



The pivotal role of monitoring for collaborative problem solving seen in interaction, performance, and interpersonal physiology

Eetu Haataja¹ · Jonna Malmberg¹ · Muhterem Dindar¹ · Sanna Järvelä¹

Received: 7 September 2020 / Accepted: 9 September 2021 / Published online: 22 October 2021
© The Author(s) 2021

Abstract

Being aware of the progress towards one's goals is considered one of the main characteristics of the self-regulation process. This is also the case for collaborative problem solving, which invites group members to metacognitively monitor the progress with their goals and externalize it in social interactions while solving a problem. Monitoring challenges can activate group members to control the situation together, which can be seen as adjustments on different systemic levels (physiological, psychological, and interpersonal) of a collaborative group. This study examines how the pivotal role of monitoring for collaborative problem solving is reflected in interactions, performance, and interpersonal physiology. The study has foci in two central characteristics of monitoring interactions that facilitate groups' regulation in reaching their goals. First is valence of monitoring, indicating whether the group members think they are progressing towards their goal or not. Second is equality of participation in monitoring interactions between group members. Participants of the study were volunteering higher education students ($N=57$), randomly assigned to groups of three members whose collaborative task was to learn to run a business simulation. The collaborative task was video recorded, and the physiological arousal of each participant was recorded from their electrodermal activity. The results of the study suggest that both the valence and equality of participation are identifiable in monitoring interactions and they both positively predict groups' performance in the task. Equality of participation to monitoring was not related to the interpersonal physiology. However, valence of monitoring was related to interpersonal physiology in terms of physiological synchrony and arousal. The findings support the view that characteristics of monitoring interactions make a difference to task performance in collaborative problem solving and that interpersonal physiology relates to these characteristics.

Keywords Metacognitive monitoring · Self-regulated learning · Collaborative problem solving · Collaborative learning · Physiological arousal · Physiological synchrony

✉ Eetu Haataja
eetu.haataja@oulu.fi

¹ Faculty of Education, Learning and Educational Technology Research Unit (LET), University of Oulu, P.O. Box 2000, 90014 Oulu, Finland

Introduction

Collaborative learning and problem solving are increasingly used in today's education. Early research demonstrated that solving computer-based problems together can initiate learning through interaction (Roschelle and Teasley 1995). More detailed descriptions in different contexts have since been made to better explain what is important when students collaborate and interact to facilitate learning through problem solving. Currently, there is a wide range of studies showing that group members who are successful in their collaborations negotiate with each other (Hmelo-Silver and Barrows 2008) and reciprocally share how they are doing with the task (Rogat and Linnenbrink-Garcia 2011; Ucan and Webb 2015). In other words, they engage in metacognitive monitoring, which can lead to more profitable strategies in collaborations (Hadwin et al. 2018). Monitoring can make learners aware of whether the task is progressing as planned or if there is a need to control and make changes to the process (Azevedo 2014; Azevedo et al. 2010). When learners engage in monitoring interactions, expressing their views to each other and acknowledge that there is a challenge in terms of their learning process, it can invite socially shared regulation, consisting of negotiated and reciprocal regulatory processes such as planning, monitoring, and controlling among the group members (Hadwin et al. 2018; Malmberg et al. 2015).

Learners' monitoring interactions have a crucial impact on deciding which regulatory actions to take in order to reach task goals (Hadwin et al. 2018), because its valence signals whether the standards and goals set for task progress, understanding, and resources are being met or not (Azevedo et al. 2010; Azevedo 2014; Sobocinski et al. 2020). Monitoring interactions with positive valence suggest that things are on track and strategies are likely working. Monitoring interactions with negative valence raise the awareness that the group is off track and control of the strategies is likely needed. Although there is an emerging interest in studying group regulatory processes in relation to valence of monitoring interactions (Sobocinski et al. 2020), the knowledge on the issue is nascent. Specifically, the empirical link between the valence of monitoring interactions and group task performance is missing. Considering this, the current study investigates how valence of monitoring is seen in interaction and its relation to collaborative task performance.

In collaborative settings, metacognitive monitoring occurs as a co-constructed, transactive activity where group members engage in reciprocal turn-taking and respond to each other to discuss and evaluate their group progress, understanding, and resources. Several studies have underlined the importance of reciprocal interactions in facilitating effective monitoring in collaborative learning (Isohätälä et al. 2020; Rogat and Linnenbrink-Garcia 2011; Ucan and Webb 2015). However, it is also known that group members do not contribute to the collaborative discourse equally (Kapur et al. 2008; Koivuniemi et al. 2018). Specifically, there is limited knowledge on how equality of participation manifests in monitoring interactions during face-to-face collaboration and its impact on task outcomes. Drawing on this, the current study further explores equality of participation in monitoring interactions and its relation to collaborative task performance.

Studying metacognitive monitoring and regulation in collaborative learning is challenging because collaboration involves multiple agents interacting in relation to temporally unfolding collaboration events. Regulation in collaborative learning is a multifaceted phenomenon involving cognitive, motivational, emotional, and behavioral aspects (Hadwin et al. 2018), and it occurs as interplay on multiple systemic levels, including physiological,

psychological, and interpersonal (Reimann 2019; Volet et al. 2009). To date, studying and gathering evidence of this temporal, multifaceted, multilevel phenomenon has been difficult for researchers. In real-time face-to-face collaborative learning studies, in particular, video data have been the most dominant data source used to study monitoring (Malmberg et al. 2017). Yet, during the past few years, there has been an increase in the development of methods and tools that are viable in making metacognitive monitoring and its underlying conditions in groups visible (Järvelä et al. 2021). Due to technological development, there are increased opportunities to apply, for example, speech recognition (Amon et al. 2019), facial expression recognition (Taub et al. 2019), or physiological sensors (Dindar, Järvelä et al. 2020), that can, to some extent, reach monitoring and its conditions in real time. Although some of these methods hold benefits in collaborative settings, such as physiological measures' unobtrusiveness and intensive sampling, more empirical work is needed to show if and how they relate to processes relevant for socially shared regulation of learning (Hadwin et al. 2018), such as monitoring. Studying metacognitive processes in relation to physiological data might help overcome the aforementioned challenges in developing a more thorough understanding of how monitoring unfolds temporally and how it interplays with regulatory processes unfolding on the multiple systemic levels (Volet et al. 2009) of a collaborative group. Therefore, in addition to studying the characteristics of monitoring interaction (i.e., equality of participation and valence), this study utilizes electrodermal activity to explore whether these characteristics of monitoring as pivotal features of regulation are also reflected in interpersonal physiology.

Background

Collaboration is a coordinated and synchronous activity that is facilitated by continuous attempts to construct and maintain a shared conception of a problem (Roschelle and Teasley 1995). Therefore, collaborative learning can elicit students' individual learning processes but can also start learning processes on a social level (Dillenbourg 1999; Salomon and Perkins 1998). However, collaborative learning does not necessarily take an effective form just by putting individuals into a group (Kuhn 2015), but rather, it requires learners to metacognitively monitor and regulate their cognition, emotion, motivation, and behavior individually and as a group (Järvelä et al. 2018). Shared understanding and planning are important for solving problems together (Eichmann et al. 2019; Häkkinen et al. 2017), but learners also need to be aware of and monitor together how they are doing with their task and what needs to be changed when challenges occur (Ackerman and Thompson 2017; Hesse et al. 2015).

Monitoring together in collaborative learning

Monitoring is at the core of metacognition and regulation because it facilitates the comparison of the current state of the cognition with the standards set for the task (Winne and Hadwin 1998). Given that metacognitive monitoring lacks direct access to cognition, it often must rely on its products (Veenman et al. 2006; Winne 2018). In problem solving, these cognitive products can take the form of progress and performance regarding the task

at hand. Therefore, thinking about progress and performance in terms of a collaborative problem-solving task can be considered a metacognitive monitoring process (Ackerman and Thompson 2017; Clark and Dumas 2016; Rogat and Linnenbrink-Garcia 2011; Ucan and Webb 2015).

Monitoring serves as a base for controlling learning and problem solving if needed and is therefore important for successful problem-solving (Ackerman and Thompson 2017; Chang et al. 2017). This is especially the case when the task is considered to be complex (Dörner and Funke 2017; Greene and Azevedo 2009) because complex problem solving involves a high level of uncertainty, temporally unfolding events, and evaluation of the effectiveness of strategies (Dörner and Funke 2017). These characteristics call learners to monitor their knowledge, progress, and strategies. For example, Rudolph et al. (2017) studied the role of monitoring in complex problem solving and found it to be positively linked to success in solving the problem. Their results suggest that problem solving requires efficient self-regulation, and in particular, monitoring. The authors also stated that it would be especially relevant to study what follows when the monitoring of the students signals challenges in the process.

Monitoring also has a fundamental role in theories of regulation in collaborative learning (Hadwin et al. 2018). For group-level regulation (co-regulation and socially shared regulation) to manifest itself, it is crucial that metacognitive monitoring is communicated and negotiated in the interactions between group members (Hadwin et al. 2018; Malmberg et al. 2017). This is especially the case during task execution because it is important that all the participants in a group know how they are progressing with the task so that they can decide together on the efficient use of strategies. Due to that, the quality of monitoring seems to be strongly associated with high-quality cognitive activities of collaborative groups (Hurme et al. 2006; Khosa and Volet 2014). Näykki et al. (2017) studied the characteristics of university students' monitoring in collaborative learning and found that groups in which learners monitored their own and their peers' understanding throughout the tasks had an advantage in terms of their learning gains over the groups who did not. In the successful groups, learners were also actively involved in monitoring each other's task progress, task understanding, and task interests. Rogat and Linnenbrink-Garcia (2011) studied in detail the characteristics of high-quality monitoring in collaborative learning. Their results suggest that groups' high-quality monitoring asks for group members' equal contributions to monitoring. This view is supported by Ucan and Webb's (2015) study, in which reciprocal monitoring interactions facilitated a beneficial knowledge co-construction process within the groups. Reciprocal interactions and equal contributions to monitoring should also support the socially shared regulation of learning, which again, should contribute to collaborative learning (Järvelä et al. 2018; Saab 2012).

