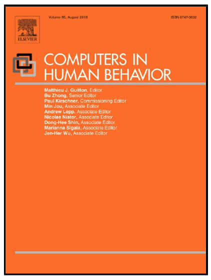
Accepted Manuscript



Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning?

Jonna Malmberg, Sanna Järvelä, Jukka Holappa, Eetu Haataja, Xiaohua Huang, Antti Siipo

PII:	S0747-5632(18)30307-8
DOI:	10.1016/j.chb.2018.06.030
Reference:	CHB 5582
To appear in:	*Computers in Human Behavior*
Received Date:	24 November 2016
Accepted Date:	24 June 2018

Please cite this article as: Jonna Malmberg, Sanna Järvelä, Jukka Holappa, Eetu Haataja, Xiaohua Huang, Antti Siipo, Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning?, *Computers in Human Behavior* (2018), doi: 10.1016/j.chb.2018.06.030

Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning?

Jonna Malmberg, Sanna Järvelä, Jukka Holappa, Eetu Haataja, Xiaohua Huang, Antti Siipo

University of Oulu

Acknowledgements: This study was supported by the Finnish Academy grant, 275440, SLAM research project

Correspondence concerning this article should be addressed to Jonna Malmberg, Faculty of Education, Department of Educational Sciences and Teacher Education, University of Oulu, Finland, P.O. Box 2000.

E-mail: jonna.malmberg@oulu.fi

Abstract

Progress in the development of technology has provided data-capturing devices that make it possible to identify detailed processes in collaborative learning. This study utilized multichannel data, namely physiological data, video observations, and facial recognition data, to explore what they can reveal about types of interaction and regulation of learning during different phases of collaborative learning progress. Participants were five groups of three members each, selected for further study from an initial set of 48 students. The collaborative task was to design a healthy breakfast for an athlete. Empatica sensors were used to capture episodes of simultaneous arousal, and video observations were used to contextualize working phases and types of interaction. Facial expression data were created by post-processing video-recorded data. The results show that simultaneous arousal episodes occurred throughout phases of collaborative learning and the learners presented the most negative facial expressions during the simultaneous arousal episodes. Most of the collaborative interaction during simultaneous arousal was low-level, and regulated learning was not observable. However, when the interaction was high-level, markers of regulated learning were present; when the interaction was confused, it included monitoring activities. This study represents an advance in testing new methods for the objective measurement of social interaction and regulated learning in collaborative contexts.

Keywords: self-regulated learning; arousal; simultaneous arousal; regulated learning, collaborative learning; facial expression recognition

1. Introduction

Decades of research on self-regulated learning (SRL) indicate that learners who have strong self-regulation skills can adapt to challenges in the learning situation (Zimmerman & Schunk, 2011). In general, contemporary perspectives view SRL as a cyclical, complex metacognitive and social process that involves adaptation in thinking, motivation, emotion, and behavior (Cleary & Zimmerman, 2012; Winne & Hadwin, 1998). According to existing models of SRL (Hadwin, Järvelä, & Miller, 2016), situated challenges create opportunities to understand how learners strategically adapt their task perceptions, goals, and strategies. The problem, however, is that it is extremely difficult to recognize and make visible how learners confront challenges that might undermine or activate their learning progress in the context of collaboration.

With the latest advances in technology, new data-capturing devices enable researchers to go beyond the ontologically flat data of observable behaviors (Channel & Muhl, 2016; Reimann, Markauskaite, & Bannert, 2014). For example, it is now becoming possible to implement physiological measures in a classroom setting, and there is increasing understanding of how physiological signals such as measures of electrodermal activity (EDA) can elucidate cognition, affect, and metacognition in educational settings (Harley, Bouchet, Hussain, Azevedo, & Calvo, 2015; Immordino-Yang & Christodolou, 2014). Because physiological signals are sensitive to contextual changes, they have the potential to advance empirical research on regulated learning in the context of collaboration (Azevedo, 2015) by providing information about cognitive demands, task difficulty, and increased attention in relation to task engagement (Henriques, Paiva, & Antunes, 2013). A substantial body of empirical research indicates that increases in EDA are associated with performance of problem-solving tasks (e.g., Munro, Dawson, Schell, & Sakai, 1987). Additionally, there are indications that the magnitude of increases in EDA depends on task difficulty (Fairclough, Venables, & Tattersall, 2005).

Recent studies also indicate that, in collaborative contexts, the individual learner's physiological reactions are dependent upon and shaped by other learners (Gillies et al., 2016), which suggests that those reactions reflect invisible social processes that co-occur with observable interactions (Palumbo et al., 2016). This invites investigation not only of students' verbal utterances in learning situations but also of the physiological reactions— normally invisible to human detection— underlying those utterances. The present study describes a novel approach that uses multichannel data to explore regulated learning as it emerges in collaborative learning.

The study builds on self-regulated learning theory (Winne & Hadwin, 1998), which is characterized as a cyclical process shaped and affected by past learning experiences and by the social learning context (Volet et al., 2009). In collaborative learning, multiple self-regulating agents seek a common understanding and task solution by engaging in different types of mutual interaction (Järvelä, Järvenoja, Malmberg, Isohätälä, & Sobocinski, 2015; Volet, Summers, & Thurman, 2009). From a methodological perspective, triangulation of multichannel video observations, physiological data, and facial expression data represents a fundamentally new approach, using both objective and subjective means to capture how self-regulating learners interact in collaborative learning contexts.

1.1. Self-regulated learning as an adaptive and cyclical process

In learning, regulation is an adaptive, cyclical process, entailing a series of contingencies over time (Cleary & Zimmerman, 2012) in responding to new challenges, situations, or failures in ways that optimize progress toward personal goals. Regulation is not spontaneous or automatic; rather, it is characterized by intent or purposeful action in response to challenges or novel situations (Hadwin, Järvelä, & Miller, 2016). Models of SRL specify phases that learners go through during task execution as they proceed toward their learning goals. Depending on the model, the number of phases varies from three to four (Winne & Hadwin, 1998; Zimmermann, 2001), varying from broader to highly specific definitions of the activities in each phase (Azevedo, Johnsson, Chauncey, & Grasser, 2011). In general, three phases can be identified in SRL: planning, strategic enactment, and evaluation

(Zimmerman, 2001). Although different models of SRL commonly present these phases as linear, this does not mean that they necessarily occur in the same order during the learning process (Malmberg, Järvelä, & Järvenoja, 2016). Indeed, in each phase of study, learners revisit and change their plans or strategies through metacognitive monitoring. It can be concluded that such changes depend on the extent to which incoming information aligns with existing knowledge structures and whether there are inconsistencies and discrepancies in the information stream (D'Mello, Lehman, Pekrun, & Graesser, 2011).

Recent empirical studies of SRL have emphasized the importance of learners' adaptations within and across learning situations (Boekaerts & Niemivirta, 2001; Johnsson, Azevedo, & D'Mello, 2011; Zimmerman, 2001). This suggests that SRL is inherently situated, as each learning situation forms its own entity, with unique challenges that may result from external task conditions (e.g., task difficulty, poor instructions) or from internal task conditions (e.g., confusion, frustration) that change within and across learning situations (Winne & Hadwin, 1998). In addition, D`Mello et al. (2011) argued that when learners become confused and encounter challenges while studying, they need to engage in effortful cognitive activities to resolve the situation. In other words, learners must adapt their SRL strategically to align with situational demands (Järvelä, Hadwin, & Miller, 2016; Malmberg, Järvelä, Järvenoja, & Panadero, 2015). The problem is that learners often do not make visible any mental processes that might actually indicate metacognitive monitoring, confusion, or task difficulty during learning that might ultimately lead to strategic adaptation. For that reason, unmasking the invisible reactions of body and brain may reveal when learners are engaging in metacognitive monitoring, followed by regulation (or a lack of regulation) in the process of learning (Järvelä, Malmberg, Sobicinsky, Haataja, & Kirschner, 2016).

