amplitude, reflecting that engagement in feedback is affected by feedback modality.

Freeth et al.,  
2013

Eye-  
tracking

The results were divided into an Ask and Answer phase, and the first 30 s of each were used to calculate the proportion of viewing time on each AOI (face or head, body, background). Values greater than 2 SD from the mean were removed, resulting in one participant's data being excluded. A 2x3x2 repeated measures ANOVA (Phase x Region x Eye Contact) was run. Since the experimenters were female, a between subjects factor of gender was tested. Multiple bivariate Pearson's correlations were conducted to test the relations between Autistic Quotient (AQ) scores and gaze patterns.

A large portion of time was spent looking at the experimenter's face, especially when being asked a question. Eyes most often gazed on the background when answering. Direct eye contact increased the likelihood of looking at the experimenter's face rather than body. AQ did not correlate with time looking at the experimenter, differing from previous video studies. Real-life interactions may present greater affordances than video-based interactions that influence participant gaze behavior in relation to their AQ.



Fu et al., 2019	Eye-tracking	Linear regression models were run to investigate whether BI status affected visual attention on the stranger (total fixation number, mean latency of reengagement, mean fixation duration, and proportion of dwell time).	A mixed-measures analysis of variance was used to test the effects of emotion and BI status on AB frequency scores (Fu et al., 2019). For significant correlations between any stationary eye-tracking AB score and social withdrawal (SW) levels, researchers then examined the effect of the AB score on SW, controlling for BI using a linear regression model (Fu et al., 2019). "Four linear regressions models were run with stranger presence duration and error margin coding as covariates, BI status as the predictor, and (a) total number of visits to the stranger, (b) mean latency of reengagement, (c) mean visit duration, and (d) proportion of dwell time on the stranger as the dependent variables (DV), respectively" (Fu et al., 2019). To investigate if BI is linked to a distinct pattern of ambulatory attention, they used a multivariate analysis of covariance while controlling for the previous model's covariates (Fu et al., 2019).
Haensel et al., 2020	Eye-tracking	"A 2 (Group: British/Irish, Japanese) × 2 (Speech: speaking, listening) × 2 (Task: introduction, storytelling) mixed ANOVA was conducted on fixation time, separately for proportional face and upper face looking...Shapiro-Wilk Tests suggested that the assumption of normality was not always met ( $p < 0.05$ ) and this could not be corrected with data transformations" (Haensel et al., 2020).	ROI and permutation analysis both revealed cultural differences in face scanning. Eye movements need to be precisely tracked in order to evaluate eye contact. Cultural modulations of face scanning occurs in this case.

Hanley et al., 2015	Eye-tracking	A priori power calculations were run to determine the number of participants for a power of 0.8 and an alpha of less than 0.05. Pearson's correlations were run between participant characteristics (performance IQ, Vocabulary and Similarities IQ, and age) and fixation frequency.	Cognitively able adults with ASD may miss non-verbal signs in social interactions. There is likely a link between social information detection and social awareness measurement in adults with ASD.
Harrison et al., 2019	Eye-tracking	A repeated-measures, factorial multiple analysis of covariance (MANCOVA) model was used to analyze data. Group was the independent variable and eye contact frequency and duration were the dependent variables. Cohen's <i>d</i> were run to determine effect sizes	Medium to large reductions in the frequency and duration of eye contact were found for AN participants, and recovered-AN participants showed an intermediate profile between AN and non-AN.
Hernández-Gutiérrez et al., 2018	EEG	Expectancy and Presentation mode served as the factors, with separate ANOVAs performed on average amplitudes within time windows determined based on visual inspection. A mixed ANOVA was used to compare the potential influence of differential face visibility during speech processing (Exp 1-3), with Experiment as a group factor and expectancy and presentation mode as within-subject factors. To determine if the N400 and late posterior positive activity were confounded by inclusion of two components of opposite polarity over similar regions, separate mixed ANOVAs were run with experiment and facial feature as group factors, and presentation mode as a within-subject factor (Hernández-Gutiérrez et al., 2018).	Closer resemblance to natural communication a communicative situation was, the higher the attentional processing, given semantic speech processing was not difficult. The N400 semantic effect was not modulated by dynamic facial information. Word expectedness did not create more semantic speech processing demand. Predictable speech, instead, generated a prolonged late-posterior positive signal, which indicates elevated motivated attention.

Hessels et al., 2017	Eye-tracking	Mean duration of eye contact was calculated from simultaneous dwells on eyes AOIs. Intraclass correlations (ICCs) were used to calculate strength of gaze agreement. A repeated-measures ANOVA on dwell time with AOI as a factor was run. In experiment 2, repeated-measures ANOVA was run to check that confederate instructions were followed. Shannon entropy was calculated for AOI distributions to determine uncertainty of gaze, a factor then used in repeated-measures ANOVAs with observer's gaze on AOIs as a factor.	In experiment 1, the ICC = 0.75, $p < .005$ , and indicated a strong correlation between time spent looking at the partner's eyes and the partner's eyes reciprocating. Gaze behavior within dyads appeared to be dyad-dependent. When interaction is possible, dyads have a bias to fixate each other's eyes. Gaze behavior of one partner does not easily manipulate the other partner's. Further investigation is required to pinpoint how correlational gaze behavior develops over time.
Hessels et al., 2018	Eye-tracking	To test directional hypotheses, one-tailed Pearson product-moment correlations investigated four hypotheses, e.g., if Autistic Quotient (AQ) and Social Anxiety Score (SAS) are correlated to total AOI dwell time (separately).	ASD and SAD traits predicted gaze duration on specific ROIs fitting with current models of ASD's gaze aversion and SAD's vigilant-avoidance. Understanding eye movements in social interactions benefit from this dual eye-tracking setup.
Hietanen et al., 2020	EDA & EMG	Change was calculated by subtracting average baseline values of 500 ms epochs per participant. Within-subjects ANOVAs compared responses of EMG and SCRs over conditions and gaze directions. EMGs were also compared over time (may spike or be prolonged depending on emotional processes). Paired samples $t$ -tests were used to probe for differences between gaze conditions when interactions were discovered between gaze direction and conditions.	Being seen, not physical presence, was prerequisite for autonomic and affective arousal: no differences were found between SCR in live and video call conditions. Direct gaze was arousing in all but the video condition. Direct gaze elicited more positive affective facial reactions than averted in all conditions.

Ho et al.,  
2015

Eye-  
tracking

This study employed the use of non-linear cross-correlations (CC) to reveal the relations between gaze and speech over time. Lag was constrained to 2 s to investigate turn-taking transitions (and pay less attention to the overall rhythm). Dyad- and participant-level analyses were carried out: correlations and maximum lags of speaking and gazing were computed for each game and then averaged across all dyads or participants. CC was carried out between a participant's own gaze and talking to detect speaking-gaze behavior at an individual level. CC was used to compute the maximum correlation between a participant's gaze and their partner's talking to investigate turn-taking signalling. Bouts of gaze and speaking were plotted in time series bar graphs and descriptively analyzed. Researchers noticed patterns of alternating gaze and speech that signal turn-taking confirmed in later quantitative analyses.