The importance of equality of participation in small groups has been acknowledged for decades for its apparent ability to predict differences in learning (Cohen 1994). In practice, equality of participation has often been operationalized with indices of statistical dispersion derived from either the number of task acts or the amount of talking time initiated by each group member (Bachour et al. 2010; Cohen and Roper 1972). In computer-supported collaborative learning environments, the length of the messages has also been used as a unit of analysis (Kapur et al. 2008; Strauß and Rummel 2021). Results from synchronous online collaboration studies (Kapur et al. 2008) suggest that equality of participation is a significant positive predictor of collaborative learning outcomes and that the level for this in a group

is quite stable. Inequal participation is hypothesized to leave little opportunity for different perspectives, strategies, and solutions to be shared and discussed, which might explain why equality of participation predicts learning outcomes.

Further, emerging evidence from studies in face-to-face collaborative learning settings suggests that productive collaboration involves distributed actions of regulation, social behaviors with positive valence, and forms of interaction, which support equal participation in a group (Pino-Pasternak et al. 2018). For example, Isohätälä et al. (2020) explored how actively students participate in cognitive and socio-emotional interactions during collaborative learning and what characterizes the moments when participation changes during transitions between the types of interaction. They categorized events from video data where all or only some of the students participated in the interaction. The qualitative analysis showed that monitoring task progress and challenges often occurred when a higher number of students participated during a transition from cognitive to socio-emotional interaction. Groups' metacognitive interaction about their performance seemed to co-occur with socio-emotional interaction and often involved increased participation in a group. The authors of the study hypothesize that these moments may help groups to coordinate and develop shared understandings of their learning and promote a positive socio-emotional climate. They also conclude that future studies should investigate how patterns of participation and types of social interaction affect regulation and outcomes of collaborative learning.

To conclude, monitoring has a special role in collaborative problem solving, because in order to collaborate, students should share their views about the task and progress towards their goal. This is also a prerequisite for socially shared regulation (Hadwin et al. 2018). However, empirical evidence about the importance of equality of participation in monitoring interactions for the group performance in face-to-face collaboration is still limited.

Valence of monitoring makes a difference to regulation

Although the aforementioned earlier research has revealed some of the characteristics of monitoring interaction, the valence of monitoring has so far gained little attention in the analysis of regulatory processes (Azevedo 2014). Because monitoring is an activity that raises participants' awareness of the current situation, it can signal two completely different states of affairs with regard to valence.

Monitoring has different valences depending on if there is a discrepancy or not with the standards and goals set. On the one hand, monitoring can signal that there is a challenge and that the cognitive activity process is not developing as it should, which requires students to control their cognitive processes. This has been called *monitoring with negative valence* (Azevedo et al. 2010; Sobocinski et al. 2020). For example, a group of students might notice that their simulation task is not progressing as it should and that they are not performing well. This serves as a sign that something needs to be changed (metacognitive control) in their cognitive processes, and they can start to discuss alternative strategies to use in the simulation. They might end up reading the instructions again or adjusting the simulation for better success. On the other hand, monitoring may also indicate that the process is currently on track, which is called *monitoring with positive valence*. This suggests that the learning is progressing as it should and that no significant adjustments are needed regarding the cogni-

tive processes of the group. Though the valence of monitoring is likely linked to emotional valence, it has a different meaning because emotional valence refers to the pleasantness of a feeling (Bradley and Lang 1994), while valence of monitoring refers to the evaluation of a cognitive process in relation to set standards (Azevedo and Witherspoon 2009). It is, for example, possible for a learner to monitor their own task progress with negative valence but still express and experience emotions with positive valence.

Monitoring with negative valence makes a difference for regulation in collaborative learning because it is likely to start the control processes of regulation. This means that the group has to activate and show an effort to change the course of the learning process (Hadwin et al. 2018), which should also be reflected as the interplay of adjustments on different systemic levels, including physiological, psychological, and interpersonal (Volet et al. 2009), of a collaborating group. Therefore, the distinction of valence as a characteristic in monitoring is important for understanding how students regulate their learning processes together (Sobocinski et al. 2020). The valence of monitoring has been acknowledged in regulation research on individual learners (Azevedo et al. 2010). However, there are only a few studies that have investigated it in collaborative learning contexts (Sobocinski et al. 2020; Koivuniemi et al. 2018). These studies suggest that becoming aware and acknowledging the challenges in approaching the goals is important for efficient collaboration and seems to be linked to collaboration dynamics such as equality of participation. Evidence for the relation between valence of monitoring interactions, collaborative task performance and marks of regulation on different systemic levels of collaborative group is still scarce.

Interpersonal physiology reflecting monitoring in collaboration

Regulation of collaborative learning is not static but rather unfolds over time. While it is metacognitively grounded (Hadwin et al. 2018), it activates adjustments on different systemic levels (physiological, psychological, and interpersonal) based on perceived or anticipated changes in the demands and conditions of the environment to meet goals (Volet et al. 2009). This means that recognizing a need for adjustment through monitoring is pivotal and should therefore be reflected in social interactions (Näykki et al. 2017) and physiology (Ullsperger et al. 2014). To have a comprehensive understanding on the nature of monitoring and regulation in collaborative learning, it is important to investigate the interplay between the aforementioned systemic levels (Reimann 2019).

So far, it has been challenging to find scalable measures to explore these processes (Hurme and Järvelä 2005). Approaching regulation through triangulation with multiple methods and data channels has been proposed as one potential solution for the challenge (Karabenick and Zusho 2015; Järvelä et al. 2019). Recent developments in technology have opened new possibilities for studying monitoring with temporal data. Still, robust measures in individual learning settings, such as computer log events (Winne and Jamieson-Noel 2002) or a think-aloud protocol (Greene et al. 2012), are not necessarily well suited to face-to-face collaborative tasks. This is because the number of log events in face-to-face collaborative settings is often low, and traditional thinking aloud directed toward the researcher cannot be carried out normally when students take turns discussing the problem with one another.

Recently, it has become possible for learning researchers to utilize data modalities that have previously only been available in highly controlled laboratory settings (Järvelä et al. 2019; Reimann et al. 2014). For example, physiological measures provide intensive temporal data linked to cognitive and emotional processes during cognitive tasks (Efklides et al. 2018; Kreibig and Gendolla 2014; Strain et al. 2013), which means they have potential in studying learning processes. Autonomic nervous system (ANS) activity does reflect cognitive processes (Critchley et al. 2013), and its measures have been used in psychological research for years to study physiological arousal in relation to the emotion, cognition, and behavior of individuals. For example, there have been attempts to find an optimal level of physiological arousal for performance (Stennett 1957; Yerkes and Dodson 1908), and empirical findings suggest that physiological arousal is linked to learning outcomes in real collaborative classroom environments (Pijeira-Díaz et al. 2018).

Research also suggests that physiological arousal relates to processes relevant to the regulation of learning (Malmberg et al. 2019). Being a goal-targeted activity, regulation of learning is likely to be reflected in ANS, which has the purpose of regulating bodily functions to meet the interpreted situational demands. For example, in addition to having cognitive and affective effects, monitoring problems in task performance (monitoring with negative valence) activates ANS, providing the somatic basis for behavioral adaptations (Ullsperger et al. 2014). Recently proposed allostatic models of regulation and adaptation of physiology also suggest that changes in ANS are not just reactions to events, but reflect individuals' and groups' predicted demands (physical, cognitive, or social) in the situation (Blair and Raver 2015; Kreibig and Gendolla 2014; Saxbe et al. 2020; Sterling 2012). This means that changes in arousal also reflect the predictions of the coming events and required actions in relation to individuals' or groups' goals.

In addition to investigating individuals' physiological activity in collaborative settings, a growing interest has risen to study the relationship between people's physiological dynamics referred as interpersonal physiology (Palumbo et al. 2017). Interpersonal autonomic physiology studies temporal interactions in ANS between multiple people. These interactions have been linked to several social cognitive and emotional phenomena relevant for collaboration, such as shared understanding (Järvelä et al. 2014), empathy (Marci et al. 2007), and emotional contagion (Pijeira-Díaz et al. 2019). However, the role of interdependence between the participants' physiological processes for social coordination remains unclear. On the one hand it seems to facilitate common social and affective space for collaboration (Cornejo et al. 2017; Danyluck and Page-Gould 2019) and on the other hand in some contexts it seems to reflect co-dysregulation in face of challenges (Saxbe et al. 2020). It has been hypothesized that these dynamics in interpersonal physiology aim towards stability through change, meaning continual predictive adjustment of multiple physiological systems to maintain homeostatic balance in a group (Saxbe et al. 2020).

Different concepts and computational procedures have been used to describe interpersonal physiology (Palumbo et al. 2017). In practice, most of the indices reveal how much interdependent or associated activity there is between the participants' physiological processes. In this study, this interdependence is referred to as physiological synchrony.

One of the potential ANS measures to inspect interpersonal physiology during temporally unfolding collaborative events is electrodermal activity (Ahonen et al. 2018; Schneider et al. 2020). Electrodermal activity (EDA) indicates the sympathetic "Fight or Flight" activity of ANS, which especially has been hypothesized to prepare the individual to face chal-

lenges identified in the situation (Dawson et al. 2017). Though EDA has often been linked with emotional reactions, it is also considered to relate to metacognitive processes, such as the monitoring of task performance (Ahonen et al. 2018; Hajcak et al. 2003; Ullsperger et al. 2014) and feelings relating to difficulty and effort (Efklides et al. 2006).