1.2. Collaborative learning and social interaction

The most widely used definition of collaboration describes it as a construction of shared understanding through interaction with others, in which the participants are committed to or engaged in shared goals and problem solving (Erkens, Jaspers, Prangsma, & Kanselaar, 2005; Teasley & Rochelle, 1993). In collaborative learning, students' behavior is more complex than in the case of individual learning (Hackman & Morris 1975). A group's learning performance is not merely a reflection of individual learners' regulation but a complex combination of all learners' contributions to the group's collective effort. That is, in a collaborative learning situation, multiple self-regulating agents interact to construct a shared problem space (Hadwin et al., 2017). In that sense, interactions in collaborative learning are necessarily reciprocal (Dillenbourg, 1999), and this should include joint attention and equal contributions to discussion of task concepts (Barron, 2003). When working collaboratively, learners must share information, search for meanings and solutions, and develop a shared understanding of the problem (e.g., Iiskala, Vauras, Lehtinen, & Salonen, 2011; Teasley & Roschelle, 1993). In other words, individual and interactive contributions are essential for successful collaborative learning (Järvelä et al., 2013).

One way of characterizing the types of interaction that emerge in collaborative learning is to categorize them as high- or low-level interactions (Volet et al., 2009). During high-level interactions, all the learners contribute to the co-construction of knowledge by asking questions and negotiating task concepts as they engage in regulation of learning. In low-level interactions, learners are instead more focused on gaining an individual understanding of the relevant topics by reading and processing information rather than interacting with their group members (Molenaar & Chiu, 2014; Volet et al., 2009). From a collaborative learning perspective, both types of interaction are appropriate, but if their collaboration is to succeed, learners must negotiate and determine common ground in relation to the task and must keep track of their joint learning progress (Ku, Tseng, & Akarasriworn, 2013).

Effective collaboration requires an environment that promotes positive interdependence and facilitates each group member's contribution. One way of enhancing collaborative learning is to structure student interaction through scripting (Morris et al., 2010). The purpose of the script is to provide opportunities to share learning materials and edit them collaboratively, engaging in group discussions and giving and receiving peer feedback on the collaborative learning progress (Miller & Hadwin, 2015). An extensive body of existing empirical research confirms that scripting of collaborative learning facilitates interaction (Weinberger, Stegmann, & Fischer, 2010), knowledge construction (Buder & Bodemer, 2008), and awareness of both social and cognitive learning activities (Phielix, Prins, & Kirschner, 2010).

Despite the effectiveness of scripting collaborative learning, task engagement and task-focused interactions are not guaranteed (Weinberger, Stegmann, & Fischer, 2010). To succeed in their collaboration, learners need to focus on task-related and cognition-focused interactions for task completion while also maintaining positive socio-emotional interactions that are relevant for self-expression (Kreijns, Kirschner, & Jochems, 2003). If, for example, individual self-regulating students differ in terms of personal goals or expectations and these differences are not resolved, the group's collaborative interactions and motivational engagement may be negatively affected (Järvelä & Hadwin, 2013). Examining what collaborative learning patterns look like in terms of the interaction between group members, Kwon, Liu, and Johnson (2014) found that positive socio-emotional interactions were associated with intensive collaboration while collaborative learning remained dormant among groups exhibiting little socio-emotional interaction. Similarly, in their study of the interplay of cognitive and socio-emotional interaction in a collaborative learning context, Järvelä et al. (2016) noted that socio-emotional interaction between the participating students increased in the early phases of collaborative learning. They argued that socio-emotional and cognitive interaction are intertwined in collaborative learning, and that it is empirically difficult to separate these constructs from learner interaction.

To date, relatively few empirical studies have explored both the types of interaction that emerge when learners collaborate and expressions of emotion in those situations (Järvenoja et al., 2017). This is largely because emotional expressions are sometimes difficult to capture because of social masking, and because there are few methods of capturing such expressions as they emerge in learning situations (D'Mello & Kory, 2015).

2. Multimodal data in research on regulated learning

Multichannel data—that is, data from different modalities—may include both subjective and objective data. For example, subjective data such as repeated and contextualized self-reports may reveal the intentions behind student learning. Conversely, objective data such as automatic facial expression data and physiological measures of the various components of EDA provide continuous information about behavioral and mental processes like confusion, increasing effort, or increased attention, which are otherwise almost impossible to capture (Henriques, Paiva, & Antunes, 2013; Winne, 2010). Facial expression recognition data can be helpful in identifying affective states (Harley et al., 2016) because EDA does not reveal valence—that is, whether changes are a result of positive or negative reactions. In addition, video observations can reveal the sequential and temporal dynamics of SRL and, in collaborative contexts, the quality of interaction between group members. Azevedo (2015) recently elaborated how different data modalities can be used to capture SRL, based on three criteria: *ideally suited*, *not ideally suited*, and *depends on the context*. For example, while facial expression recognition data are ideally suited to capturing affective states, physiological measures like EDA are ideally suited to capturing cognition and affect, and video observations are ideally suited to capturing cognition and metacognition.

Until now, the use of physiological sensors and facial expression recognition data has focused almost entirely on individual learner signals, ignoring opportunities to monitor more elusive phenomena such as the quality of social interactions informed by a theoretical understanding of regulated learning (Järvelä et al., 2016). Here, we argue the need to investigate whether physiological sensors in combination with facial expression recognition data can shed light on the quality of social interaction during collaborative learning (e.g., Azevedo et al., 2015; Hernández-García, González-González, Jiménez-Zarco, & Chaparro-Peláez, 2014). The availability of wearable physiological sensors now provides an opportunity to study physiological markers during collaborative interaction as they emerge in a learning context.

2.1. Sympathetic arousal as a marker of challenge and engagement

EDA (also referred to as *galvanic skin response*) provides an indication of the activation of sweat glands innervated by the sympathetic nervous system. Measured skin conductance is considered one of the best indicators of sympathetic arousal (Hernández-García et al., 2014). For example, EDA during sympathetic arousal can signal cognitive demand related to task difficulty, cognitive load, and potential for learning (Fairclough, Venables, & Tattersall, 2005; Ferreira et al., 2014), as well as engagement related to increased mental effort (Hernández-García et al., 2014). Increased mental effort in particular can potentially improve performance, especially when the task is complex (Hockey, 1997).

EDA includes three different modes, reflecting different psychological processes: tonic activity levels (baseline activity); phasic responses (sudden impact of a stimulus); and spontaneous fluctuations (nonspecific skin conductance responses). Specifically, changes in the level and occurrence of nonspecific skin conductance responses (NSSCRs) reflected as peaks at EDA can signal increased mental effort related to task difficulty, cognitive load, or engagement, and these are indicators of physiological arousal (Azevedo et al., 2014; Hernández-García et al., 2014). Nikula (1991) posited that NSSCRs relate in particular to cognitive processes, especially "negatively tuned cognitive activity." It follows that changes in NSSCRs may reflect changes in situational context or dispositional factors (Mendez, 2009). Such changes in physiological processes have typically been examined from the perspective of individual students (e.g., Fairclough, Venables, & Tattersall, 2005) but not in terms of how NSSCRs may occur simultaneously among group members (Elkins et al., 2009). It seems important, then, to understand how changes in NSSCRs (reflected in EDA peaks) may signal a need to activate SRL in the face of learning challenges.