Speakers end their turn with a gaze towards partner, and begin their turn with an averted gaze relative to their partner, which then returns after establishing that the turn has been taken. The relationship between speech and gaze has a significant task-dependent component.

Hoehl et al.,  
2014

EEG

Continuous wavelet transformations were used to compute total-induced oscillations. Data was trimmed and baseline was subtracted, then the grand average was calculated for EC and NEC conditions. A two-way ANOVA was run with the factors EC condition and electrode between 5-7 Hz 400-800 ms after object onset (Hoehl et al., 2014).

Similar brain networks play roles in joint attention interactions in adults and infants, as indicated by EC leading to desynchronization of alpha-band activity after object presentation in a live joint-attention setting. Eye contact could act as a primer for infants to access their emerging semantic knowledge system (Hoehl et al., 2014).

Holleman et al., 2020	Eye-tracking	Relative total dwell time was calculated for each participant. Mann-Whitney-Wilcoxon test (MWW) examined "whether the median relative total dwell time on the eyes-AOI was significantly smaller for the live-instruction group compared with the pre-recorded instruction group" (Holleman et al., 2020).	Participants varied greatly in their gaze behavior to the confederates; only weak evidence exists in the present study for the 'social risk' hypothesis.
Honma et al., 2012	Eye-tracking & pupillary dilation	Repeated-measures ANOVAs with Bonerroni correction examined the differences in error distance (ED) and pupillary dilation between 27 points (Honma et al., 2012). Welch's <i>t</i> -tests compared pupillary dilation between perceived pupil or outside-pupil fixation by viewer. Distance from pupil (DfP) and ED or pupillary dilation relations were assessed with single logarithmic and linear regression analyses. Pearson's coefficients checked for within-gender correlations and then a parallel line analysis confirmed lack of interaction in regression lines. ANCOVA probed for gender differences in pupillary dilation and ED.	Perceivers were unable to correctly judge gaze direction of the viewers, yet pupillary diameter was significantly larger during perceived pupil fixation from the viewer. Females showed a higher accuracy of gaze direction than men.

Hu et al., 2018	EEG	<p>For IBS, two-way ANOVAs were run per electrode pair, with FDR. Pearson correlation coefficients evaluated the link between cooperation performance and IBS. Mediation analysis was conducted to test for perceived cooperativeness as a mediator for IBS, with age as covariate (Hu et al., 2018). Control analyses were performed to evaluate the effects of cooperation condition on strategy choice, presumed human/machine on IBS, or order of trials using 2-way repeated measures ANOVAs. Pearson correlation analyses were used to assess the relations between selection rates and IBS (all bands).</p>	<p>Neural evidence of cooperative activity generating IBS during interactive decision making. The context of cooperation impacts IBS. Certain cooperative contexts allow eventual collection of socially-relevant information to enable high-level social cognitive processing.</p>
Ioannou et al., 2014	Eye-tracking & facial temperature	<p>Kappa measure of agreements were run to assure inter-rater reliability, and ANOVAs were run on the results between raters. Correlations were run between all six sites on the face for each condition. All ROI were averaged, and a 2 x 2 x 2 mixed repeated measures ANOVA was conducted to observe a pattern on the effects of interpersonal distance and gaze on facial temperature. Individual region analyses were done by averaging all individuals' six ROIs, running a 2 x 2 x 2 ANOVA with order of the conditions, eye contact, and proximity as the factors.</p>	<p>Direct and averted gaze increased facial temperature regardless of the interpersonal distance in the study. An approach from social to intimate distance elevated facial temperature of the participant, but order was crucial to maintain the effect. Skin temperature did not recover after the most arousing condition until exposure to the least arousing. The results suggest a preparatory autonomic response was initiated due to the social interactive cues of specifically ordered proximity and direct gaze.</p>

Järvelä et al., 2014	EDA & HR	<p>Indices of physiological linkage were averaged over the whole session. Analyses were conducted on the dyad-level, with means and differences (between members' measurements) per moment. Regular <i>t</i>-tests were used to determine differences from zero (due to normalization procedures). A Linear Mixed Model was used to analyze the condition changes and associations between physiological and questionnaire data. Determined, through the use of Hurvich and Tsai's criterion, that first-order autoregressive covariance structure produced the best-fit for the physiological data.</p>	<p>Physiological linkage exists between players of a multiplayer game, and the linkage changes based on social interaction and context, which supports the notion of emotional contagion. Competitive and cooperative modes had no influence on linkage, but it might be due to the context (light-hearted game). Presence of an AI agent can disturb social presence effects and engagement. AI agent actions are perceived as less socially relevant.</p>
Jiang et al., 2015	fNIRS	<p>IBS analysis: The Wavelet transform coherence (WTC) method assessed the cross-correlation between two participants' fNIRS time series. Granger causality analysis (GCA) was conducted on time series channel data to assess directionality of synchronization. One-way <i>t</i>-tests showed significant difference from zero. Communication behavior's relation to IBS was tested with fisher linear discrimination analysis for selected INS peaks that matched video instances of communication for LF and FF pairs. Cumulative IBS over time was tested as a predictor of LF or FF by leave-one-out cross-validation.</p>	<p>The LF pairs in the LGD situation presented elevated IBS, specifically during verbal interactions. Leader-initiated interactions led to greater IBS. Discriminant analyses helped identify LF or FF pairs shortly after session initiation.</p>

Jongerius et al., 2021	Eye-tracking	<p>A multilevel analysis was conducted "with face gaze (average dwell time) as an independent variable and patients' reported trust in physicians as a dependent variable" (Jongerius et al., 2021). "To test if social anxiety moderated the effect of face gaze on trust, we examined the interaction between social anxiety and face gaze" with repeated use of multilevel models (Jongerius et al., 2021).</p>	<p>Mobile eye-tracking can allow highly detailed study of gaze for a high number of participants. Initial gaze behavior of the first minute provides a relatively strong prediction of the entire session. Face gaze duration was longest on average in the first decile and lowest in the ninth, suggesting a temporal relationship that may play roles in many naturalistic settings.</p>
Kajopoulos et al., 2021	Eye-tracking	<p>Mean fixation times were calculated for each AOI and analyzed with ANOVA with Distance, Gaze, and AOI as within-subjects factors and Germany or Japan as a between-subjects factor. Probability of gaze to AOIs were calculated as relative frequencies. "Fixation Time [%] scores were examined in a mixed-design analysis of variance (ANOVA) with Distance (Far, Near), Gaze (With Gaze, Without Gaze), and AOI (Face, Body, Target, and White Space) as within-subjects factors and Experiment (Germany and Japan) as [a] between-subjects factor" (Kajopoulos et al., 2021). Another ANOVA was run to test for Gender effects instead of Experiment.</p>	<p>This study provides evidence for a substantial top-down component to gaze behavior in social situations, i.e., intentionally ignoring irrelevant gestures. Culture and gender showed no effects.</p>



Kendrick &  
Holler, 2017

Eye-  
tracking

The time between the latter two points is the aversion-completion distance (in ms). The data were analyzed using mixed effects logistic regression models. Different predictor effects were compared through model comparisons.