Ahonen et al. (2018), studying collaborative learning in coding with dyads, found that the amplitude of EDA responses contains information about collaboration dynamics in relation to the role participants have in collaboration. They also found preliminary evidence suggesting that EDA provides insight into the valence of pivotal task events. In the study, students who ran the code showed significantly higher EDA responses before running unsuccessful code than those running successful code. Authors of the study frame this as a link to emotional valence, but further explain it with participants' mental model of the code goodness, which could be considered to rely on metacognition. The fact that the arousal arose before testing the non-successful code could mean that prediction of the coming performance and adjustments needed after that was made beforehand, which would align with allostatic models of physiological regulation (Sterling 2012).

Based on prior research, Malmberg et al. (2019) investigated how physiological arousal relates to monitoring events during collaborative exams. They found that, at a session level, the frequency of monitoring utterances was strongly related to physiological arousal, seen as EDA peaks.

Recently, Schneider et al. (2020) found physiological synchrony to be positively linked with interaction characteristics relevant to regulation in collaborative learning, such as reciprocal interactions and dialogue management. They also found that groups that managed to reach a consensus decreased their physiological synchrony over time. Further, group effort demands might play a role in this, as Dindar, Järvelä et al. (2020) found monitoring of mental effort to be positively related to physiological synchrony during collaboration. Physiological synchrony has also been found to reflect the similarity between subjective self-reported evaluations of cognition between group members in collaboration (Dindar, Malmberg et al. 2020). Sobocinski et al. (2020) studied the valence of monitoring and the following reactions by collaborative group members as markers of adaptation during a collaborative physics exam. They found that groups that monitored more with positive valence also tended to show more transitions in their group-level physiological states, as derived with vector quantization from their heart rate signals. Though the strength of the correlation was limited, the relations between monitoring and interpersonal physiology seem to have potential to be explored further.

Frequently found relations between monitoring interactions and physiology suggest that the hypothesized interplay (Reimann 2019; Volet et al. 2009) between systemic levels of physiology and social interaction in the regulation of collaborative learning exists. However, the findings of interpersonal physiology in relation to monitoring expressed in collaborative learning are still somewhat inconsistent, and there seem to be group- (Haataja et al. 2018) and task-dependent variations (Dindar et al. 2019). It remains unclear whether interpersonal physiology is more likely to reflect the dynamics of roles and equality of participation in collaborative interactions (Ahonen et al. 2018; Schneider et al. 2020) or the recognized regulatory need for adjustments (monitoring with negative valence) in a group (Sobocinski et al. 2020; Volet et al. 2009). The valence of monitoring should be pivotal for the following regulation process (Azevedo 2014), making it important to study valence in relation to monitoring interactions, arousal, and physiological

synchrony. More research is needed to better understand how physiological synchrony, which in some contexts seems to facilitate adaptation (Feldman 2007) but in others is coupled with challenges (Malmberg et al. 2019; Saxbe et al. 2020), relates to monitoring and regulation in collaborative learning.

Aims

This study acknowledges the pivotal role of monitoring for collaborative problem solving and hypothesizes that it can be evidenced in interactions, performance, and interpersonal physiology. Therefore, the study aims to investigate how valence and equality of participation in monitoring interactions relates to collaborative problem-solving performance, physiological arousal, and physiological synchrony.

The specific research questions and hypotheses are as follows:

RQ1. Do valence and equality of participation in monitoring interactions relate to collaborative task performance?

H_1 . Valence of monitoring interactions predicts collaborative task performance.

H_2 . Equality of participation in monitoring interactions predicts collaborative task performance.

RQ2. Do physiological arousal and physiological synchrony relate to valence of monitoring interactions?

H_3 . Physiological arousal predicts valence of monitoring interactions.

H_4 . Physiological synchrony predicts valence of monitoring interactions.

RQ3. Do physiological arousal and synchrony relate to equality of participation in monitoring interactions?

H_5 . Physiological arousal is positively linked to equality of participation in monitoring interactions.

H_6 . Physiological synchrony is positively linked to equality of participation in monitoring interactions.

Methodology

Participants and context

The subjects of the study were volunteer university students ($N=77$, $M_{\text{age}}=27.84$, $SD=5.51$, Male=33) who were randomly assigned to groups of three. In total, data were collected from 26 groups. However, seven groups were excluded from the dataset, either due to a participant leaving the task before it was over or because of the poor quality of a participant's EDA data. One of the excluded groups included only 2 participants. Therefore, the final dataset includes data from 19 groups ($N=57$, $M_{\text{age}}=27.29$, $SD=4.89$, Male=25).

The collaborative task was a business simulation in which participants were required to run a shirt production company (Danner et al. 2011). The simulation's goal was to increase the value of the company as much as possible. This depended on the relationships between 24 variables (e.g., employee wages, storage costs, store locations, advertisement costs, shirt price), which could be adjusted by the group members. The simulation included two distinct

phases: exploration and performance. The current study concentrates on the exploration phase, wherein the participants ran the simulation for six simulated months with the aim of learning how the system worked, as well as understanding how adjusting the variables would affect the value of the company. Adjustments for each month were followed by a transition to the next month, after which the participants could immediately see the company's current value based on the decisions and adjustments made during the previous month. This study focuses on one-minute episodes that took place immediately after each transition as those were considered to be pivoting parts of the simulation where the students would likely spontaneously monitor their problem-solving processes. The participants were not prompted on how to interact after the transitions.

Data

The data were collected in a classroom-like research infrastructure specifically designed for collaborative learning. The research space was separated into three rooms with sound-proof walls, which made it possible to collect data from three groups simultaneously. First, participants were asked to fill in consent forms. Second, for EDA recording, the researchers attached Shimmer 3GSR+sensors (Burns et al. 2010) to the participants' non-dominant hands so that the gel electrodes were placed on the thenar and hypothenar eminences on their palms (Dawson et al. 2017). Third, the researchers introduced the participants to their randomly assigned group members and guided them to the data collection room. In the room, the group members were seated at a table with a desktop computer. Fourth, the researcher read the task instruction and then left the participants to complete the simulation collaboratively. On average, it took 41 min and 14 s for the groups to complete the exploration phase of the task ($SD=17$ min 5 s).

The data of the current study consist of video recordings, physiological data, and task performance measures for 19 groups. Altogether, the video recordings include 27 h and 3 min of data. However, this study concentrates on the exploration phase of the simulation, and the (1 min) episodes following the transitions between the simulated months altogether provide 114 one-minute video recordings. The original EDA data consist of 57 recordings with a sampling frequency of 128 Hz. Log data include the timestamps of the transition episodes and the company values before and after the transitions in the simulation.

Analysis of the performance

Since the goal of the simulation is to maximize the company value, the changes in the company value after each simulated month are considered as performance indicators for the Tailorshop simulation (Danner et al. 2011). Although a trend in the change seems to be the most reliable measure of individuals' dynamic decision-making skill, this study was more interested in the specific moments of collaboration in the simulation interactions. Therefore, a change in the company value ($M = -12373.26$, $SD=17493.20$) for the month, following the transition minute, was used as an indicator of task performance for each month. Because the change is presented as a simulated monetary unit, being possibly difficult to interpret, the values were standardized to help interpretation of the results.

Analysis of the video data

The analysis of the current study concentrates on the (1 min) time segments when the students transitioned to the subsequent month in terms of exploring the simulation and, therefore, when they had the chance to monitor their progress with the task. Monitoring utterances and subjects participating were identified and coded from the transition (1 min) episodes in the video, and was based on prior coding schemes used for monitoring coding (Haataja et al. 2018; Whitebread et al. 2009). Verbalizations related to the ongoing on-task assessment of the quality of the task performance, understanding, and the degree to which performance was progressing towards a desired goal were considered as monitoring interactions (Whitebread et al. 2009). After this, a coding scheme on the valence of monitoring was developed based on theory (Azevedo 2014; Azevedo and Witherspoon 2009) and prior research (Sobocinski et al. 2020).

The valence of monitoring included three categories: monitoring with positive, neutral, and negative valence. Defining characters for monitoring with positive valence were that the students considered that the problem-solving process was progressing towards their goal or that they acknowledged knowing how the simulation worked. Monitoring with neutral valence was more about students verbally pointing out the current state of the problem-solving process without considering if it was progressing towards the goal or not. Monitoring with negative valence signaled that the process of problem solving was not advancing towards the goal or that the students did not understand how the system operated. Table 1 explains the key differences between each of these categories in further detail.

In the current study, verbal expressions of monitoring were considered as participation in monitoring and, therefore, equality of participation into monitoring was operationalized from the variation in summed durations of expressed monitoring utterances between the students. In practice, equality of participation in monitoring interactions between the students was calculated for each simulation transition minute as the standard deviation (*SD*) of the three members' monitoring proportions (the duration of the monitoring by a member as a proportion of the total duration of the monitoring) within each group (Kapur et al. 2008). This would have meant that a lower *SD* indicated more equal participation in the monitor-

Table 1 Description of coding categories for the valence of monitoring interactions

Code	Description	Examples
Monitoring with positive valence	Utterances stating that the group is advancing towards their goal or that they understand something concerning the task. Positive comparison to the previous state of the task.	<i>"Our company value increased"</i> <i>"We are doing well!"</i> <i>"We know that this relates to ..."</i> <i>"We made more profit!"</i>
Monitoring with neutral valence	Utterances pointing out the current state of the task without considering if it is progressing towards the goal or not. No comparison to the previous state of the task.	<i>"We are now in month 5"</i> <i>"Our bank account value is ..."</i> <i>"We made 1500 shirts"</i>
Monitoring with negative valence	Utterances stating that the group is not advancing towards their goal or that they do not understand something concerning the task. Negative comparison to the previous state of the task.	<i>"I don't understand how this works"</i> <i>"We are going bankrupt"</i> <i>"We are making less money"</i> <i>"Our value decreased"</i>

ing interactions between the students in a group. Therefore, the variable was reversed for easier interpretation. First, each value was subtracted from the theoretical maximum of the proportional *SD* that three group members could have, referring to a case of least equal participation where one student would be responsible for all the monitoring. Second, resulting values were then divided by theoretical maximum to adjust the range of the scale to 0–1. In the resulting variable, value 0 indicates the lowest possible equality of participation where one student is responsible for all the monitoring in a group, and value 1 indicates the highest possible equality of participation where all three group members attend to exactly the same amount.