2.1. Emotion recognition from facial expressions

Facial expression is the most important channel for automatically detecting emotions (Azevedo 2015; Kortelainen et al., 2012). In a learning context, facial expression may reflect negative emotions such as boredom, confusion, frustration, or anxiety, depending on the challenge encountered. In contrast, positive facial expressions may reflect interest or engagement (D'Mello et al., 2011). There is also increasing evidence of the importance of emotional experiences in promoting learning (Linnenbrink-Garcia & Pekrun, 2011); for example, research has established a correlation between feelings of anxiety and decreased learning performance (Schutz & Davis, 2000).

However, differences in facial expression across various emotions can be very subtle and are rarely classified in terms of prototypical, universally recognized categories such as happiness, sadness, and surprise as described by Ekman (1982). Faces create different expressions by activating different combinations of independently moving muscles. A system known as the Facial Action Coding System (FACS) was proposed by Ekman and Friesen (1978), describing 44 different Action Units that can be used to encode human facial expressions. The FACS does not classify the meaning of or reason for different facial expressions but only the physical actions needed to generate them.

Clearly, emotion signaling is only one use of facial expressions, and over the years, various methods have been proposed in the literature to detect emotion-related expressions by using computer vision techniques. Some of these methods attempt to detect a subset of FACS-compatible Action Units directly from facial images (e.g., Valstar & Pantik, 2006), which can then be used to infer the matching emotional state. Other systems classify facial expressions by using some other intermediate feature space (e.g., Eleftheriadis et al., 2015; Huang et al., 2016). These intermediate feature spaces, or at least mappings from them to target expressions, are then optimized using machine learning techniques to obtain better classification results. The facial expression classifier used in this experiment falls into the latter category; it produces an estimate of valence (the negative-positive emotional axis), which is one of the most important dimensions for distinguishing between different emotions (Fontaine et al., 2007).

2.2. Aims

In the present study, observational data, along with physiological measures and facial expression recognition data, were used to capture interactions during the collaborative learning process. Skin conductance sensors were used to track episodes of simultaneous arousal among group members, along with facial expression recognition data to capture affect. The study addressed three research questions. 1) How do phase of working and type of interaction relate to simultaneous occurrence of arousal among group members? 2) What types of facial expression (positive, negative, or neutral) are observed when arousal occurs simultaneously among group members? 3) How does regulation of learning appear during interaction when arousal occurs simultaneously among group members?

3. Methods

3.1. Participants and context

The participants (*N* = 48, 27 females) were high school students (*Mean age* = 17.4 years; *SD* = .67) from a teacher training school. Participation was voluntary. During the study, the students collaborated in groups of three (16 groups in total) on a collaborative task to design a "healthy breakfast," using the weSpot learning environment.

3.1. Study design

The total duration of the collaborative task was 75 minutes. At the outset, the students were assigned to groups on the basis of their prior knowledge of the topic and their individual scores in the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, Smith, Garcia, & McKeachie, 1993). Prior knowledge was assessed by ten items, asking such questions as "What nutrients should breakfast include?" "Why are fibers important?" and "What type of food includes carbohydrates?" These ten multiple choice questions reflected exactly the knowledge needed to complete the collaborative task; each question had one or more correct answers. The MSLQ is a 7-point Likert-like instrument for measuring students' understanding of their SRL. It comprises 81 items and a total of 15 scales in two sections, with Cronbach's α ranging from .45 to .88. On the basis of individual scores on the MSLQ (*M* = 357, *SD* = 53.4, *Range* = 245–467) and prior knowledge of the topics (*M* = 75, *SD* = 10.65, *Range* = 50–95, *Max* 100), the students were divided into heterogeneous groups of three.

Following an introductory explanation of the experiment, Empatica E3 sensors were placed on the students to measure their EDA. Sitting at separate tables, each group was provided with an iPad for task execution, as well as other resources to complement the task (Figure 1). Once organized, they were given instructions about what to do, along with instructions for using the weSpot Learning environment for collaboration.

WeSpot is a cloud-based environment for collaborative inquiry learning that allows learners to perform scientific investigations (Mikrodyannis et al., 2013). It also provides a flexible tool for instructors to arrange and script collaborative inquiry learning (Figure 1).

Figure 1 here

The collaborative task was to "Design a perfect breakfast for a marathon runner." The weSpot learning environment included 1) a case description of a hypothetical person, along with that person's daily energy needs. The problem statement at the beginning of the task also included 1a) information in percentage terms of how much fiber, calories, fat, and carbohydrates the breakfast should include. The students' collaborative task was to complete 1b) an Excel sheet that included a detailed list of nutrients that a marathon runner should eat for breakfast (Figure 1).

In addition to the above instructions and information, the learning environment included 2) a script to guide collaboration, comprising five phases: 2a) activate prior knowledge and plan your collaborative working; 2b) set criteria for task completion; 2c) search for information; 2d) discuss and complement the findings in the learning environment; and 2e) communicate the results. The task outcome was specified as a detailed list that included a description of the nutrients that a healthy breakfast should include.

Activate prior knowledge asked the students to think first about what they already knew about the topics and to define concepts relevant to the task.

Set criteria for task completion asked the students to set four rules to be followed when designing the breakfast—that is, they needed to calculate the exact proportion of fiber, fat, calories, and carbohydrates the breakfast should include.

Search for information asked the students to find actual breakfast products, either from resources available in the classroom (empty food packages) or from the Internet, and to complete an Excel sheet listing nutrients that a marathon runner should eat for breakfast.

Discuss and complement the findings in the learning environment asked the students to check the correctness of their calculations on the Excel sheet and to evaluate the correctness of their answer.

3.2. Data collection

During the experiment, three different data modalities were collected in LeaForum (http://www.oulu.fi/eudaimonia/node/19394), which is a classroom-like space with modern equipment that includes a spherical, 360º point-of-view MORE video system and Empatica E3 skin conductance sensors that can be used to track EDA. The collected data included observational, physiological, and facial expression data (processed from the video data).

The *observational data* consisted of video recordings of the students' collaborative learning using the MORE video system, which can record 30 speech tracks and three video tracks simultaneously through spherical, 360º point-of-view cameras.

Physiological data were collected using Empatica E3 skin galvanic sensors that tracked the students' EDA. NSSCR data, which indicate the strength of arousal to external stimuli, were isolated from EDA data.

Facial expression data were created by post-processing the video recordings from the MORE system. Following recording, these data were processed using the face analysis component of the

MORE system, which detects and tracks faces throughout the sequence and estimate valences (positive, negative, or neutral) for each face (Kortelainen et al., 2012).

4. Analysis

The analysis focused on physiological data, observation data, and facial expression data reflecting group-level events rather than events at the level of the individual student. As the sampling rate differed for each of the three data channels, the time window was set to one minute. The one-minute unit of analysis was selected because of the focus on temporally unfolding events and because NSSCR rates can typically vary from 1–3 responses per minute (low arousal) to 20–25 responses per minute (high arousal). The one-minute segment facilitated investigation of episodes where two or more students simultaneously registered a high-arousal NSSCR rate.