Gaze aversion was more frequent in dispreferred responses and in their reception as seen by an analysis of self-repairs by questioners in the transition space between turns. Preference and complexity influenced gaze direction. The formation that participants sat in increased the likelihood of mutual gaze and hence gaze aversion, whereas other natural orientations would reduce mutual gaze at the outset.

Kraus et al.,  
2020

EDA & EEG

All data was averaged from repetitions of the same condition. All data was log transformed to achieve normality. EEG: ANOVA and linear trend analyses were conducted with experimental conditions (Alone, Stranger, and Partner) as factors to see effects on EEG power band (alpha, beta, delta, gamma, and theta) activity. ANOVAs were run to detect hemispheric differences of EEG power. Tukey tests were used post-hoc. EDA: To assess the effects of aversive visual stimuli on EDA, number of fluctuations (NSF) of NSSCRs were analyzed; baseline compared via *t*-tests to imagery were compared in the Alone condition. Pearson product moment correlations were run on each baseline and image blocks across all conditions between EEG frequency bands and EDA.

This study evidenced social touch reduces theta power band activity, and is modulated by intimacy, as shown by stranger and attachment style conditions and factors. Frontal theta power reflected affect-laden processing in interpersonal space, which was moderately correlated to NSSCRs, here a measure of emotional arousal and apprehension (Kraus et al., 2020).

Lenoble & El Haj, 2021	Eye-tracking	Calculated variables: fixation number (/min), fixation duration, saccade number (/min), saccade duration, and total amplitude (angle saccade covered). Mann-Whitney <i>U</i> tests were used to compare between-subject designs, and Cohen's <i>d</i> was used to report effect size.	The participants had more fixations, shorter fixations, more saccades, and longer duration of saccades while wearing eye-tracking glasses compared to looking at a 21" blank screen while recounting an autobiographical memory.
Magrelli et al., 2013	Eye-tracking	Mixed-effect model regression using the maximum likelihood method (ML) was employed; a series of mixed-effect linear models were assessed to evaluate the effect of the social cue, autism, age, task, and gender on the gaze-to-face distance across time (Magrelli et al., 2013). Models accounting for the same factors also looked at effects on first fixations lengths (FELs) and reaction times (RTs). Nested models were compared to determine one with the best fit. Various models correct for number of estimated parameters and/or observations.	Orientation was not requested; the study presented a naturalistic approach in which spontaneous orientations to social cues were observed. Children with ASC orient slower and less to social cues.
Mitkidis et al., 2015	HR	T-tests were run on MVRQA and %Determinism between participants with the between-participant factor of PGG (yes/no). To determine if heart rate synchrony or arousal predicted subjects' performance in the TC, regression models were run with heart rate predictors and expectations of returns and investments as dependent variables. For this analysis due to repeated measures (x4), heart rate synchrony, expectations of returns and investments for each PGG per dyad was averaged to obtain one value.	Heart rate synchrony significantly predicted PGG expectations and may indicate a potential physiological marker for interpersonal trust. Heart rate was an indicator of positive excitement, possibly denoted an increased awareness of the subject's partner, and the synchrony itself is a marker of this reciprocation. This study evidences the potential for heart rate synchrony as a proxy for the process of trust formation.

Mønster et al., 2016	EDA, EMG & HR	CRQA estimated synchrony between measures. Determinism, average diagonal line length, and longest diagonal line length were used to analyze the plots. PCA used to reduce variables and CRQA measure dimensionality for interpretation. False-pair surrogate analysis was conducted to assess which physiological measures show synchrony above-chance.	In this study, emotional induction and collective choice were sensitively detected by measures of synchrony. Higher synchrony was associated with instances of negative group affect.
Nelson & Mondloch, 2019	Eye- tracking	Total looking time was a proportion for each participant. A post-hoc power analysis resulted in 0.99 for detection of a medium-sized effect. ANOVAs with Bonferroni-corrected post hoc tests examined the variation of visual attention (on ROIs) with age and experimenter emotion. Correlations were run between gaze proportion and accuracy of response.	Children spent less time looking at the face than adults, but patterns of ROI gazing were similar to adults. Children scanned distractor objects more than adults when discrimination of facial expression difficulty increased, suggesting the task was more difficult for children than for adults.

Nguyen et al., fNIRS  
2020

GLMM analysis assessed cortical activation patterns throughout the varied conditions. The onset and duration of each condition was modeled in seconds (Nguyen et al., 2020). WTC was used to calculate IBS/INS due to its ability to consider global activity in an in-phase and lagged-phase fashion. Task duration was established based on template completion duration, visual inspection, and spectral analyses. Then, the frequency band 0.02 - 0.10 Hz (corresponding to 10 - 50 s) was shown as task-related. Average coherence (IBS) was calculated for the phases. 3 conditions x 16 channels coherence values were found for each dyad. Neural synchrony values were Fisher's z-transformed, random pair analysis was conducted with 1,000 permutations, and  $p$ -values were corrected with a FDR for multiple comparisons.

Frontal and temporal areas of mother-child dyads showed significant IBS during a naturalistic cooperation task. Behavioral reciprocity, an indicator of parent-child quality of interaction, was positively correlated with IBS. Beyond behavioral reciprocity, IBS was associated with cooperative task performance; further, IBS was influenced by child agency and maternal stress. Investigation of the individual and dyadic factors for IBS must be extended. Father and other caregiver-child dyads' IBS can be explored, as well as factors such as preterm birth, postnatal depression, or the average quality and quantity of time spent with the child.

Noy et al,  
2015

HR

Co-confident motion (CC) and non-CC were automatically detected by an algorithm. Round-level analyses were conducted by calculating medians for each measure, as well as proportion of round CC. For physiological measures' round-level analysis, "arousal was calculated as a median value of individual HR time-series (HR) in each round" (Noy et al., 2015). Physiological coupling was calculated by Pearson correlations between HR time-series of each player in each round. For segment-level analysis, z-normalization of all data points of each player was done to remove individual differences in baseline HR. The arousal of each segment was indexed as average zHR. Segment-by-segment analyses pooled all players and rounds and compared CC vs. non-CC and SR-high vs. SR-low (subjective ratings of togetherness high and low). Non-parametric bootstrapping was used to overcome the inherent dependence between dyads and consecutive segments. Actual mean difference of two groups vs distribution of mean differences of random bootstrapped samples was done 10,000 times. The *p*-values were estimated from control distributions and were subjected to correction.