Reliability of the video data coding

The reliability analysis was performed on two levels: first, (a) to identify monitoring utterances at the student level and second, (b) to reveal the valence of the monitoring utterances at the student level. Both the monitoring utterances and the valence of the monitoring utterances were coded by two researchers. After the principal coder had completed the coding, the interrater coder assessed 20 % of the data from randomly selected one-minute episodes, altogether consisting of 23 min of video from 15 groups. The interrater coder was asked to demarcate monitoring utterances and the valence of the monitoring utterances from the students' verbal interactions (i.e., when the monitoring began and ended) and to name the type of valence (negative, neutral, or positive). In the first round of the reliability coding, four random episodes were selected. In this phase of the analysis, the rules and criteria for identifying monitoring utterances with valence were also clarified. In the second round of the analysis, 10 episodes from 23 randomly assigned episodes were analyzed, yielding agreement between the coders with a kappa value of 0.59. After negotiating and specifying the coding schema, the rest of the episodes were coded, and the overall interrater reliability kappa value for 168 monitoring utterances was 0.64, which can be considered as demonstrating a substantial level of agreement. The resulting video codes used in the later steps of the analysis were all coded by a principal coder who was more experienced with these types of data. The duration of each valence of monitoring for each one-minute episode and for each student was retrieved from Observer XT12.5 software.

Analysis of the physiological data

The EDA signal was pre-processed with a previously used approach (see, e.g., Di Lascio et al. 2018). First, the EDA data were visually inspected for clear signs of movement artifacts (e.g., lost electrode contact seen as a drop in the signal), and the artifacts found were manually corrected with Ledalab toolbox. Second, the signals were downsampled to 4 Hz with the purpose of making long recordings computable for recurrence quantification analysis. Third, signals were standardized in order to make them comparable with each other (Ben-Shakhar 1985). Standardization has been successfully used for EDA (Ben-Shakhar 1985; Dawson et al. 2017) and is also suggested for recurrence-based analysis to ensure that its measures are based on the sequential similarity of the time series (Wallot and Leonardi 2018). Fourth, the signal was decomposed with Ledalab continuous decomposition analysis, including the adaptive smoothing of the signal (Benedek and Kaernbach 2010), which resulted in a rapidly changing phasic signal component (seen as peaks in the original signal)

and a more slowly changing tonic signal component (seen as a base level in the original signal).

Physiological arousal was measured by the change in the tonic signal component of EDA, which has been suggested to be applicable to unfolding events such as collaborative learning (Mendes 2009). In practice, physiological arousal was derived from the slope of the tonic EDA component by fitting a linear model for each signal in each transition minute. Because this gave values for the individuals in a group, an aggregation for the group level had to be carried out. For this, the best unbiased linear predictor method in the MicroMacro R package was used (Croon and van Veldhoven 2007; Lu et al. 2017). This resulted in slope values for each group for each transition, reflecting physiological arousal at a group level. Positive values signaled an increase and negative values signaled a decrease in the group's arousal.

Physiological synchrony aims to reveal interdependence in physiology between the individuals. In this study, the explored time windows for interdependence were quite short (1 min), and therefore, the phasic signal component of EDA was used as the signal for calculation (Mendes 2009). Multidimensional recurrence quantification analysis (MdrQA) is one of the few methods that can quantify the synchrony between more than two signals

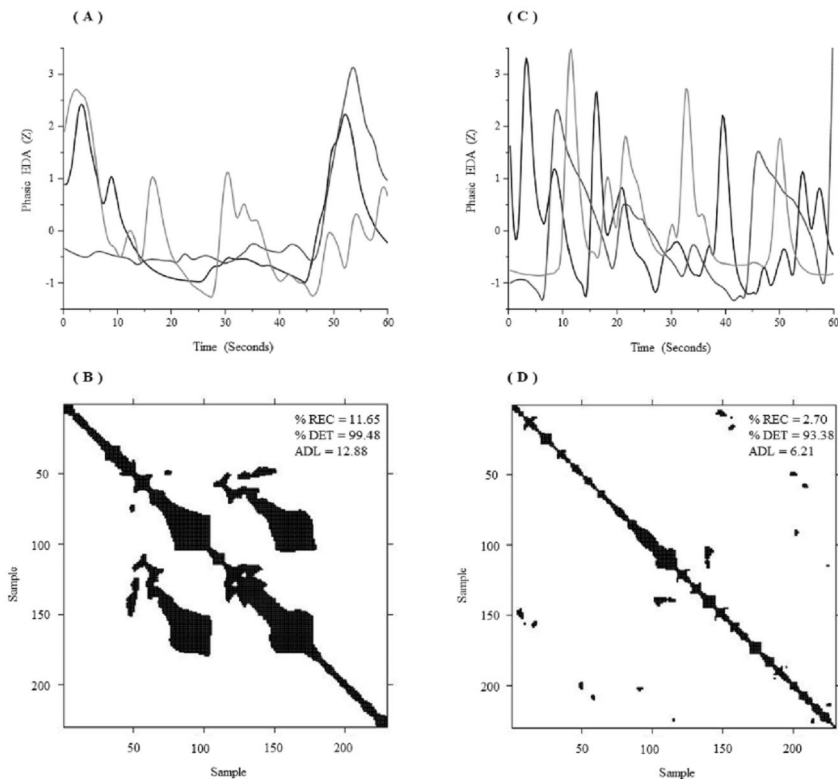


Fig. 1 Example of EDA signals (A,C) and the resulting recurrence plots (B,D) of one-minute episodes with high synchrony (A,B) and low synchrony (C,D)

(Wallot et al. 2016), and therefore, it was used to quantify the physiological synchrony between the students. MdrQA is a nonlinear time series analysis method that allows the investigation of synchrony patterns between two or more time series, which do not necessarily have to be stationary. MdrQA statistics are based on recurrence plots, which graphically display the dynamics of a multidimensional phase space of a system, such as a collaborating group (see the example in Fig. 1). In general, MdrQA statistics derived from the plot can be considered to indicate synchrony between the signals. For example, percent recurrence (%REC) indicates how many individual elements between three signals are shared, percent determinism (%DET) is the degree to which elements between the three signals repeat in terms of larger connected synchrony patterns, and average diagonal line (ADL) indicates the average size of the repeated synchrony patterns (Wallot et al. 2016).

The parameters for running the MdrQA analysis were decided based on the suggestions in the RQA literature (Wallot et al. 2016). First, the delay (DEL) parameter was estimated using the average mutual information function for each individual EDA signal. Second, the false nearest neighbor function was used for each EDA signal to estimate the embedding dimension parameter. With both functions, the first local minimum was determined for each signal and then averaged and rounded up for all the signals. In this case, the resulting values were divided by two as the signals were embedded together in the MdrQA. This means that not all the dimensions had to be reconstructed by time-delayed embedding because they were available as separately measured signals (Wallot et al. 2016). The parameters used were delay=10 and embedding=2. The radius parameter was set to 0.30, which kept the mean of percent recurrence close to the suggested 5 % (Wallot and Leonardi 2018).

To verify that the synchrony occurred due to collaboration and not due to chance or task constraints, false groups were formed so that phasic EDA signals were randomly matched with participants from other groups. In practice, each signal was randomly assigned to a false group so that none of the signals stayed in the same group with the original group members. Then, MdrQA was performed for the false groups to compare them with the real ones. The Mann-Whitney U test showed a significant difference between the real and the false groups for percent recurrence (%REC, $U=5\,385.5$, $n_1 = 112$, $n_2 = 112$, $z=-2.032$, $p=.042$), percent determinism (%DET, $U=5\,116$, $n_1 = 112$, $n_2 = 112$, $z=-$, $p=.010$) and average diagonal line length (ADL, $U=5\,407$, $n_1 = 112$, $n_2 = 112$, $z=-1.988$, $p=.047$). However, for maximum diagonal line length (MDL, $U=6\,306$, $n_1 = 112$, $n_2 = 112$, $z=-0.159$, $p=.874$) and percent lamarinity (%LAM, $U=5428.5$, $n_1 = 112$, $n_2 = 112$, $z=-1.944$, $p=.052$), no significant difference was found, and therefore, these were excluded from the later phases of analysis.

In order to use all the observed values from transition moments for each group in the analysis, the effects of repeated measures and possible serial dependency had to be taken into account in the statistical analysis. Generalized estimating equations (GEE) with a robust covariance estimator were used to model the standardized independent variables as predictors of task performance, equality of participation in monitoring, and valence of monitoring interactions. GEE attempts to accommodate the covariance that exists between the observations (i.e., repeated observations or clusters) and yields regression coefficient estimates with standard error estimates being corrected for nested or repeated types of data (McNeish et al. 2017). Because the dependent variables were continuous, normal, gamma, and inverse gaussian distributions with logarithmic and identity link functions were tested to select the best fitting working correlation matrix indicated by the lowest quasi-likelihood under the independence criterion (QIC) value (Garson 2012). Independent variables were

standardized before being included in the model. The independent covariates for each model were included based on the lowest values of corrected quasi-likelihood under the independence model criterion (QICC), indicating the best fit for the model (Cui and Qian 2007). The results of each model with the working correlation matrix, distribution, and link functions are reported in each table. Variance inflation factor values for MdRQA measures ($VIF < 2.2$) and different valences of monitoring ($VIF < 1.2$) were below commonly used cutoff limits (4–10, O'Brien 2007) suggesting that multicollinearity does not exist between the variables.