4.1. EDA analysis

The purpose of EDA analysis was to identify responses signaling individual physiological arousal. From the EDA, NSSCRs per minute were selected for further inspection, as the study design included events that unfolded over time, and there were no specific time-locked events of interest (Blaschovich, Vanman, Mendes, & Dickerson, 2011; Mendes, 2009). The aim was to identify arousal episodes that occurred simultaneously across two or three group members during a one-minute time window.

To ensure the reliability of the data, the Matlab toolbox Ledalab (Benedek & Kaernbach, 2010) was used to pre-process EDA data. In the first step, the quality of the recordings was checked manually, and any recordings that clearly included constant movement artifacts or abnormally low values were removed from the analysis. Because of the poor quality of the EDA signal, only five groups (*n* = 15) were included for further investigation.

The data were first standardized as z-scores. The EDA signal was then smoothed out, using an adaptive Gaussian filter, and was separated into tonic and phasic components (Benedek & Kaernbach, 2010). NSSCRs with a minimum amplitude of 0.05μS (Boucsein et al., 2012) were then identified, and one-minute time segments were used to calculate NSSCR frequency ($M = 7.54$, $SD = 6.89$, *Range* = 0–29) for each student in order to identify arousal episodes. A segment was considered to be a arousal episode for the student if the NSSCR frequency/min value rose above the individual's mean. Frequencies in arousal episodes varied from 7 to 29 NSSCRs/min ($M = 15.17$, $SD = 4.51$). The one-minute segment was classified as a simultaneous arousal episode if several group members simultaneously experienced arousal. In other words, we selected only those one-minute episodes during which arousal occurred at the same time across two or three group members during a one-minute time window. In total, arousal was experienced simultaneously by two group members on 102 occasions, and by all three group members on 29 occasions, yielding 131 simultaneous one-minute arousal episodes from the five video observations.

4.2. Observational analysis

The video data were analyzed in two levels, using qualitative content analysis (Jordan & Henderson, 1995) based on data- and theory-driven categories. The analysis took account of both visible activity and the quality of the interaction. To begin, high-arousal episodes involving at least two students were located from the video data on the basis of video timestamps and timestamps for observational data. This process yielded 131 one-minute episodes from the five video observations.

At the first level of the analysis, work phases were identified in terms of progress on the collaborative task and how the simultaneous arousal episodes related to these phases. The qualitative descriptions in Table 1 identify the different phases, indicating progress on the task. The identified phases are a) work instructions; b) searching for information; c) communication; d) learning environment; e) adding information; and f) off-topic conversations. The categories represent progress on the task according to the script; for example, the experiment commenced with work instructions, followed by use of the learning environment according to the script, searching for information, and adding information.

In contrast, communication or off-topic conversations could take place at any time while working on the collaborative task.

Table 1 Here

At the second level of the analysis, each phase of working was then further coded by interaction type. The purpose of defining interaction types was to distinguish and characterize the quality of collaborative interactions and regulated learning during different phases of working. This yielded three different modifiers: a) low-level interaction, b) high-level interaction, and c) confusion. Off-topic conversations were not linked to activity quality because they were seen as inhibiting progress on the task. Table 2 presents more detailed descriptions of the qualitative modifiers as adapted from D'Mello and Graesser (2016), Molenaar and Chiu (2014), and Volet et al. (2009).

Table 2 Here

The category "low-level interaction" refers to *reading and processing of information to acquire knowledge, accompanied by low interaction*. This means that group members were either silent when one person was talking or agreed silently but did not participate in the conversation, and there was no visible regulation of learning. The category "high-level interaction" refers to activities related to *the construction of meaning, such as generating new ideas, elaborating ideas, critiquing ideas, and connecting them to prior knowledge, and which feature high interaction and regulated learning* (e.g., Volet et al., 2009). High-level interaction means that each group member was participating and contributing to the conversation, and regulation of learning became visible through learners' engagement in cognitive activities. In contrast, "confusion" could lead to either high or low levels of interaction, depending on whether it was resolved and regulated (D'Mello & Graesser, 2011)—that is, confusion involves *markers of metacognitive monitoring and prompting of other group members to regulate learning* (Hadwin, Järvelä, & Miller, 2016).

In summary, while high-level interaction related to mutual and shared interaction and regulated learning during collaboration (Van Boxtel 2004; Volet et al., 2009a), low-level interaction exhibited no visible markers of regulation but helped students to share their mutual understanding of topics. Confusion exhibited some markers of metacognitive monitoring, with potential for improved learning (D'Mello & Graesser, 2011; Hadwin, Järvelä & Miller, 2011). Inter-rater reliability for the qualitative content analysis was assessed by calculating Cohen's kappa values based on the data as a whole. For phases of working and interaction type, the kappa value was 0.65. For the unsolved cases, differences were negotiated and the coded video episodes were reviewed again until consensus was reached.

4.3. Facial expression recognition analysis

Facial expression recognition analysis made use of a MORE system tool that automatically detects and analyzes faces visible in the video recording. Faces are detected and tracked throughout each video sequence. Each face is then geometrically normalized using detected eye and mouth points. After normalization, dynamic texture features (LBP-TOP features) are extracted (for more detail, see Zhao & Pietikäinen, 2007), and a support vector machine (SVM) classifier model is used to estimate the valence class for each processed face. The valence model is trained using natural facial expressions from the MAHNOB-Implicit-Tagging database. For each face, the model estimates valence using one of three classes (positive, negative, or neutral), and the outcome is expressed as the frequency of positive, neutral, or negative facial expressions recognized by the system during the one-minute time window.

The reliability and validity of facial expression recognition analysis has been thoroughly evaluated by Zhao and Pietikäinen (2007) and by Huang et al. (2016), where LBP-TOP features are compared with state-of-the-art algorithms. Additionally, the method was evaluated using the Cohn-Kanade facial expression database and achieved a recognition rate of 96.26%. Huang et al. (2016) conducted further experiments on the MAHNOB-Implicit-Tagging database to evaluate the reliability and validity of the valence model. Keskinarkaus et al. (2015) provided more details of system implementation and performance evaluation.

5. Results

5.1. How do phase of working and type of interaction relate to episodes of simultaneous arousal?

The first research question related to phases of working and quality of interaction during episodes of simultaneous arousal. As each episode lasted for one minute, the phase of working and type of interaction represents what was actually going on during these episodes. In total, there were 131 high-arousal episodes that related to phases of working as follows: a) work instructions (f = 37), searching for information (f = 18), communication (f = 33), learning environment (f = 20), adding information (f = 19), and f) off-topic conversation (f = 4). Each of the five activities involved a) low interaction (f = 90), b) high interaction (f = 11), and c) markers of confusion (f = 26). However, off-topic conversations were not associated with interaction type. Each phase of working involved high-arousal episodes during the collaborative task.

Figure 2 details the frequencies of phases of working during the collaborative work and how interaction types were distributed across phases of working as percentages. The calculated distribution of the three interaction types was based on the total duration of the interaction type during each phase of working. Quantifying the proportional distribution of the three interaction types enabled investigation of how the interaction types typically emerged in specific phases for the five collaborative groups. This facilitates a better understanding of phases of working and interaction type when high-arousal occurs simultaneously across group members. Figure 2 shows how types of interaction occur during phases of working.

Figure 2 Here

In each phase of working, *low-level interaction* occurred with the greatest frequency. It was most frequent during the “work instructions” phase (89%), followed by the “adding information” phase (84%) and the “searching for information” phase (67%). It was less frequent in the “learning environment” (50%) and “communication” phases (56%).