High kinematic and subjective togetherness are associated with elevated HR, higher inter-player HR correlations, and higher motion intensity (frequency and max velocity). Controlling for motion intensity, both higher HR and higher inter-player HR correlations are still found. Normalized HR was higher in kinematic and subjective measures of togetherness.

Oberwelland et al., 2016	Eye-tracking & fMRI	A mixed ANOVA model evaluated effects of experimental condition and random effects (within-participant and group). Eye-tracking: A Mixed 2 x 2 x 2 ANOVA was run on number of saccades, errors and latency (separately) differences between within-group factors of JA and Familiarity and age as between-group. Paired sample <i>t</i> -tests were run on proportion of self-initiated gaze shifts on different targets. A mixed ANOVA evaluated differences between target fixation proportions during JA or Familiarity conditions.	Demonstrated modulation of the JA network by familiarity of the interaction partner and initiation type. Self-initiated JA with a familiar partner showed the highest magnitude of modulatory activity from JA ROIs, but not other-initiated JA. Familiarity significantly elevated activity in ROIs associated with social-communicative stimuli processing. Developmental trajectories of JA were established to a limited extent.
Pérez et al., 2019	EEG	Synchronization of each frequency band was estimated between 27 channels for each speaker and listener (each subject was a speaker and listener per block, so they are assessed independently). To compare IBS to brain-to-speech synchronization between language conditions, a nonparametric bootstrap-based <i>t</i> -test method was used (10,000 permutations) with FDR for multiple comparisons (Pérez et al., 2019).	Linguistic context impacted IBS (in this context of dissimilarly proficient subjects). Linguistic factors that mediate IBS should be explored. Preventing subjects seeing each other can provide a more controlled setting, yet may lower quality of communication, lower ecological validity, and result in different brain-to-brain (B2B) signals.
Peshkovskaya & Myagkov, 2020	Eye-tracking	Eye-gaze parameters used: revisits, fixation count, dwell time, fixation time, and average fixation duration. ANOVAs and Mann-Whitney <i>U</i> -tests were run to explore the relations between the experimental variables and gaze behavior. Phase, game role, strategy, and their combined effects were also investigated.	Gaze behavior predicted strategic-based behavior. Cooperative subjects showed more attention to all stimuli and revisited scanned areas more. Defectors show long fixation durations on areas and low numbers of revisits. The time spent gazing at the payoff matrix was not significantly different between cooperators and defectors.

Peters et al.,  
2014

Pre-ejection  
period  
(PEP),  
cardiac  
output (CO)  
and total  
peripheral  
resistance  
(TPR)

"Data were analyzed in 2 (Emotion Regulation Condition: suppression vs. expression) × 2 (Role: sender vs. receiver) mixed ANOVAs" (Peters et al., 2014). Data before conversation was treated as independent because dyads were not formed until the conversation; during and after conversation, data were analyzed to account for non-independence. Mixed ANOVAs assessed affective responses to watching the video, partner attributions and interaction attributions. Physiological measures: Pre-ejection period (PEP) measures sympathetic activation, and cardiac output (CO) and total peripheral resistance (TPR) allow differentiation between "approach-motivated challenge and avoidance-motivated threat states" (Peters et al., 2014). Baselines of these measures were compared. Mixed ANOVAs assessed differences of all measures to baseline during anticipation and conversation.

Suppression of affective signals, even if negatively-valenced, caused threat-related physiological responses for both regulators and their partners. Statistically, gender did not impact the observed effects. Studies employ disparate conceptualizations of suppression during expression, some via a conversation allow/disallow paradigm. Emotional expression norms and effects may vary across cultures and race, though no effects were evidenced in the study (with low power of analysis, however).

Pfeiffer et al.,  
2012

Eye-  
tracking

A two-way repeated measures ANOVA was run to compare the effects of latency on perceived relatedness of gaze reaction between JA and non-joint-attention groups (NJA). A one-way repeated measures ANOVA between latency and relatedness of gaze reaction in the shared attention experiment was also conducted. *t*-tests were run to compare JA in 2 experiments for each latency.

Sense of agency was more sensitive to latency in the shared attention condition, meaning it declined more rapidly as latency increased than the joint attention experiment. Participants' sense of agency was less sensitive to latency when their confederate had action alternatives. Peak relatedness rating was reached earlier (400 ms) and declined linearly when confederates had no alternatives, and not before 800 ms with alternatives. SA took longer to establish and almost always included gaze back to the confederate's face, though in JA this rarely occurred. SA was highly variable between individuals; thus, measures of empathy, systemizing, agreeableness, etc. could be linked.



Potts et al., 2019	EDA	<p>ANOVAs were applied to the effects of stress conditions on SCL, and interactions were explored with Bonferroni-Holm corrected planned <i>t</i>-tests. To assess acute stress impacts on sympathetic-adrenal-medullary (SAM) activation, a 3 (Stress Group: Control, CPT, SECPT) x 3 (Exposure Phase: Baseline, Exposure, Post-Exposure) x 2 (Latency: Fast, Slow) x 2 (Sex: Female, Male) mixed ANOVA was performed on square-root transformed SCL (Potts et al., 2019). The independent variable of SAM reactivity was analyzed. The continuous covariate of z-standardized SCL represented SAM reactivity. The use of GEE here was to explore only the impact of physiological reactivity and latency on propensity to trust.</p>	<p>SAM reactivity did not vary by stressor type. Acute stress overall was associated with reduced trust. Models of predicted trust were complex, involving SAM reactivity and other signals (cortisol) in a time-dependent manner. The ordinal logistic model predicted reduced trust likelihood as SAM engagement increased, which is also reflected in the negative correlation between SCL and number of investments (Potts et al., 2019). Acute stress can lower and increase the likelihood of trust based on specific profiles of physiological reactivity.</p>
Prinsen et al., 2019	EDA	<p>SCR differences across gaze conditions and subjective ratings were evaluated with repeated measures ANOVAs. Separate ANCOVA models for each social anxiety dimension were run to investigate if SCRs and subjective ratings were affected by State Adult Attachment Measurement (SAAM) scores between subjects. Gender and age of the participants were then included in all models as covariates of no interest to evaluate robustness of observed effects.</p>	<p>SCRs and subjective ratings of arousal were higher in response to reciprocal gaze, and low-attachment and high-avoidance subjects showed greater SCRs when viewing the confederate, regardless of gaze reciprocity. Thus, internal working models of attachment modulate psychophysiological reactions in a paired gaze context (Prinsen et al., 2019).</p>