Results

Descriptive statistics

Descriptive statistics for the monitoring utterances with different valences are introduced as durations and frequencies in Table 2. Most of the monitoring had a neutral valence ($M = 16.88$ s, $SD = 11.45$ s). Monitoring with positive valence occurred the least ($M = 5.45$ s, $SD = 5.52$ s). On average, half of the duration of the transition minutes included monitoring interactions (see Table 2).

The Friedman test was run to see if there were differences between six subsequent transitions in terms of the duration of the monitoring and the three different valence categories. Durations of monitoring with neutral valence ($\chi^2(5) = 2.81$, $p = .730$) and with positive valence ($\chi^2(5) = 6.85$, $p = .232$), and monitoring overall, showed no differences between transitions. Only monitoring with negative valence showed a significant difference between transition episodes ($\chi^2(5) = 17.80$, $p = .003$). However, a post hoc comparison of Wilcoxon signed-rank tests with Bonferroni correction could not identify significant differences in pair-wise comparisons between the transitions. Variance partitioning coefficients suggested that differences between groups explained 21 % of the variance in all monitoring, 12 % in monitoring with negative valence, 27 % in monitoring with neutral valence, and 4 % in monitoring with positive valence. This supported the use of GEE for the later steps of the analysis.

The values for equality of participation varied from 0 to 1, with a mean of 0.52 and a SD of 0.20. A repeated measures ANOVA showed no significant difference between the means of the six subsequent transition episodes ($F(5, 90) = 1.72$, $p = .14$), which would suggest that on average there was no change or trend in terms of how equally the students took part in monitoring interactions during the simulation. The variance partitioning coefficient sug-

Table 2 Monitoring interactions with different valences for all the transitions

Type of monitoring	Duration (s)						Frequency					
	Min	Max	<i>M</i>	<i>SD</i>	Sum	%	Min	Max	<i>M</i>	<i>SD</i>	Sum	%
Monitoring with negative valence	0	38.29	9.06	7.76	1015.28	29	0	7	2.3	1.58	258	31
Monitoring with neutral valence	0	47.55	16.88	11.45	1890.28	54	0	11	3.63	2.29	407	49
Monitoring with positive valence	0	21.19	5.45	5.52	610.05	17	0	6	1.48	1.46	166	20
All monitoring	0	51.8	31.39	12.81	3515.6	100	0	16	7.42	3.01	831	100

gested that 27% of the variance in equality of participation in monitoring was explained by the group. This is to say, there were consistent differences between the groups (see Fig. 2) in terms of how equally the monitoring interactions were distributed between the group members, which also supported use of GEE for the later analysis.

RQ1. Do valence of monitoring interactions and equality of participation in monitoring interactions relate to collaborative task performance?

Equality of participation in monitoring and monitoring with negative, neutral, and positive valence (duration) were used as independent variables to fit a GEE model predicting task performance. Based on the QIC and QICC, the best fitting model predicting task performance included equality of participation in monitoring and monitoring with positive valence as significant positive predictors (see Table 3). Intercept made the fit of the model worse and was therefore excluded from it. This result means that higher equality of participation in monitoring interactions and more monitoring with positive valence predicted better task performance following the transition episodes.

RQ2. Do physiological arousal and physiological synchrony relate to valence of monitoring interactions?

The slope of the tonic EDA signal, representing physiological arousal and %REC, %DET, and ADL, representing physiological synchrony, were used to fit models predicting each valence category of monitoring interactions. The GEE model output for monitoring with negative, neutral, and positive valence shows that physiological arousal and physiological synchrony were both related to the valence of monitoring interactions (see Table 4). First, the model predicting monitoring with negative valence included the tonic EDA slope and %DET as significant positive predictors, meaning that increase in physiological arousal and higher physiological synchrony were related to more monitoring interactions with negative valence. Second, the model predicting monitoring with neutral valence included the tonic EDA slope as a significant negative predictor, meaning that an increase in arousal was related to less monitoring interactions with neutral valence. Third, the model predicting monitoring with positive valence indicates that %REC and %ADL were significant negative predictors, meaning that higher physiological synchrony was related to less monitoring with positive valence.

RQ3. Do physiological arousal and synchrony relate to equality of participation in monitoring interactions?

Fig. 2 Equality of participation presented for each group during each transition episode

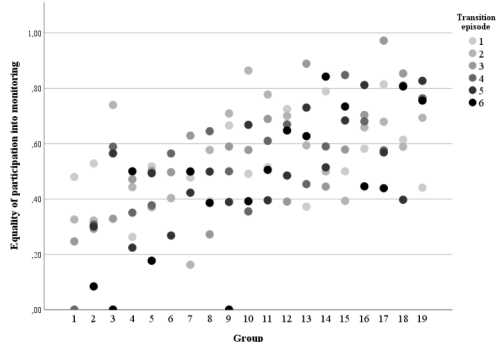


Table 3 GEE model output for following performance change

Dependent variable	Following performance change (standardized)																	
Covariance Structure		Exchangeable																
Distribution	Normal with identity link																	
	Model 0			Model 1			Model 2			Model 3			Model 4			Model 5		
Covariate	β (SE)	Wald χ^2	95% CI	β (SE)	Wald χ^2	95% CI	β (SE)	Wald χ^2	95% CI	β (SE)	Wald χ^2	95% CI	β (SE)	Wald χ^2	95% CI	β (SE)	Wald χ^2	95% CI
Intercept	0.01 (0.17)	0.001	[-0.34, 0.35]															
Equality of participation in monitoring				0.15 (0.06)	5.35*	[0.02, 0.27]	0.17 (0.07)	5.24*	[0.02, 0.31]	0.14 (0.07)	4.98*	[0.02, 0.27]	0.16 (0.06)	7.22**	[0.04, 0.27]	0.17 (0.07)	6.93**	[0.04, 0.30]
Monitoring with positive valence													0.19 (0.06)	8.52**	[0.06, 0.31]	0.17 (0.07)	5.29*	[0.02, 0.30]
Monitoring with neutral valence										-0.10 (0.08)	1.61	[-0.25, 0.05]				-0.09 (0.09)	0.94	[-0.26, 0.09]
Monitoring with negative valence							-0.07 (0.07)	1.01	[-0.22, 0.07]							-0.08 (0.06)	1.58	[-0.20, 0.04]
Model criteria																		
QIC	97.76		88.30		87.26		91.40		82.61		85.27							
QICC	94.00		89.65		89.40		93.68		85.12		89.25							

Note. * $p < .05$, ** $p < .01$, QIC = Quasi-likelihood under the independence criterion (lower is a better fit), QICC = Corrected quasi-likelihood under the independence model criterion

Note * $p < .05$, ** $p < .01$, QIC = Quasi-likelihood under the independence criterion (lower is a better fit), QICC = Corrected quasi-likelihood under the independence model criterion

Table 4 GEE model output for monitoring with negative, neutral, and positive valence

Dependent variable	Monitoring with negative valence			Monitoring with neutral valence			Monitoring with positive valence		
Covariance Structure	First order autoregressive			Unstructured			First order autoregressive		
Distribution	Gamma with identity link			Gamma with identity link			Gamma with identity link		
Covariate	β (SE)	Wald χ^2	95% CI	β (SE)	Wald χ^2	95% CI	β (SE)	Wald χ^2	95% CI
Intercept	8.93 (0.86)	108.21**	[7.24, 10.61]	16.55 (1.29)	165.36**	[14.03, 19.07]	5.60 (0.63)	79.63**	[4.37, 6.83]
Tonic EDA slope	1.42 (0.20)	51.27**	[1.04, 1.81]	-3.40 (0.26)	165.35**	[-3.91, -2.88]			
REC%							-0.51 (0.06)	79.61**	[-0.62, -0.40]
DET%	1.31 (0.34)	14.60**	[0.64, 1.98]						
ADL							-0.86 (0.10)	79.60**	[-1.05, -0.67]
Model criteria									
QIC		286.43			191.80			504.24	
QICC		290.95			194.85			508.45	

Note. * $p < .05$, ** $p < .01$, QIC = Quasi-likelihood under the independence criterion (lower is a better fit), QICC = Corrected quasi-likelihood under the independence model criterion

The slope of the tonic EDA signal, representing physiological arousal, and %REC, %DET, and ADL, representing physiological synchrony, were used as independent variables to predict equality of participation in monitoring interactions. None of the independent physiological data variables improved the fit of the GEE model for predicting equality of participation in monitoring interactions. The best fitting model therefore only included the intercept ($\beta = 0.241$, $\chi^2 = 333.72$, $p < .001$, 95 % CI [0.22, 0.27]), which suggests that neither arousal nor physiological synchrony has the potential to predict equal participation in monitoring interactions.

Discussion

Research into regulation in learning has emphasized the need to differentiate the characteristics of monitoring in different settings (Azevedo 2014). Equality of participation (Isohäätä et al. 2017; Rogat and Linnenbrink-Garcia 2011) and the valence of monitoring interactions (Sobocinski et al. 2020) have been considered as important characteristics for successful collaborative learning, but few studies have examined these systematically. Theories of regulation in collaborative learning have also emphasized the multifaceted (Hadwin et al. 2018) and multilevel (Volet et al. 2009) nature of regulation in collaborative learning. Prior research suggests that interpersonal physiology does relate to monitoring and regulation in collaborative learning. However, this relation has not held true for all tasks and all groups (Dindar et al. 2019; Haataja et al. 2018). In general, researchers have called for more empirical evidence to confirm the methodological relevancy of multimodal data for conceptual and theoretical progress in the field of regulated learning (Järvelä et al. 2019; Reimann 2019).