Confusion occurred more frequently during the “learning environment” (50%), “communication” (27%) and “searching for information” phases (17%). It occurred less frequently during the “adding information” (6%) and “work instructions” phases (11%).

High interaction occurred more frequently during the “communication” (18%), “searching for information” (16.7%), and “adding information” phases (10.5%). It did not occur at all during the “learning environment” and “work instructions” phases.

To summarize, low-level interaction and confusion featured in all phases of working, but high-level interaction did not. Low-level interaction occurred most frequently during the work instructions phase and when working with the weSpot learning environment. Confusion occurred most frequently when working with the learning environment and during communication. High-level interaction occurred most frequently during communication and when searching for information.

5.2. What types of facial expression (positive, negative, or neutral) are observed when arousal occurs simultaneously among group members?

Since the number of recognized facial expressions during a one-minute time segment varied from zero to 445 (M = 108, SD = 112), the frequency of recognized faces was transformed to percentage form to elaborate in more detail the distribution of facial expressions during the one-minute episodes (Table 3).

Table 3 here

Figure 3 shows the types of facial expression occurring during episodes of simultaneous arousal. The occurrence of negative, neutral, and positive facial expressions is presented in terms of relative time, as these occur in each of the five groups. The Y-axis indicates the percentage occurrence of negative, positive, and neutral facial expressions for all students; the X-axis shows how the negative, positive, and neutral facial expressions occur during the 75-minute collaborative learning session for all students during simultaneous arousal episodes. Figure 3 shows that each of the 131 simultaneous arousal episodes occurring during the 75-minute collaborative learning task included negative, positive, and neutral facial expressions. However, there were five one-minute episodes where no emotions were recognized. The mean percentage occurrence of negative facial expressions during simultaneous arousal episodes was 40% (*SD* = 13); for neutral facial expressions, it was 33% (*SD* = 13); and for positive facial expressions, it was 22% (*SD* = 11). In other words, the collaborating groups made negative faces most frequently during the simultaneous arousal episodes, but each type of facial expression (negative, neutral, and positive) occurred during the simultaneous arousal episodes.

Figure 3 here

Next, to investigate whether there were differences in the facial expressions when the type of interaction (high-level, low-level, confusion) varied, we used interaction as independent variable and percentage of recognized facial expressions as a dependent variable. A one-way ANOVA indicated a significant difference in expression of negative emotions at the level of $p < .005$ for the three types of interaction (F [2,12] = 5.92, $p = .003$). Using the Bonferroni test, post hoc comparisons indicated that the mean score of negative emotions for confusing interactions (M = .490, SD = .03) was significantly different to that for low-level interaction (M = .398, SD = 0.15). However, the mean score for negative emotions during high-level interaction (M = .50, SD = .03) did not differ significantly from low-level or confused interaction episodes. In other words, there was a significant difference between the amount of negative emotions expressed when the interaction was confused and when it was low-level.

5.3.) How does regulation of learning appear during interaction when arousal occurs simultaneously among group members?

To illuminate the different types of interaction, three examples will now be discussed. First, regulation of learning was visible only when the interaction was either high-level or involved confusion—that is to say, 37 of the 131 episodes involved regulation of learning. The three examples here are a illustrative sample (Patton, 2002) that helps to better understand what multichannel data can reveal about regulated learning in a collaborative learning context. Each case represents a different interaction type during a one-minute episode, and in each sample, negative facial expressions occur most frequently.

Example 1. Simultaneous arousal in low interaction during work instructions

In the first example, negative facial expressions were identified 437 times during the one-minute episode. Here, the students have just arrived, and they are listening to the instructions. There is no interaction between the group members during that one minute. During the episode, the instructor explains how to log in to the learning environment (weSPOT) and starts to explain the collaborative learning task.

Instructor: *Ok, and now, when you have all managed to log in to the learning environment, I will briefly explain to you what your collaborative task is about. Your task is to find out what Jussi, a 27-year-old male, should eat for breakfast. Jussi needs to plan his breakfast carefully because he is training for a marathon.*

Altogether, simultaneous arousal occurred during work instructions in 37 episodes; of those, 82% involved low interaction. In this case, it can be concluded that the change in EDA was due to an anticipation effect, which is also identified in earlier research (Smith, 1989). The students did not know what their task was, or what to expect from the experiment, which caused excitement and a change in their EDA during the task instructions.

Example 2. Simultaneous arousal in high interaction during communication

In the second example, negative facial expressions were recognized 357 times during a one-minute episode. The students had already made some progress in the task but had not yet been able to solve it completely. In this example, the group members are negotiating what they should do next and how they should progress with the task.

Iina: *The breakfast should include 25% of all calories needed for the day. So, one quarter of … what was it … 1600 calories?*

Jussi: *If it says 1600, and so far we have 302 calories, it is not working.*

Iina: [laughs] *We have to add a banana.*

Leena: *So is that like a daily estimate or …?*

Iina: [reads out loud] *Breakfast should include 25% of all calories needed for the day.*

Jussi: *So, if 1600 is needed, there should be…*

Iina: Something like 400.

Jussi: So, how many calories are there in a banana?

Iina: We need more calories ...

Altogether, simultaneous arousal occurred during communication in 18 episodes; of those, 17.5% involved high interaction. In Example 2, the collaborating group is discussing aspects of the task. The participants realize how they need to complete the task, and they are also in the process of formulating a solution. In this case, each of the three students is engaged with the task and negotiates a shared understanding in terms of how to progress with the task.

Example 3. Simultaneous arousal related to confusion during communication

In the third example, negative facial expressions were recognized 417 times during the one-minute episode. The students had just started to work with the task, and they were adding information to the learning environment and looking at the template for the first time, which they needed to use to progress with the task.

Satu: *My name is not there! What is your name in here?* [The students can see each other's Google names in the learning environment, but they do not know each other's Google names].

Marja: *Well, your name is not there.*

Satu: *What is your name there?*

Satu: *I just wrote something there ... Should I mark the information here in grams?*

Marja: *I really do not know.*

[Both are laughing.]

Jussi: *Protein.*

Satu: *Is it okay now?*

Maria: *I do not know. Was it 345?*

Maria: *What … or do we have to calculate this somehow? The proportions?*

Satu: *It says here "proportion"* [showing the food packet to Maria].

Maria: *Oh ... I did not look for it there.*

Satu: *That is 138.*

Jussi: *How much protein was there?*

Satu: *How should I know that?*

Maria: *Doesn't it say it there?*

Satu: *It should be here somewhere, perhaps on the other side?*

Jussi: *Or maybe we need to calculate that?*

In total, simultaneous arousal occurred during communication in 18 episodes, of which 26.5% involved confusion. In Example 3, the collaborating group is confused about how to use the learning environment, what information is relevant, and where to find it.

6. Discussion

This study aimed to understand what multichannel data such as EDA signals, facial expression data, and video observations can reveal about regulated learning in the context of collaborative learning. In the analysis, particular attention was paid first to identifying arousal episodes that occurred simultaneously among two or three group members during one-minute time windows. First, the phases of working in which high-arousal occurred were identified, as were the types of interaction (low, high, or confusion) involved in simultaneous arousal episodes. Second, the types of facial expression (negative, positive, and neutral) involved in simultaneous arousal episodes were investigated, along with any differences in facial expression when the type of interaction varied.