Rahal et al., 2020	Eye-tracking	A mixed effects repeated measures linear regression analyzed the proportion of attention to own payoffs from Social Value Orientation (SVO) angle, group setting, and the respective interaction. Three mixed effects repeated measures linear regressions were run to predict effort in information search of components number of inspected informatoin, number of fixations, and decision times with predictors SVO angle, group setting, and respective interaction.	Behavioral ingroup-based generosity was evidenced; eye-tracking can inform cognitive processes underlying these types of decisions. Decision makers weighed in-group outcomes more than out-group based on fixation data.
Rauchbauer et al., 2019	Eye-tracking & fMRI	fMRI time series were analyzed with the GLMM: single-subject models had one regressor for the 1 min discussion for the social interaction and another for image presentation. A mixed-model ANOVA with sessions and participants as random factors and interacting agent as the factor of interest for inferences at the population level (Rauchbauer et al., 2019). The main effect of the conversation agent was compared to baseline, followed by a comparison between conversation agents (HHI vs HRI), focusing on mentalizing and social motivation ROIs. A threshold of $p < 0.05$ FDR-corrected at the cluster-level for the whole brain was used for all statistical inferences (Rauchbauer et al., 2019).	Preserving the reciprocal dynamics of the interaction via utilization of a robot head with a retroprojected face in the control condition allowed comparisons between a HHI and HRI using fMRI. HHI was associated with social motivation and mentalizing. The exploration of human social cognition could use this type of experimental paradigm; as well, further work with humanistic robot interactions can improve their social competence.

Rösler et al.,  
2021

EDA & HR

For HR, HRV, SCL, and gaze data, 4 separate ANCOVAs were conducted to determine the effects of social anxiety. Fixation proportions were modeled using the mean-centered SIAS score and phase of experiment as predictors and a two-level factor of ROI (head or body). HR, HRV and SCL were the same apart from the ROI. To analyze if participants exhibit particular gaze patterns, split-half consistencies on head fixations per phase were computed and Pearson's correlation coefficients were run within subject.

The likelihood of a confederate returning gaze or initiating a social interaction impacted gaze behavior in sub-clinical socially anxious subjects (i.e., when the confederate was on the phone compared to when it was waiting). Social anxiety levels did not affect gaze behavior. HR consistently correlated with social anxiety scores, indicating ability of subjects to keep socially perceivable and relevant behaviors such as eye contact in a normal state at the cost of less perceivable physiological changes. The conversation elicited highest levels of HR, HRV, and SCL. SCL elevated also during the phone call. The naturality precludes investigation of responses to emotional valence. An increase in movement during speech could also arouse subjects. A higher baseline heart rate while in the presence of another is possible. Ambulatory heart rate readings may shed light on variability throughout every-day situations and interactions.

Ruissen & de  
Bruijn, 2015

EEG

The late phase of the P3 component was analyzed. N2, early P3, and late P3 amplitudes were analyzed using a General Linear Model (GLM) with Context, Compatability, Stimulus Type and Electrode as within-subjects factors and Condition as the between-subjects factor (Ruissen & de Bruijn, 2015). Repeated measures ANOVAs tested for effects of all factors on N2, early P3, and late P3 component separately.

There are clear indications of shared representations (through observation of the simon effect, SSE) at the electrophysiological level, but very small differences behaviorally. The N2 component showed oxytocin-induced modulations, evidencing an enhancement of joint task processes at the electrophysiological level (Ruissen & de Bruijn, 2015). The authors suggest low-level processes of perceived similarity, self-other integration, and response conflict underlie higher-level processes, all of which have been shown to be modulated by oxytocin.

Santamaria et EEG  
al., 2020

Connectivity analysis utilized multivariate models using model order 5, which explained most of the adult and infant data. Validation of the fitted model was performed with the Ding and r-squared methods, which both indicated good model estimation. Surrogate data analysis was performed to ensure results showed significantly above-chance intra-brain connectivity. Intra-brain and inter-brain metrics were chosen (4 per) from the graph theoretical indices of network topology, where nodes are EEG electrodes and edges are connections to gain insight into connectivity properties. Intra- and inter-brain densities were calculated as the ratio of existing, significant edges to the total number of possible edges. Paired t-tests, corrected for multiple comparisons, were run of intra-brain network density for each subject and metric. Inter-brain density (IBD) was compared across conditions with an ANOVA. Differences in topography of inter-brain connectivity were examined through the threshold connectivity matrices of Strength and Divisibility, and across both metrics, showed the inter-brain network was stronger during the positive condition. Then, the directionality of PDC allowed discernment of mother-to-infant and vice versa strength during each condition. Control analyses included loudness as a covariate to rule out differences in connectivity were not due to utterance volume.

Child (infant) and mother showed stronger neural integration during positive than negative emotional states, but no relationship was found for emotional valence. Inter-brain metrics reveal that mothers had greater influence over the dyadic network during positive emotional states and infants more during the negative. Inter-brain graph matrices can be applied to explain the effects of emotional quality and tone during dyadic social interactions.

Sariñana-González et al., 2019

HR & HRV

Detection of group effects by task, outcome, and gender, a 3 x 2 x 2 ANOVA was run, with Bonferroni post hoc tests. A repeated-measures ANOVA with a general linear model was utilized to examine differences in HR and HRV within groups between periods. Estimation of the magnitude of HR and HFnu responses from task was done through application of the trapezoidal rule and calculation of the area under the curve with respect to the ground (AUCg), which evaluates the distance to the ground for all measures.

Participants who cooperated had higher HRs than those who completed a nonsocial task or competed and had lower HRV-HF levels than those who worked alone. There were gender differences present, i.e. women had higher HRs than men, suggesting an ANS sensitivity to the lab environment. Men were more affected by their performance outcome in cooperation tasks. This is a valid model for analyzing cooperation and competition of a controlled task and its outcome. The building block task was not stressful but activating to the ANS.

Shalom et al., 2015

EDA & HR

One-way within subjects ANOVAs were conducted for HR and SC separately to observe differences between baseline and conditions. A 2 x 2 (anxiety level x condition) mixed model ANOVA was run for HR and SC separately.

Online text interaction (CMC) was intense enough to stimulate arousal in subjects across anxiety levels. No objective (physiological response) differences in arousal were found between FTF and CMC conditions, despite disagreement with subjective reports of arousal and anxiety. Though CMC was reported to induce less anxiety, be more controllable, less threatening, and facilitate success, the objective measures showed no significant differences. No differences in HR or SC were observed between subjects of different anxiety levels.

Soto-Icaza et al., 2019 EEG

TF analysis was made with a Wavelet transform. Analysis for TPJ and the Intraparietal Sulcus (IPS) ROIs because they were identified as the hotspot between significant source of JA regressor for the TD group (Soto-Icaza et al., 2019). Non-parametric tests determined differences between means. A permutation was run on randomized data points to estimate that the results were due to chance. T-value TF charts served as inputs for each regressor and subject, followed by a Wilcoxon test, the results of which were corrected by cluster-based permutations. For ERP analyses, modulation in the Nc component was assessed. To avoid false-positive results, spatial ROI lacking temporal assumptions were compared. This identified the spatio-temporal ROIs of the Nc component. Broad band activity due to muscular movement in the appropriate ROIs was identified with a control analysis.