This study supports the view that the different types of valence occurs in monitoring and can be recognized in the interactions of the students (Sobocinski et al. 2020). However, in the collaborative problem-solving context studied, much of the monitoring was neutral, without indicating a clear negative or positive valence. This distribution between different valence categories likely depends, to a great extent, on the contextual task demands and on the competence of the group. Because the task of the current study was novel for the participants, they were likely to monitor a lot of elements of which not all were related to the cognitive goals of the task. Complex problem solving also involves a high level of uncertainty (Dörner and Funke 2017), which is also likely to trigger knowledge construction that could be reflected as neutral monitoring. Still, in general monitoring with different valences stayed constant, and no significant differences between the transition episodes were found as the students progressed with the task.

The results suggest that groups differ in how equally participants take part in monitoring interactions, and this seems to be a characteristic that is somewhat “fixed” through the collaboration. This means that in some groups, participants continuously take more equal responsibility in terms of monitoring interactions. Similar findings for equality of participation in collaborative learning have been made in prior research (Cohen 1994, Kapur et al. 2008). This could indicate that initial individual and/or social conditions such as self-efficacy or interest might have a significant role in terms of equality of participation in monitoring interactions. This should be considered when interventions aiming to support monitoring are designed. For example, it might be important that prompts, which have been considered as a prominent approach to facilitate metacognitive interaction (Malmberg et al. 2015), are introduced early in the collaboration and that these would also aim to influence equality of participation in monitoring interactions.

Group members’ equality of participation in monitoring and monitoring with positive valence were found to be significant predictors of task performance. First, this supports the importance of shared regulatory processes in collaborative learning (Hadwin et al. 2018). When more participants monitor the task process, it is likely that different points of views are presented to construct shared knowledge (Roschelle and Teasley 1995), which also supports the reciprocal and negotiated use of strategies and joint responsibility for the progress of the task (Isohätälä et al. 2020). It is also possible that inequality in participation has a hindering effect on collaboration because it can involve overruling and social loafing types of phenomena (Linnenbrink-Garcia et al. 2011). Importantly, although this study gives some support to the view that participation in monitoring can (in some cases) signal good-quality collaboration (Jeong and Hmelo-Silver 2016), it does not account for every variety of other quality characteristics in monitoring, such as targets and accuracy of monitoring, which affect how successfully students construct knowledge together (Rogat and Linnenbrink-Garcia 2011). Second, the result underlines the role of the valence of monitoring in revealing pivotal moments in collaborations. An explanation for the latter could be that when group members understand something relevant for the task, signaled as monitoring with positive valence, their following performance is likely to be better. Still, though the GEE analysis adjusts to the repeated nature of the data, this result should be interpreted carefully because of the possible relatedness between transitions.

The results of the current study also show that the valence of monitoring relates to the interpersonal physiology of the students. For example, monitoring with negative valence was positively related to increase in physiological arousal (EDA slope) and physiological

synchrony (%DET). This means that when monitoring interactions suggest that there is a need to act and change something in a group's strategies, learners simultaneously show the effortful allocation of physiological resources and attune with each other. This aligns with the hypothesized interplay of regulation on different systemic levels in collaborative groups (Reimann 2019; Volet et al. 2009). Arousal itself might also have a special adaptive role in collaborative learning, since it seems to increase information sharing (Berger 2011), which might be especially useful when different strategies are considered. In contrast, monitoring with neutral valence was related to a decrease in physiological arousal, and monitoring with positive valence was negatively linked to physiological synchrony (%REC and ADL), which signals that synchrony is lower when things are considered to be on-track. These results can also explain why some of the prior studies have observed variance in the relation between monitoring and physiological synchrony for different tasks and groups (Dindar et al. 2019; Haataja et al. 2018). Because the valence of monitoring interactions reflects task demands, it is likely that tasks with different degrees of difficulty and groups with different levels of competence show different relationships between monitoring and physiological synchrony if the valence of monitoring is not considered. This is also in line with recent findings showing a link between the monitoring of mental effort (Dindar, Järvelä et al. 2020) and task difficulties (Malmberg et al. 2019) with physiological synchrony. It seems that physiological synchrony occurs as a condition, especially when the group as a whole considers that efforts and changes in the collaborative process are needed. When the strategies are changed, physiological synchrony tends to decrease (Mønster et al. 2016). Therefore, it could be hypothesized that groups that continuously show high synchrony and arousal throughout their collaboration might actually be unable to adapt and find efficient strategies and could therefore benefit from support.

Earlier research has found interpersonal physiology to reflect reciprocal contributions to collaborative learning (Schneider et al. 2020), and therefore, this study explored the possible relationship between equality of participation in monitoring and physiological synchrony and arousal. However, the current study could not find a relationship between physiological synchrony or arousal and equality of participation in monitoring interactions. The difference in these results with those of prior research might be due to the different indices used for measuring physiological synchrony, different task types, and conditions, or the fact that monitoring interactions make up only a small part of all the interactions in collaborations. The result also suggests that, although joint contributions to monitoring interactions seem to be important for performance, it is not a prerequisite for students to interpret similarly the demands of the ongoing situation.

In relation to physiological arousal, it should be noted that constant increased physiological arousal is a taxing condition for the body (Dawson et al. 2017). Therefore, if a task involves a lot of monitoring with negative valence caused, for example, by uncertainty or difficulty, this is likely to be exhausting at the mental and physiological levels (Barrett et al. 2016; Stephan et al. 2016). Also, it could be hypothesized that if the physiological state of the individual or group is not optimal to start with, it might be difficult to keep demanding monitoring and regulation processes going. Considering this, physiological data hold the potential to reveal a more holistic picture of the conditions within which self- and socially shared regulation processes emerge (Ben-Eliyahu and Bernacki 2015; Järvelä et al. 2019).

The challenge of using physiological data as a direct proxy to study any mental-level self-regulatory process remains (Winne 2019). In the case of arousal, though it is likely

related to monitoring of some current or forthcoming cognitive demands (Dawson et al. 2017), it is very challenging to say if these interpreted demands actually are cognitive or, for example, socio-emotional. Therefore, this allocation of physiological resources to meet the demands seen as changes in arousal cannot itself indicate whether it stems only from metacognition, or whether it is related to the learning task at all. This means that other data about the context are needed to make (at least somewhat) correct predictions of specific cognitive processes with physiological data (Järvelä et al. 2021). However, current investigations into the relations between physiology and social interactions in collaborative groups, as in this study, are likely to increase our understanding of how regulation involves adjustments on the multiple systemic levels of a collaborative group (Reimann 2019; Volet et al. 2009).

Further, because the valence of monitoring is likely, but not necessarily always, linked to emotional valence during collaborative learning, it would be important to investigate these processes together (Törmänen et al. 2021). Arousal originating from monitoring of cognition, or some other cognitive process, can also be considered as an affective ingredient when emotional experience is being constructed (Barrett 2016). This might partly explain some of the relationships found between emotions and metacognitive processes (Taub et al. 2019). One way to control and inspect these relations further in the future would be to separately capture fine-grained data of emotional expressions (e.g., facial expressions), metacognitive monitoring (e.g., interactions, or think-a-loud), and physiological arousal (e.g., electrodermal activity) and inspect the discrepancies and relations between them.

Future studies should also consider the nonlinear nature of emergence in these learning processes. For example, monitoring with different valence characteristics triggers different types of feedback loops of regulation, which are not linear but are likely to greatly affect the following learning process (Azevedo 2014). It is also important to investigate the temporal changes of interpersonal physiology such as moving in and out of synchrony (Likens and Wiltshire 2020), because these seem to be prominent in revealing the quality of the collaboration (Schneider et al. 2020) and might reflect adaptation or mal-adaptation of a group (Saxbe et al. 2020; Sobocinski et al. 2020). Because regulation in collaborative learning is a dynamic process and emerges on different systemic levels, which are likely to constantly interact with each other (Reimann 2019; Volet et al. 2009), a complex dynamical systems approach might offer potential methodological tools (e.g., MdRQA) for researching it in the future (Hilpert and Marchand 2018; Jacobson et al. 2016).

In conclusion, this study shows the importance of monitoring interactions for successful collaborative learning. It also provides evidence that levels of metacognitive interactions and interpersonal physiology are linked in collaborative learning. This suggests that monitoring interactions that serve groups' regulation of learning towards a shared goal are linked to a different type of regulatory process on another systemic level: interpersonal physiology. Though the strength of this link may be limited, it is likely that research involving multiple data modalities such as video and physiological data advances understanding of how regulation on these different levels intertwine and facilitate or hinder groups in progressing towards their goals in problem solving and learning.

Limitations

The current study has several limitations. First, the data for this study were gathered with a simulation task, which is rather specific and not usual for the participants to be working with. Therefore, it might have a novelty effect not seen in other contexts. Also, though the task performance measures of the simulation are likely to reflect some of the learning gains, they are not a direct measure of these.

The focus of the current study was limited to transition moments in the simulation process. This was due to limited possibilities to code the entire video corpus with the level of detail used in this study. As a result, some of the monitoring that focused more on content understanding during exploration was likely left out of the analysis. Additionally, the study only concentrated on one central process of regulation—monitoring—and two of its characteristics. Considering the full cycles of regulation, including different phases, would be important in the future.

The transitions between simulated months are also likely to trigger monitoring and therefore results might not apply to monitoring during tasks with less structure. Due to the nature of the task, this monitoring also mostly targeted cognitive performance in contrast to, for example, knowledge about strategies. Further, in a real classroom the learners might monitor progress towards a variety of goals, from which only some might be related to learning. Though monitoring of cognition has been shown to be linked with arousal, other factors are likely to be linked to physiological arousal and synchrony during collaborative learning.

Acknowledgements Oulu University LeaF research infrastructure has been used in the data collection of this study

Funding Open access funding provided by University of Oulu including Oulu University Hospital. This work was supported by the Academy of Finland [Grant numbers, 324381, 308809, 297686] and University of Oulu.