Third, examples were drawn from the data to better illustrate when and how to make connections for SRL in situations where negative facial expressions occurred with simultaneous arousal.

The results indicate that simultaneous arousal occurred among group members in each phase of working. However, the groups mostly experienced simultaneous arousal during task instructions. Earlier research has shown that when the learners are not sure about what to expect, this can affect arousal reflected in their EDA levels (Smith, 1989). To avoid this anticipation effect, measures of EDA should perhaps be taken into account after the learners have acquired sufficient information about what they are expected to do.

For the most part, the interaction was low-level in each phase of working. Based on models of regulated learning, especially when considering interaction (Hadwin et al., 2016), it can be concluded that there was hardly any “sharing” during low-level interaction. This may also mean that group members were focused and engaged in progressing with the task individually during low-level interaction (Järvelä, Järvenoja, Malmberg, Isohätälä, & Sobocinski, 2016). As individual students’ contributions are needed for collaborative learning (Zhao & Chan, 2014), low-level interaction should not be discouraged, as this also helps group members to accomplish their learning goals.

High-level interaction occurred least frequently during simultaneous arousal episodes. Previous research has indicated that changes in EDA are elicited during stressful tasks (e.g., Bandura, 1982; Pecchinenda & Smith, 1996; Tomaka, Blascovich, Kelsey, & Leitten, 1993). In other words, situations that are considered challenging typically trigger increased activity in the sympathetic nervous system, as reflected in EDA (Giromini et al., 2016). However, changes in EDA may also be caused by engagement or interest (Henriques, Paiva, & Antunes, 2013). On the basis of these qualitative examples, the results suggest that students were engaged in collaborative learning through interaction during episodes of high-level interaction, and it is possible to locate SRL from this type of interaction. Similar findings were reported by Isohätälä, Järvenoja, and Järvelä (2016) in their analysis of students' participation in collaborative interaction. Investigating how regulation of learning emerged during the fluctuation of participation in interaction, they noticed that active participation helped in the emergence of regulated learning, which in turn contributed to the coordination of collaboration.

Interactions that included confusion were detected in 26 episodes. On the basis of theoretical models of SRL (e.g., Hadwin et al., 2016; Winne & Hadwin, 1998), this type of interaction seems to activate SRL in collaborative contexts (Malmberg, Järvelä, & Järvenoja, 2016), as it invites students to activate regulation. The learners most often made negative faces during episodes that included markers of confusion, which may shed some light on the socio-emotional interactions that occur when a situation is confusing. That is, students did not engage in positive socio-emotional interactions with each other, which also manifests as low-level interaction between group members (Kwon, Liu & Johnson, 2014). This finding may also reflect the study's exploratory setting. As the students collaborated in groups that were determined beforehand, the data collection setting was artificial, and perhaps the students were not given sufficient space to establish a safe socio-emotional space for collaborative learning (Fransen, Kirschner, & Erkens, 2011). However, the qualitative examples show that the episodes labeled confusing included markers of metacognitive monitoring (Winne & Hadwin, 1998), which suggests that when EDA rises and learners make negative faces, markers of regulated learning may be identified following those episodes.

6.1. Conclusion

By using multichannel data during collaborative learning, this work extends the possibility of recognizing either regulated learning episodes or episodes that call for regulated learning. However, of the 131 episodes, 37 involved regulated learning. Although this study was exploratory, and the sample size was small, the results offer some guidelines for future research into SRL. First, while there is a possibility that EDA can be used to detect arousal events involving regulated learning, it cannot directly measure or locate episodes where SRL is needed, and this approach to researching regulated learning is still in its early stages. Second, although physiological data can be a useful indicator of student reactions in learning situations (Gillies et al., 2016; Worsley & Blikstein, 2015), it is essential to contextualize carefully those episodes where simultaneous arousal occurs among group members. It is important to ask what types of interaction occur during those episodes, especially if the goal is to develop new methodological solutions for capturing regulated learning "on the fly" and to find new ways of utilizing the power of learning analytics to support regulated learning as it occurs (Worsley & Blikstein, 2015). Despite the emergence of new data-capturing devices, the analytical procedure remains very extensive and painstaking despite the small sample size. In order to understand the relation between EDA data and regulated learning, different methodological solutions, such as the Hidden Markow Models, could be used to better understand the underlying structure of the data.

By leveraging new data-capturing devices such as Empatica E3 sensors and facial expression recognition software, we were able to identify how simultaneous arousal relates to phases of working and types of interaction during collaborative learning. We illuminated the occurrence of simultaneous arousal episodes and the types of facial expression that related to those episodes. We established that a combination of multiple data channels has the potential for use as a method of capturing regulated learning during collaboration. The examples show that it is possible to make careful connections for regulated learning, especially in the episodes that included either high-level interaction or confusion. Future efforts should analyze the correlation of simultaneous arousal and instances where multiple students contribute to joint interactions. Promising new technologies can be used not only to detect when multiple students contribute to joint discussions but also to recognize emotional states from students' speech acts (Väyrynen, Kortelainen, & Seppänen, 2013).

This study demonstrates that new technologies such as facial recognition software and physiological signals can be used unobtrusively in collaborative learning contexts to capture episodes that may reveal regulated learning or the need to activate it. In the future, this type of multichannel data should be used for validating contextual data, and students' own interpretations of simultaneous arousal episodes would provide valuable insights for data interpretation. To that extent, this work represents an important first step toward more reliable and objective measurement of regulated learning in collaborative contexts.

One limitation of this study was the small sample size, which hinders generalization which is a result of bad quality of physiological data. The bad quality of the physiological data was due mainly to the fact that sensors were not attached properly at the beginning of the study or became loose as the task progressed. In addition, as the unit of analysis was the group as a whole, good quality data were needed from all the students in the same group. However, even small samples are very labor-intensive, as there is little if any empirical research in the field of SRL that could be benchmarked for analysis or comparison with our findings. Today, facial expression recognition is an active research area in signal processing and computer vision (Huang, Zhao, Hong, Zheng, & Pietikäinen (2016), and further multidisciplinary collaboration in the learning sciences, machine vision, and signal processes may offer promising new methodological solutions that can illuminate SRL in new ways in the future.

References

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, *50*(1), 84–94. http://doi.org/10.1080/00461520.2015.1004069

Azevedo, R., Johnson, A., Chauncey, A., & Graesser, A. (2011). Use of hypermedia to convey and assess self-regulated learning. In B. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 393–407). New York, NY: Routledge.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122–147. http://doi.org/10.1037/0003-066X.37.2.122

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences, 12*, 307–359.

Blascovich, J., Vanman, E., Mendes, W., & Dickerson, S. (2011). Social psychophysiology for social and personality psychology. UK: SAGE Publications. http://doi.org/10.4135/9781446287842

Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, *47*(4), 647–658. http://doi.org/10.1111/j.1469-8986.2009.00972.x

Boekaerts, M., & Niemivirta, M. (2000). Self-regulated learning. Finding a balance between learning goals and ego-protective goals. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 417–450). San Diego: Academic Press.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. http://doi.org/10.1111/j.1469-8986.2012.01384.x

Buder, J., & Bodemer, D. (2008). Supporting controversial CSCL discussions with augmented group awareness tools. *International Journal of Computer-Supported Collaborative Learning, 3*, 123–139. http://doi.org/10.1007/s11412-008-9037-5

Cleary, T. J., Callan, G. L., & Zimmerman, B. J. (2012). Assessing self-regulation as a cyclical, context-specific phenomenon: Overview and analysis of srl microanalytic protocols. *Education Research International, 2012*, 1–19. http://doi.org/10.1155/2012/428639

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative learning: cognitive and computational approaches* (pp. 1–19). Oxford: Elsevier.