A possible neural mechanism contributing to the initiation of JA to explicit mentalization was found: the Nc ERP shows modulation of attentional processing related to JA behavior. Importantly, ASD children show a different modulation. Explicit mentalization is associated with beta band activity in the TPJ.

St. John, et al., 2016	EEG	Repeated-measures ANOVAs were conducted for 4-6 Hz and 6-9 Hz with condition (nonsocial and joint attention), region, and hemisphere as within-participants factors. If the ANOVAs showed significant differences, ANOVAs of language-only and social engagement conditions were run to tease apart effects.	Controlled social and social conditions allowed the systematic analysis of functional neural activation. In the joint attention condition, frontal, parietal, and temporal lobes all experienced an increase in activation compared with the nonsocial condition. 4-6 Hz band results were the most pronounced. Language input did not explain any EEG data differences. No hemispheric interactions were detected, but future research might explore infant sub-groups, e.g. temperament. Conditions changed rapidly and could influence on another in sequence, but this could mirror reality.
Stevanovic et al., 2019	EDA, EMG & HRV	Mean values were calculated for each 64-s segment. Mixed Models procedure of SPSS 24 was conducted, i.e. a linear over-time actor-partner interdependence model for EDA, heart rate, and HRV-HF per dyad type. Dyad types were analyzed separately for suitable model specification. Low and high dominance (below and above mean) were used as categorical variables for model simplicity. False discovery rate procedure was used to protect against type I error.	Affiliation affected ANS responses in several fashions. High levels of affiliation from partner are calming. Giving of affiliation is arousing and taxing, moments of dominance serve as calming respites. Affiliations are associated with smiling. Affiliation from NT to AS participants may overwhelm with emotional information and induce anxiety.



Stevanovic et al., 2021	EDA	<p>Pearson correlation coefficients for within-dyad EDA signals measured psychophysiological synchrony, with differentiation of proposal and non-proposal time intervals. The mean of dyad correlations was taken after fisher transformation. Dependent t-tests were run to test differences of synchrony correlations between proposal and non-proposal sequences. A GLMM assessed effects of the participant's role and depression diagnosis on EDA response rates (Stevanovic et al., 2021). Chose GLMM due to non-independence (i.e., a proposer is also a receiver). A GLMM was also run to test for effects of whole task duration. To assess the EDA effect of recipient's response on the proposer, another GLMM was used.</p>	<p>EDA response and synchrony were higher during proposal sequences. This study demonstrated the ability to capture these turn-by-turn sequential phenomena through a granular approach. Proposer EDA was higher than the recipient's. Emergence of a joint decision (accepting responses) was arousing and led to higher EDA for the proposer. Proposal-making was more stimulative to EDA for depressed than non-depressed.</p>
Strang et al., 2013	Anterior-posterior head position & IBI	<p>Relations between PBC, perceived team attributes, and team performance were examined; as well, a set of time-series measures (CC, CRQA, and cross-fuzzy entropy [CFEn]) were compared in terms of their characterization of PBC in different comparisons (Strang et al., 2013).</p>	<p>PBC is impacted by team-task environment. PBC was negatively associated with team performance and attributes. Differentiation of roles and strategies may modulate the relationship between PBC, team attributes and performance. Linear (CC) and nonlinear (%REC and CFEn) measures may not always agree on PBC characterizations.</p>

Sun et al.,  
2021

fNIRS

WTC computed IBS. Cohen's d was calculated for t-tests and ANOVAs. Individual differences were ruled out with independent sample t-tests and chi-square tests. A series of 2 x 2 ANOVAs was run to identify channels showing relations between dyad and task condition (S-S/T-S vs alone/paired). To address the IBS changes over time, a two-way ANOVA was run to test for differences between IBS between groups and times (time block defined by time needed to complete a task). Kendall's tau correlations analyses were conducted between IBS and behavioral indices.

A dyad member's social experience was found to affect cooperative task performance. No significant changes in IBS were observed across time. IBS was negatively correlated with task performance.

Tsuji et al.,  
2021

Eye-tracking

Time window analysis was conducted and a regression was run between groups. Growth curve analysis (GCA) was fitted to compare the shape and latency of the gaze curves between groups. Comparison levels were made with all conditions to test whether word recognition was above-chance.

Unfamiliarity with video chat contexts may underlie the failure to recognize words above-chance. Virtual agent group toddlers learned the worst, and in-person the best. "Human" temporal and gestural variations still existed in the video chat scenario that may have contributed to toddler performance being better than in the virtual agent group.

Vanderhasselt et al., 2018

HRV

Data from the resting times of the experiment was used to establish the baseline (15 min), which was then log-transformed and analyzed with a mixed repeated measures ANOVA with time (within) and (between) gender as factors.

Gender-specific results were evident in all measures. HRV was higher for women across the entire duration, which may confer emotion regulation or demonstrate motivation differences; yet, they may also not represent cognitive differences. HRV increase was larger in men than women across all blocks, and within men, negative feedback blocks had greater HRV increases. Self-report measures of stress could be unreliable, but HRV provides an insight into emotional regulation efforts that corroborate the evidence of the former method.

Vanutelli et al., 2017

EDA and HR

A repeated measure ANOVA observed the modulation of dependent variables (SCL, SCR, and HR) throughout the task, and intersubject correlational indices were computed for each pair for each physiological measure. These indices were plugged into ANOVAs for each physiological measure with independent factor feedback and block to find differences in synchrony strength across conditions (Vanutelli et al., 2017). Degrees of freedom, post hoc comparisons, and Bonferroni tests were applied where appropriate.

Elevated intrasubject HR before feedback followed by lower HR after social reinforcement. After feedback, intersubject EDA synchrony increased. SCL synchrony was elevated and SCR was modulated across blocks (elevated synchrony after social feedback and an exponential increasing within the second half) (Vanutelli et al., 2017).

Varlet et al., 2020	EEG & finger movement	<p>Repeated measures ANOVAs were used to compare movement synchronization across conditions. Two-way repeated measures ANOVAs were performed to compare individual EEG self and other condition signals over leadership condition and harmonic (1-5 for self and 1-3 for other). One-way ANOVAs were run to observe the differences of inter-dyadic signals across leadership conditions. Multiple linear regressions analyses (MLRAs) were performed on the mean and standard deviation of the relative phase angles and distance between the dyads' member's fingers. Absolute relative phase values were used to measure deviation from synchrony. MLRAs were conducted to examine the role of roles on EEG signals and general EEG signal to movement synchronization (Varlet et al., 2020).</p>	<p>Leader-follower roles selectively modulated the magnitude of EEG responses related to processing movements of the self and other. Individual differences were significant. "Neural self-other integration" manifested in maximal amplitude EEG signal from participants when no leader was assigned in a joint improvised movement scenario (Varlet et al., 2020).</p>
Vranjes et al., 2019	Eye-tracking	<p>Interpreter and Therapist, separately: descriptive statistics on the absolute and frequency of Nonverbal, Verbal, and Combined listener's responses were computed, as well as their relation to mutual gaze windows. Qualitative analysis of embodied movements: used conversational analysis (CA) to uncover the use of head nods by the interpreter and therapist as resources to affiliate during listening, as well as its interaction and relation with gaze and verbal utterances.</p>	<p>CA allowed uncovering of the use of dual-feedback, verbal and non-verbal listener affiliative actions in this context. A mixed methods approach allowed revelation that the interpreter discreetly coordinates the conversation and affiliates with the patient, with interpreter's nods strongly linked to patient's gaze and served as an affiliation-inviting marker. The therapist actively aims to manifest triadic affective interaction through dual-feedback.</p>