Declarations

Declaration of interest None

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, 21(8), 607–617. <https://doi.org/10.1016/j.tics.2017.05.004>
- 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>

- Amon, M. J., Vrzakova, H., & D'Mello, S. K. (2019). Beyond dyadic coordination: Multimodal behavioral irregularity in triads predicts facets of collaborative problem solving. *Cognitive Science*, 43(10), 1–22. <https://doi.org/10.1111/cogs.12787>
- Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self- and socially-regulated learning. *Metacognition and Learning*, 9(2), 217–228. <https://doi.org/10.1007/s11409-014-9123-1>
- Azevedo, R., & Witherspoon, A. M. (2009). Self-regulated learning with hypermedia. In Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.), *Handbook of metacognition in education* (pp. 319–339). Routledge
- Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist*, 45(4), 210–223. <https://doi.org/10.1080/00461520.2010.515934>
- Bachour, K., Kaplan, F., & Dillenbourg, P. (2010). An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *IEEE Transactions on Learning Technologies*, 3(3), 203–213. <https://doi.org/10.1109/TLT.2010.18>
- Barrett, L. F. (2016). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(1), nsw154. <https://doi.org/10.1093/scan/nsw154>
- Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160011. <https://doi.org/10.1098/rstb.2016.0011>
- Ben-Shakhar, G. (1985). Standardization Within Individuals: A Simple Method to Neutralize Individual Differences in Skin Conductance. *Psychophysiology*, 22(3), 292–299. <https://doi.org/10.1111/j.1469-8986.1985.tb01603.x>
- Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1), 80–91. <https://doi.org/10.1016/j.jneumeth.2010.04.028>
- Ben-Eliyahu, A., & Bernacki, M. L. (2015). Addressing complexities in self-regulated learning: a focus on contextual factors, contingencies, and dynamic relations. *Metacognition and Learning*, 10(1), 1–13. <https://doi.org/10.1007/s11409-015-9134-6>
- Berger, J. (2011). Arousal increases social transmission of information. *Psychological Science*, 22(7), 891–893. <https://doi.org/10.1177/0956797611413294>
- Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: a developmental psychological approach. *Annual Review of Psychology*, 66, 711–731. <https://doi.org/10.1146/annurev-psych-010814-015221>
- Bradley, M., & Lang, P. J. (1994). Measuring Emotion: The Self-Assessment Semantic Differential Manikin and the. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Burns, A., Greene, B. R., McGrath, M. J., O'Shea, T. J., Kuris, B., Ayer, S. M. ... Cionca, V. (2010). SHIMMER™ – A Wireless sensor platform for noninvasive biomedical research. *IEEE Sensors Journal*, 10(9), 1527–1534. <https://doi.org/10.1109/JSEN.2010.2045498>
- Chang, C. J., Chang, M. H., Chiu, B. C., Liu, C. C., Chiang, F., Wen, S. H. ... Chen, W., C.-K., &. (2017). An analysis of student collaborative problem solving activities mediated by collaborative simulations. *Computers & Education*, 114(300), 222–235. <https://doi.org/10.1016/j.compedu.2017.07.008>
- Clark, I., & Dumas, G. (2016). The regulation of task performance: A trans-disciplinary review. *Frontiers in Psychology*, 6(JAN). <https://doi.org/10.3389/fpsyg.2015.01862>
- 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>
- Cohen, E. G., & Roper, S. S. (1972). Modification of Interracial Interaction Disability: An Application of Status Characteristic Theory. *American Sociological Review*, 37(6), 643. <https://doi.org/10.2307/2093576>
- Cornejo, C., Cuadros, Z., Morales, R., & Paredes, J. (2017). Interpersonal Coordination: Methods, Achievements, and Challenges. *Frontiers in Psychology*, 8(September), 1–16. <https://doi.org/10.3389/fpsyg.2017.01685>
- Critchley, H. D., Eccles, J., & Garfinkel, S. N. (2013). Interaction between cognition, emotion, and the autonomic nervous system. In *Handbook of Clinical Neurology* (Vol. 117, Issue October, pp. 59–77). <https://doi.org/10.1016/B978-0-444-53491-0.00006-7>
- Croon, M. A., & van Veldhoven, M. J. P. M. (2007). Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods*, 12(1), 45–57. <https://doi.org/10.1037/1082-989X.12.1.45>
- Cui, J., & Qian, G. (2007). Selection of Working Correlation Structure and Best Model in GEE Analyses of Longitudinal Data. *Communications in Statistics - Simulation and Computation*, 36(5), 987–996. <https://doi.org/10.1080/03610910701539617>

- Danner, D., Hagemann, D., Holt, D. V., Hager, M., Schankin, A., Wüstenberg, S., & Funke, J. (2011). Measuring Performance in Dynamic Decision Making. *Journal of Individual Differences*, 32(4), 225–233. <https://doi.org/10.1027/1614-0001/a000055>
- Danyluck, C., & Page-Gould, E. (2019). Social and Physiological Context can Affect the Meaning of Physiological Synchrony. *Scientific Reports*, 9(1), 8222. <https://doi.org/10.1038/s41598-019-44667-5>
- Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 217–243). Cambridge University Press. <https://doi.org/10.1017/9781107415782.010>
- Dillenbourg, P. (1999). What do you mean by collaborative learning?. In Dillenbourg, P. (Ed.), *Collaborative learning: Cognitive and Computational approaches* (pp. 1–19). Elsevier
- Dillenbourg, P., Baker, M. J., Blaye, A., & O'Malley, C. (1995). The evolution of research on collaborative learning. In Spada, E., & Reiman, P. (Eds.), *Learning in Humans and Machine: Towards an interdisciplinary science* (pp. 189–211). Oxford: Elsevier
- Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3), 1–21. <https://doi.org/10.1145/3264913>
- 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>
- Dindar, M., Malmberg, J., Järvelä, S., Haataja, E., & Kirschner, P. A. (2020). Matching self-reports with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies*, 25(3), 1785–1802. <https://doi.org/10.1007/s10639-019-10059-5>
- Dindar, M., Järvelä, S., & Haataja, E. (2020). What does physiological synchrony reveal about metacognitive experiences and group performance? *British Journal of Educational Technology*, 51(5), 1577–1596. <https://doi.org/10.1111/bjet.12981>
- Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8(JUL). <https://doi.org/10.3389/fpsyg.2017.01153>
- Efklides, A., Kourkoulou, A., Mitsiou, F., & Ziliaskopoulou, D. (2006). Metacognitive knowledge of effort, personality factors, and mood state: their relationships with effort-related metacognitive experiences. *Metacognition and Learning*, 1(1), 33–49. <https://doi.org/10.1007/s11409-006-6581-0>
- Efklides, A., Schwartz, B. L., & Brown, V. (2018). Motivation and affect in self-regulated learning: Does metacognition play a role?. In *Handbook of self-regulation of learning and performance* (2nd ed., pp. 64–82). Routledge/Taylor & Francis Group
- Eichmann, B., Goldhammer, F., Greiff, S., Pucite, L., & Naumann, J. (2019). The role of planning in complex problem solving. *Computers & Education*, 128(July 2018), 1–12. <https://doi.org/10.1016/j.compedu.2018.08.004>
- Feldman, R. (2007). Parent-infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *Journal of Child Psychology and Psychiatry*, 48(3–4), 329–354. <https://doi.org/10.1111/j.1469-7610.2006.01701.x>
- Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology*, 34(1), 18–29. <https://doi.org/10.1016/j.cedpsych.2008.05.006>
- Greene, J. A., Hutchison, L. A., Costa, L. J., & Crompton, H. (2012). Investigating how college students' task definitions and plans relate to self-regulated learning processing and understanding of a complex science topic. *Contemporary Educational Psychology*, 37(4), 307–320. <https://doi.org/10.1016/j.cedpsych.2012.02.002>
- 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(July 2017), 337–347. <https://doi.org/10.1016/j.chb.2018.06.007>
- Hadwin, A. F., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation and shared regulation in collaborative learning environments. In Schunk, D. H., & Greene, J. A. (Eds.), *Handbook of self-regulation of learning and performance* (pp. 83–106). Routledge/Taylor & Francis Group
- Hajcak, G., McDonald, N., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology*, 40(6), 895–903. <https://doi.org/10.1111/1469-8986.00107>
- Häkkinen, P., Järvelä, S., Mäkitalo-Siegl, K., Ahonen, A., Näykki, P., & Valtonen, T. (2017). Preparing teacher-students for twenty-first-century learning practices (PREP 21): a framework for enhancing collaborative problem-solving and strategic learning skills. *Teachers and Teaching*, 23(1), 25–41. <https://doi.org/10.1080/13540602.2016.1203772>

- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (Vol. 9789400723, pp. 37–56). Springer Netherlands. https://doi.org/10.1007/978-94-017-9395-7_2
- 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>
- Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction*, 26(1), 48–94. <https://doi.org/10.1080/07370000701798495>
- Hurme, T. R., & Järvelä, S. (2005). Students' activity in computer-supported collaborative problem solving in mathematics. *International Journal of Computers for Mathematical Learning*, 10(1), 49–73. <https://doi.org/10.1007/s10758-005-4579-3>
- Hurme, T., Palonen, T., & Järvelä, S. (2006). Metacognition in joint discussions: an analysis of the patterns of interaction and the metacognitive content of the networked discussions in mathematics. *Metacognition and Learning*, 1(2), 181–200. <https://doi.org/10.1007/s11409-006-9792-5>
- Isöhätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research*, 81, 11–24. <https://doi.org/10.1016/j.ijer.2016.10.006>
- Isöhätälä, J., Näykki, P., & Järvelä, S. (2020). Convergences of joint, positive interactions and regulation in collaborative learning. *Small Group Research*, 51(2), 229–264. <https://doi.org/10.1177/1046496419867760>
- Jacobson, M. J., Kapur, M., & Reimann, P. (2016). Conceptualizing debates in learning and educational research: Toward a complex systems conceptual framework of learning. *Educational Psychologist*, 51(2), 210–218. <https://doi.org/10.1080/00461520.2016.1166963>
- 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., Hadwin, A. F., Malmberg, J., & Miller, M. (2018). Contemporary perspectives of regulated learning in collaboration. In F. Fischer, C. Hmelo-Silver, S. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (1st ed., pp. 127–136). Routledge.
- Järvelä, S., Järvenoja, H., & Malmberg, J. (2019). Capturing the dynamic and cyclical nature of regulation: Methodological 10.1007/s11409-021-09279-3 Progress in understanding socially shared regulation in learning. *International Journal of Computer-Supported Collaborative Learning*, 14(4), 425–441. <https://doi.org/10.1007/s11412-019-09313-2>
- 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(March), 101203. <https://doi.org/10.1016/j.learninstruc.2019.04.004>
- Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: how to support collaborative learning? How can technologies help? *Educational Psychologist*, 51(2), 247–265. <https://doi.org/10.1080/00461520.2016.1158654>
- Kapur, M., Voiklis, J., & Kinzer, C. K. (2008). Sensitivities to early exchange in synchronous computer-supported collaborative learning (CSCL) groups. *Computers & Education*, 51(1), 54–66. <https://doi.org/10.1016/j.compedu.2007.04.007>
- Karabenick, S., & Zusho, A. (2015). Examining approaches to research on self-regulated learning: conceptual and methodological considerations. *Metacognition and Learning*, 10(1), 151–163. <https://doi.org/10.1007/s11409-015-9137-3>
- Khosa, D. K., & Volet, S. E. (2014). Productive group engagement in cognitive activity and metacognitive regulation during collaborative learning: can it explain differences in students' conceptual understanding? *Metacognition and Learning*, 9(3), 287–307. <https://doi.org/10.1007/s11409-014-9117-z>
- Koivuniemi, M., Järvenoja, H., & Järvelä, S. (2018). Teacher education students' strategic activities in challenging collaborative learning situations. *Learning, Culture and Social Interaction*, 19(November), 109–123. <https://doi.org/10.1016/j.lcsi.2018.05.002>
- Kreibig, S. D., & Gendolla, G. H. E. (2014). Autonomic nervous system measurement of emotion in education and achievement settings. In Pekrun, R., & Linnenbrink-Garcia, L. (Eds.), *International Handbook of Emotions in Education* (pp. 625–642). Routledge
- Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, 44(1), 46–53. <https://doi.org/10.3102/0013189X15569530>
- Likens, A. D., & Wiltshire, T. J. (2020). Windowed multiscale synchrony: modeling time-varying and scale-localized interpersonal coordination dynamics. *Social Cognitive and Affective Neuroscience*, 22(7), 117–122. <https://doi.org/10.1093/scan/nsaa130>
- Linnenbrink-Garcia, L., Rogat, T. K., & Koskey, K. L. K. (2011). Affect and engagement during small group instruction. *Contemporary Educational Psychology*, 36(1), 13–24. <https://doi.org/10.1016/j.cedpsych.2010.09.001>

- Lu, J. G., Page-Gould, E., & Xu, N. R. (2017). *MicroMacroMultilevel R package*.
- Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior*, 52, 562–572. <https://doi.org/10.1016/j.chb.2015.03.082>
- 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>
- Malmberg, J., Haataja, E., Seppänen, T., & Järvelä, S. (2019). 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>
- Marci, C. D., Ham, J., Moran, E., & Orr, S. P. (2007). Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *The Journal of Nervous and Mental Disease*, 195(2), 103–111. <https://doi.org/10.1097/01.nmd.0000253731.71025.fc>
- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22(1), 114–140. <https://doi.org/10.1037/met0000078>
- Mendes, W. B. (2009). Assessing autonomic nervous system activity. In Harmon-Jones, E., & Beer, J. S. (Eds.), *Methods in social neuroscience* (pp. 118–147). Guilford Press
- Mønster, D., Håkansson, 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>
- Näykki, P., Järvenoja, H., Järvelä, S., & Kirschner, P. (2017). Monitoring makes a difference: quality and temporal variation in teacher education students' collaborative learning. *Scandinavian Journal of Educational Research*, 61(1), 31–46. <https://doi.org/10.1080/00313831.2015.1066440>
- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41(5), 673–690. <https://doi.org/10.1007/s1135-006-9018-6>
- 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*, 21(2), 99–141. <https://doi.org/10.1177/1088868316628405>
- Pijeira-Díaz, H. J., Drachler, H., Kirschner, P. A., & Järvelä, S. (2018). Profiling sympathetic arousal in a physics course: How active are students? *Journal of Computer Assisted Learning*, 34(4), 397–408. <https://doi.org/10.1111/jcal.12271>
- Pijeira-Díaz, H. J., Drachler, 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(September 2018), 188–197. <https://doi.org/10.1016/j.chb.2018.11.008>
- Pino-Pasternak, D., Whitebread, D., & Neale, D. (2018). The role of regulatory, social, and dialogic dynamics on young children's productive collaboration in group problem solving. *New Directions for Child and Adolescent Development*, 2018(162), 41–66. <https://doi.org/10.1002/cad.20262>
- Reimann, P. (2019). Methodological progress in the study of self-regulated learning enables theory advancement. *Learning and Instruction*, xxx, 101269. <https://doi.org/10.1016/j.learninstruc.2019.101269>
- 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>
- 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>
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer Supported Collaborative Learning* (Vol. 128, pp. 69–97). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-85098-1_5
- Rudolph, J., Niepel, C., Greiff, S., Goldhammer, F., & Kröner, S. (2017). Metacognitive confidence judgments and their link to complex problem solving. *Intelligence*, 63(May 2016), 1–8. <https://doi.org/10.1016/j.intell.2017.04.005>
- 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., & Perkins, D. N. (1998). Chapter 1: Individual and Social Aspects of Learning. *Review of Research in Education*, 23(1), 1–24. <https://doi.org/10.3102/0091732X023001001>
- Saxbe, D. E., Beckes, L., Stoycos, S. A., & Coan, J. A. (2020). Social allostasis and social allostatic load: A new model for research in social dynamics, stress, and health. *Perspectives on Psychological Science*, 15(2), 469–482. <https://doi.org/10.1177/1745691619876528>

- 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>
- 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*. <https://doi.org/10.1007/s11409-020-09224-w>
- Stennett, R. G. (1957). The relationship of performance level to level of arousal. *Journal of Experimental Psychology*, 54(1), 54–61. <https://doi.org/10.1037/h0043340>
- Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A. E., Paliwal, S., Gard, T. ... Petzschner, F. H. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, 10(NOV2016). <https://doi.org/10.3389/fnhum.2016.00550>
- Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, 106(1), 5–15. <https://doi.org/10.1016/j.physbeh.2011.06.004>
- Strain, A. C., Azevedo, R., & D'Mello, S. K. (2013). Using a false biofeedback methodology to explore relationships between learners' affect, metacognition, and performance. *Contemporary Educational Psychology*, 38(1), 22–39. <https://doi.org/10.1016/j.cedpsych.2012.08.001>
- Strauß, S., & Rummel, N. (2021). Promoting regulation of equal participation in online collaboration by combining a group awareness tool and adaptive prompts. But does it even matter? *International Journal of Computer-Supported Collaborative Learning*, 16(1), 67–104. <https://doi.org/10.1007/s11412-021-09340-y>
- Taub, M., Azevedo, R., Rajendran, R., Cloude, E. B., Biswas, G., & Price, M. J. (2019). How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system? *Learning and Instruction*, July 2018, 101200. <https://doi.org/10.1016/j.learninstruc.2019.04.001>
- 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*, 0123456789. <https://doi.org/10.1007/s11423-021-10037-0>
- Ucan, S., & Webb, M. (2015). Social regulation of learning during collaborative inquiry learning in science: How does it emerge and what are its functions? *International Journal of Science Education*, 37(15), 2503–2532. <https://doi.org/10.1080/09500693.2015.1083634>
- Ullsperger, M., Danielmeier, C., & Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. *Physiological Reviews*, 94(1), 35–79. <https://doi.org/10.1152/physrev.00041.2012>
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning*, 1(1), 3–14. <https://doi.org/10.1007/s11409-006-6893-0>
- 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>
- Wallot, S., & Leonardi, G. (2018). Analyzing multivariate dynamics using cross-recurrence quantification analysis (CRQA), diagonal-cross-recurrence profiles (DCRP), and multidimensional recurrence quantification analysis (MdrQA) – A Tutorial in R. *Frontiers in Psychology*, 9(December), 1–21. <https://doi.org/10.3389/fpsyg.2018.02232>
- 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. <https://doi.org/10.1371/journal.pone.0168306>
- Wallot, S., Roepstorff, A., & Mønster, D. (2016). Multidimensional recurrence quantification analysis (MdrQA) for the analysis of multidimensional time-series: a software implementation in matlab and its application to group-level data in joint action. *Frontiers in Psychology*, 7, 1–13. <https://doi.org/10.3389/fpsyg.2016.01835>
- Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S. ... Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning*, 4(1), 63–85. <https://doi.org/10.1007/s11409-008-9033-1>
- Winne, P. H. (2018). Cognition and metacognition within self-regulated learning. In Schunk, D. H., & Greene, J. A. (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 36–48). Routledge
- 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. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In Hacker, D. J., Dunlosky, J., & Graesser, A. (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Erlbaum

- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27(4), 551–572. [https://doi.org/10.1016/S0361-476X\(02\)00006-1](https://doi.org/10.1016/S0361-476X(02)00006-1)
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5), 459–482. <https://doi.org/10.1002/cne.920180503>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.