D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion, 25*(7), 1299-1308

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153–170.

Eleftheriadis, S., Rudovic, O., & Pantic, M. (2015). Discriminative shared Gaussian processes for multiview and view-invariant facial expression recognition. *IEEE Transactions on Image Processing*, *24*(1), 189–204,

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*(4), 238–252.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Investigator's guide*. Palo Alto, CA: Consulting Psychologists Press.

Elkins, A. N., Muth, E. R., Hoover, A. W., Walker, A. D., Carpenter, T. L., & Switzer, F. S. (2009). Physiological compliance and team performance. *Applied Ergonomics, 40*(6), 997–1003. http://doi.org/10.1016/j.apergo.2009.02.002

Erkens, G., Jaspers, J., Prangsma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior*, *21*(3), 463–486. https://doi.org/10.1016/j.chb.2004.10.038

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, *56*(2), 171–184. http://doi.org/10.1016/j.ijpsycho.2004.11.003

Fontaine, J.R., Scherer, K.R., Roesch, E.B., & Ellsworth, P. (2007). The world of emotion is not two-dimensional. *Psychological Science, 18*, 1050–1057.

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

Grau, V., & Whitebread, D. (2012). Self and social regulation of learning during collaborative activities in the classroom: The interplay of individual and group cognition. *Learning and Instruction*, *22*(6), 401–412. http://doi.org/10.1016/j.learninstruc.2012.03.003

Hackman, R., & Morris, C (1975). Group tasks, group interaction process, and group performance effectiveness. *Advances in Experimental Social Psychology*, 8, 45–99. https://doi.org/10.1016/S0065-2601(08)60248-8

Hadwin, A. F., Järvelä, S. & Miller, M. (2017). *Understanding the role of self-regulation, co-regulation and shared regulation in collaborative learning environments*. Manuscript in preparation.

Hadwin, A., & Oshige, M. (2011). Socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record*, *113*(2), 240–264.

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). Computers in human behavior: A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system q. *Computers in Human Behavior*, *48*, 615–625. http://doi.org/10.1016/j.chb.2015.02.013

Henriques, R., Paiva, A., & Antunes, C. (2013). Accessing emotion patterns from affective interactions using electrodermal activity. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 43–48. http://doi.org/10.1109/ACII.2013.14

Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., & Chaparro-Peláez, J. (2014). Applying social learning analytics to message boards in online distance learning : A case study. *Computers in Human Behavior, 47*(2015), 68–80. http://doi.org/10.1016/j.chb.2014.10.038

Hockey, R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1–3), 73–93. http://doi.org/10.1016/S0301-0511(96)05223-4

Huang, X., Zhao, G., Hong, X., Zheng, W., & Pietikäinen, M. (2016). Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing*, *175*, 564–578.

Huang, X., Kortelainen, J., Zhao, G., Li, X., Moilanen, A., Seppänen, T., & Pietikäinen, M. (2016). Multi-modal emotion analysis from facial expressions and electroencephalogram. *Computer Vision and Image Understanding*, *147*, 114–124.

Iiskala, T., Vauras, M., Lehtinen, E., & Salonen, P. (2011). Socially shared metacognition of dyads of pupils in collaborative mathematical problem-solving processes. *Learning and Instruction*, *21*(3), 379–393. https://doi.org/10.1016/j.learninstruc.2010.05.002

Immordino-Yang, M. H., Christodoulou, J. A., Pekrun, R., & Linnenbrink-Garcia, L. (2014). Neuroscientific contributions to emotion measurement in educational contexts. In

In R. Pekrun & L.Linnenbrink-Garcia (Eds.) *Handbook of emotions in education* (pp. 607–624). New York: Routledge.

Isohätälä, J., Järvenoja, H., & Järvelä, S. (2016). Socially shared regulation of learning and participation in social interaction in collaborative learning. Submitted.

Johnson, A. M., Azevedo, R., & D'Mello, S. K. (2011). The temporal and dynamic nature of self-regulatory processes during independent and externally assisted hypermedia learning. *Cognition and Instruction*, *29*(4), 471–504. http://doi.org/10.1080/07370008.2011.610244

Järvelä, S., & Hadwin, A. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist*, *48*(1), 25–39. doi:10.1080/00461520.2012.750227

Järvelä, S., Järvenoja, H., Malmberg, J., & Hadwin, A. F. (2013). Exploring socially shared regulation in the context of collaboration. *Journal of Cognitive Education and Psychology*, *12*(3), 267–286.

Järvelä, S., Järvenoja, H., Malmberg, J., Isohätälä, J., & Sobocinski, M. (2015). How do types of interaction and phases of self-regulated learning set a stage for collaborative engagement? *Learning and Instruction*, *43*(2016), 39–51. http://doi.org/10.1016/j.learninstruc.2016.01.005

Järvelä, S., Malmberg, J., Sobocinski, M., Haataja, E., & Kirschner, P. (2016). What multimodal data can tell us about the self-regulated learning process. Submitted.

Järvenoja, H., Järvelä, S., & Malmberg, J. (2017). Supporting groups' emotion and motivation regulation during collaborative learning. *Learning and Instruction*, in press. http://doi.org/10.1016/j.learninstruc.2017.11.004

Keskinarkaus, A., Huttunen, S., Siipo, A., Holappa, J., Laszlo, M., Juuso, I., ... Laukka, S. (2015). MORE—a multimodal observation and analysis system for social interaction research. *Multimedia Tools and Applications*, *75*(11) 1–25.

Kory, J., & D'Mello, S. (2015). A review and meta-analysis of multimodal affect detection. *ACM Computing Surveys, 47*(3). http://doi.org/10.1145/2682899

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, *19*(3), 335–353. http://doi.org/10.1016/S0747-5632(02)00057-2

Ku, H. Y., Tseng, H. W., & Akarasriworn, C. (2013). Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning. *Computers in Human Behavior*, *29*(3), 922–929. doi:10.1016/j.chb.2012.12.019

Kwon, K., Liu, Y. H., & Johnson, L. P. (2014). Group regulation and social-emotional interactions observed in computer supported collaborative learning: Comparison between good vs. poor collaborators. *Computers and Education*, 78. http://doi.org/10.1016/j.compedu.2014.06.004

Linnenbrink-Garcia, L., & Pekrun, R. (2011). Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology*, *36*(1), 1–3. http://doi.org/10.1016/j.cedpsych.2010.11.004

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 lowperforming groups. *Computers in Human Behavior*, *52*, 562–572. http://doi.org/10.1016/j.chb.2015.03.082

Malmberg, J., Järvelä, S., & Järvenoja, H. (2016). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. Submitted.

Mendes, W. B. (2009). Assessing autonomic nervous system activity. In E. Harmon-Jones & J. S. Beer (Eds.), *Methods in social neuroscience* (pp. 118–147). New York: Guilford Press.