Wagner-  
Altendorf et  
al., 2020

EEG

Repeated measures ANOVA with Huynh-Feldt correction was performed for ERP data analysis. Early ERP components N200 and P300 were not analyzed due to uncertainty of the influence of the audio stimuli used for "truth" or "lie" cues. Electrode positions were analyzed separately with repeated measures 2 x 3 ANOVAs (truth vs. lie) x (left vs. central vs. right laterality) with Huynh-Feldt correction.

Accuracy of detection was higher for true vs false statements. Detection was better than chance, likely due to informant cues. Analysis of the frontal late negativity component (CNV) showed that convincing statements were associated with attenuated CNV compared with unconvincing statements. The interactive nature of this paradigm, i.e. that a partner classified the lie statements, enriches the relations between CNV and deception with the dimension of social interaction. Difference between the truth and lie conditions was not found.

Wallot et al.,  
2016

Hand  
movement  
& HR

MdRQA was used to calculate synchrony of the activity measures and heart rates within dyads (can measure 4 simultaneous signals, whereas CRQA cannot). To further discern whether values are not synchronized by chance, %Determinism was used. This measure captures the adjacent diagonal structures/trajectories over time in an MdRQA plot. False pair analysis was performed to isolate active coordination from effects of the task constraints. Data were analyzed with linear mixed models.

In the hierarchical condition, the degree of hand-movement synchrony was negatively associated with assessed variables of car size and aesthetic appeal. In the free interaction condition, synchrony was positively associated with objective and subjective outcomes. Synchrony was positively related to feelings of cooperation and fun (rapport). Synchrony may be a function of the goal of interaction and may vary in effect with task complexity (Wallot et al., 2016). The social dimension of the interaction may impact the role of synchrony in the quality of said interaction. Heart rate synchrony was not observed potentially due to task (failed to evoke emotions, etc.). Since BPM was averaged across 5-s windows, some relevant dynamics might have been smoothed out. Leader-follower relations cannot be investigated with MdQRA analysis.

Xue et al.,  
2018

fNIRS

WTC was used to compute the relationship between Hbo time series within dyads. Time-averaged WTC resting state values were subtracted. After WTC analysis, significant changes in IBS were found and, in combination with noise considerations, used for frequency band selection. One-sample *t*-tests were used to probe for differences in IBS across channels, and *t*-maps were generated and smoothed. If a channel was found to have a significant IBS, one-way ANOVAs were run with Group as a between-subject factor. Pearson correlations between IBS and creativity scores were run.

Tasks were performed equally as well across dyads. Low-low dyads showed elevated IBS, other groups lacked IBS. Results suggest cooperation as compensation for creativity deficits.

Yamamoto et  
al., 2019

Eye-  
tracking

A generalized linear mixed model (GLMM) was used to estimate factors' effects on the response variable. The analyses estimated factors affecting eye contact quantity with EC bout quantity in EC session as the response variable; and, they estimated factors affecting the quantity of eye-contact initiations, with the quantity of infant- or parent-led EC bouts in the EC session as the response variable. The most parsimonious model was chosen by application of the widely applicable information criterion (WAIC).

Interpersonal distance plotted was inverse U-shaped, which showed the interpersonal distance preferences of both infant and adult; this study was the first to investigate interpersonal distance on infant-parent dyad interactions.

Yu et al., 2017	EDA & Eye-tracking	A linear mixed model (LMM) was used to assess variation of the factors based on condition, participants and confederates. For EDA data, 60 s was extracted and averaged for each 1 s window, normalized the data, transformed the data, then with SCL, used a linear mixed regression model with group, condition, time, and gender as fixed factors and the participants and confederates as random factors.	Interpersonal guilt reduces transgressor's gaze to the eyes and redirects it to the nose, signalling an avoidance motivation in social interaction. The partner's eyes were covered in the Nose group, which may lead to reduced perception of partner pain and hence self-guilt, but authors argue that most information is in the cheeks and nose. Due to the nose indicating pain, aversion from the eyes to the nose region in experiment 1 could be explained by participants seeking the nose region to observe their partner's pain.
-----------------	--------------------	---	---



Zamm et al.,  
2021a

EEG

Pause duration was defined as time between keystrokes in ms and variability of each subject's pause duration was calculated as the coefficient of variation (CV) and represented as percentage of variability of the mean pause duration in that task. Beta ERD was computed for each pause on each channel, and its linear changes across time windows of solo and duet performances were examined at two ROIs (parietal and central) using a linear mixed model (Zamm et al., 2021b). Pause effects on duet synchrony were assessed with a one-way repeated-measures ANOVA, where pause location was denoted as 1 or 0 (after pause or in other location, respectively). A two-way repeated-measures ANOVA examined pause duration with pause number and task as factors within pairs. Effect of task on pause variability was assessed with a two-way repeated-measures ANOVA, with task and pause location as factors within pairs. Spearman's correlation calculated the relations between duet pause duration and duet asynchronies. Effects of task and pause duration on beta ERD were found from linear mixed model predictions.

Pauses were shorter in duet performances compared to solo, and shorter pauses used in duet performances were associated with elevated levels of synchrony. Beta ERD was steady between pauses for both playing conditions, but was enhanced for shorter compared to longer pauses, suggesting that duet performance afforded a brain state of elevated action readiness.

Zamm et al.,  
2021b

EEG

Neural entrainment of duets was defined by EEG Power Spectral Density (PSD) of each pianist. EEG oscillation amplitude envelopes were calculated to evaluate the dynamics of the partners' neural responses over time. Pianists' diversity of melody duration and hence EEG samples was rectified using resampling. Inter-partner correlations of EEG amplitude envelopes (AECs) were assessed per dyad. Multiple regression models and ANOVAs were used to test differences of performance rate and spectral density, respectively. Pearson correlations were run on AEC values. Re-pairing was conducted to test for above-chance effects.

Spontaneous rate differences (solo) predicted duet signed asynchrony when not paced by an external cue. Enhanced spectral power at the duet frequency was associated with music partners' accuracy of synchronization. "Period coupling was observed in pianists' neural activity during duet performance" (Zamm et al., 2021b). Synchronization accuracy was best predicted by leader's spectral peaks.

*Appendix F: Relative frequencies of data analyses techniques*

**TABLE 7.** Relative frequencies of data analysis techniques for studies outside education sorted by biodata type and utility

<b>Biodata</b>	<b>Utility</b>	<b>Data analysis method</b>	<b>Count</b>	<b>RF per biodata utility</b>
EDA	Sympathetic autonomic arousal	Linear mixed model	2	14.29%
		One-way ANOVA	3	21.43%
		Mixed model ANOVA	2	14.29%
		Repeated measures ANOVA	1	7.14%
		ANCOVA	2	14.29%
		Paired t-test	3	21.43%
		GEE	1	7.14%
		Synchrony	Two-tailed t-test	3
	Synchrony	Pearson correlation	2	20.00%
		Linear mixed model	1	10.00%
		Cross-Recurrence Quantification Analysis (CRQA)	1	10.00%
		Repeated measures ANOVA	2	20.00%
		False pair surrogate analysis	1	10.00%

HRV-HF	Arousal	Mixed model ANOVA	1	25.00%
		Linear mixed model	1	25.00%
		Repeated-measures ANOVA	1	25.00%
		Mixed repeated measures ANOVA	1	25.00%
	Synchrony	Linear mixed model	1	100.00%
HRV-RMSSD	Arousal	Multiple linear regression	1	100.00%
HR	Arousal	One-way ANOVA	1	33.33%
		Mixed model ANOVA	1	33.33%
		Linear mixed model	1	33.33%
	Synchrony	Linear regression model	1	33.33%
		Multivariate Recurrence Quantification Analysis (MVRQA)	1	33.33%
		Independent sample t-test	1	33.33%
OO/ZM and CS activity	Valence expressions	One-way ANOVA	1	100.00%
	Valence synchrony	CRQA	1	20.00%
		False pair surrogate analysis	1	20.00%

		Pearson correlation	1	20.00%
		Repeated-measures ANOVA	1	20.00%
		Two-tailed t-test	1	20.00%
Eye movements	Gaze direction & movement	Mixed repeated measures ANOVA	6	16.67%
		Non-linear cross correlations (CC)	1	2.86%
		Repeated-measures ANOVA	2	5.71%
		One-way ANOVA	2	5.71%
		Linear regression model	1	2.86%
		Intraclass correlations (ICC)	1	2.86%
		Pearson correlation	4	11.43%
		Multilevel models	1	2.86%
		Mixed random effects models	1	2.86%
		Chi-square likelihood ratio	1	2.86%
		Two-way repeated measures ANOVA	2	5.71%

		Independent sample t-test	1	2.86%
		Mixed model ANOVA	4	11.43%
		(One-tailed) Pearson product-moment correlation	1	2.86%
		Linear mixed model	2	5.71%
		Mixed effects logistic regression models	1	2.86%
		Conversational analysis	1	2.86%
		Mixed effects repeated measures linear regression	1	2.86%
		Mann-Whitney U test	1	2.86%
		Cohen's d	1	2.86%
Pupillary dilation		Repeated-measures ANOVA	1	33.33%
		ANCOVA	1	33.33%
		Welch's t-test	1	33.33%
HbO & Hb concentration	Individual brain activity	Mixed model ANOVA	1	50.00%
		ANCOVA	1	50.00%
Magnetic fields from electrical brain activity	Synchrony	Weighted phase lag index (WPLI)	1	100.00%

Electrical brain activity	Synchrony	Multilevel models	1	2.94%
		Pearson correlation	2	5.88%
		Multiple linear regression	2	5.88%
		Linear mixed model	1	2.94%
		One-way repeated measures ANOVA	5	14.71%
		Two-way repeated measures ANOVA	4	11.76%
		Spearman's correlation	1	2.94%
		Paired t-tests	3	8.82%
		Mixed repeated measures ANOVA	2	5.88%
		Wilcoxon test	1	2.94%
		Two-way ANOVA	2	5.88%
		One-way ANOVA	3	8.82%
		Pearson product moment correlations	1	2.94%
		Linear trend analysis	1	2.94%
		Mixed model ANOVA	2	5.88%
General linear model	1	2.94%		

	Nonparametric bootstrap-based t-test	1	2.94%
	Weighted phase lag index (WPLI)	1	2.94%
Individual event related potential (ERP)	Wilcoxon test	2	14.29%
	Two-way repeated measures ANOVA	1	7.14%
	Mixed repeated measures ANOVA	4	28.57%
	One-way ANOVA	1	7.14%
	Mixed model ANOVA	1	7.14%
	Paired t-tests	1	7.14%
	Repeated measures general linear model	1	7.14%
	Linear mixed model	1	7.14%
	Threshold-free cluster enhancement	1	7.14%
	Spearman's correlation	1	7.14%
Individual brain activity	Mixed repeated measures ANOVA	3	37.50%
	ANCOVA	1	12.50%



		Pearson correlation	1	12.50%
		Wilcoxon test	1	12.50%
		Two-way repeated measures ANOVA	1	12.50%
		Mixed model ANOVA	1	12.50%
Haemodynamic response	Blood-oxygen-level dependent (BOLD) response	Mixed model ANOVA	2	50.00%
		Paired t-tests	1	25.00%
		General linear model	1	25.00%
Pre-ejection period (PEP)		Mixed model ANOVA	1	100.00%
Cardiac Output (CO)		Mixed model ANOVA	1	100.00%
Total Peripheral Resistance (TPR)		Mixed model ANOVA	1	100.00%
Respiration rate	Synchrony	Two-tailed t-test	1	100.00%
IBI	Synchrony	Cross correlation	1	25.00%
		CRQA	1	25.00%
		Cross-fuzzy entropy	1	25.00%
		Two-tailed t-test	1	25.00%
Embodied movements	Postural sway synchrony	Cross correlation	1	16.67%
		CRQA	1	16.67%

		Cross-fuzzy entropy	1	16.67%
	Finger movement synchrony	Repeated-measures ANOVA	1	16.67%
		Multiple linear regression	1	16.67%
	Hand movement synchrony	Multidimensional recurrence quantification analysis (MdRQA)	1	16.67%
fNIRS	Synchrony	Wavelet transform coherence (WTC)	3	18.75%
		Cross correlation	1	6.25%
		Granger causality analysis	1	6.25%
		One-way t-test	2	12.50%
		One-way ANOVA	1	6.25%
		Pearson correlation	1	6.25%
		Cohen's d	1	6.25%
		Mixed model ANOVA	1	6.25%
		Two-way ANOVA	1	6.25%
		Independent sample t-test	1	6.25%
		Chi-square test	1	6.25%
		Kendall's tau correlation	1	6.25%

		Fisher linear discrimination analysis	1	6.25%
Facial temperature	Arousal	Mixed repeated measures ANOVA	1	33.33%
		One-way ANOVA	1	33.33%
		Pearson correlation	1	33.33%