Mikroyannidis, A., Okada, A., Scott, P., Rusman, E., Specht, M., Stefanov, K., … Chaimala, F. (2013). WeSPOT: A personal and social approach to inquiry-based learning. *Journal of Universal Computer Science*, *19*(14), 2093–2111. http://doi.org/10.3217/jucs-019-14-2093

Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior, 52*, 573–588. http://doi.org/10.1016/j.chb.2015.01.050

Molenaar, I., & Chiu, M. M. (2014). Dissecting sequences of regulation and cognition: Statistical discourse analysis of primary school children's collaborative learning. *Metacognition and Learning*. http://doi.org/10.1007/s11409-013-9105-8

Morris, R., Church, H., Hadwin, A. F., Gress, C. L. Z., & Winne, P. H. (2010). The use of roles, scripts, and prompts to support CSCL in gStudy. *Computers in Human Behavior*, *26*, 815–824.

Munro, L. L., Dawson, M. E., Schell , A. M., & Sakai, L. M. (1987). Electrodermal lability and rapid performance decrement in a degraded stimulus continuous performance task. *Journal of Psychophysiology*, *1*, 249–257.

Nikula, R. (1991). Psychological correlates of nonspecific skin-conductance responses. *Psychophysiology*, *28*(1), 86–90. http://doi.org/10.1111/j.1469-8986.1991.tb03392.x

Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., & Goodwin, M. S. (2016). Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review, 21*(2):99–141. http://doi.org/10.1177/1088868316628405,

Pecchinenda, A. (1996). The affective significance of skin conductance activity during a difficult problem-solving task. *Cognition & Emotion, 10*(5), 481–504. http://doi.org/10.1080/026999396380123

Phielix, C., Prins, F. J., Kirschner, P. a., Erkens, G., & Jaspers, J. (2011). Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool. *Computers in Human Behavior, 27*(3), 1087–1102. http://doi.org/10.1016/j.chb.2010.06.024

Pintrich, P. R., Smith, D. A., Garcia, T. & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated for Learning Strategies Questionnaire (MSLQ). *Education and Psychological Measurement*, *53*(3), 801–814.

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. http://doi.org/10.1111/bjet.12146

Schutz, P. A., & Davis, H. A. (2000). Emotions and self-regulation during test taking. *Educational Psychologist*, *35*(4), 243–256. http://doi.org/10.1207/S15326985EP3504_03

Teasley, S. D., & Roschelle, J. (1995). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S.P.Lajoie & S.J Derry (Eds.), *Computers as cognitive tools* (pp.229-257). Hillsdale, NJ: Lawrence Erlbaum.

Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology*, *65*(2), 248–260.

Valstar, M., & Pantic, M. (2006). Fully automatic facial action unit detection and temporal analysis. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)* (pp. 149–149). IEEE.

Van Boxtel, C. (2004). Studying peer interaction from three perspectives: The example of collaborative concept learning. In J. van der Linden and P. Renshaw (Eds.), *Dialogic Learning* pp. 125–143.Kluwer:Academic Publishers, Dordrecht.

Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction*, *19*(2), 128–143. http://doi.org/10.1016/j.learninstruc.2008.03.001

Volet, S., Vauras, M., & Salonen, P. (2009). Self- and social regulation in learning contexts: An integrative perspective. *Educational Psychologist*, *44*(4), 215–226. http://doi.org/10.1080/00461520903213584

Väyrynen, E., Kortelainen, J., & Seppänen, T. (2013). Classifier-based learning of nonlinear feature manifold for visualization of emotional speech prosody. *IEEE Transactions on Affective Computing*, *4*(1), 47–56.

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior, 26*(4), 506–515.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, *45*(4), 267–276. http://doi.org/10.1080/00461520.2010.517150

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah: Lawrence Erlbaum.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 1–38). New York, NY: Lawrence Erlbaum Associates.

Acknowledgements: This study was supported by the Finnish Academy grant, 275440, SLAM research project

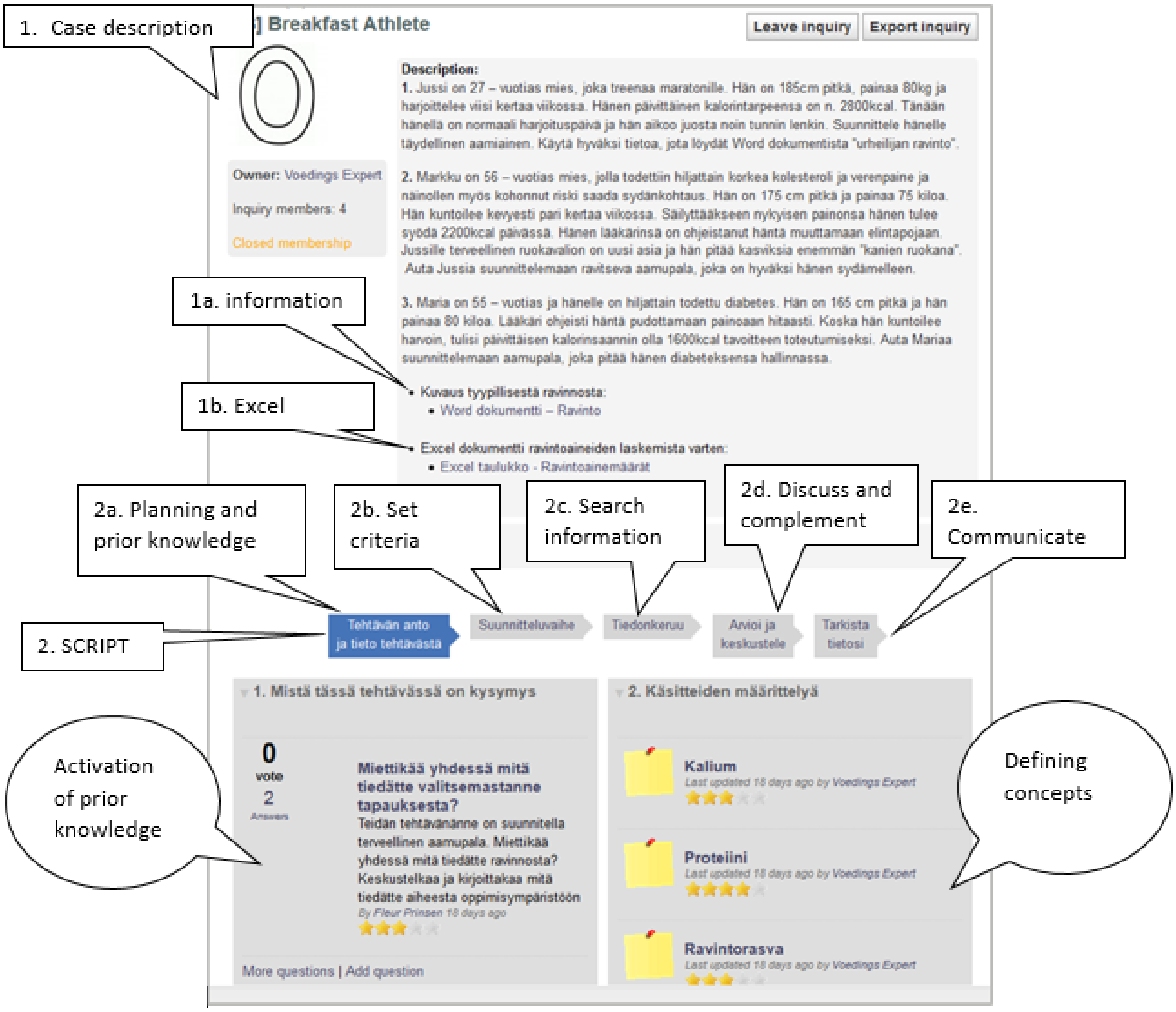


Figure 1. Screen snapshot of WeSpot Learning environment

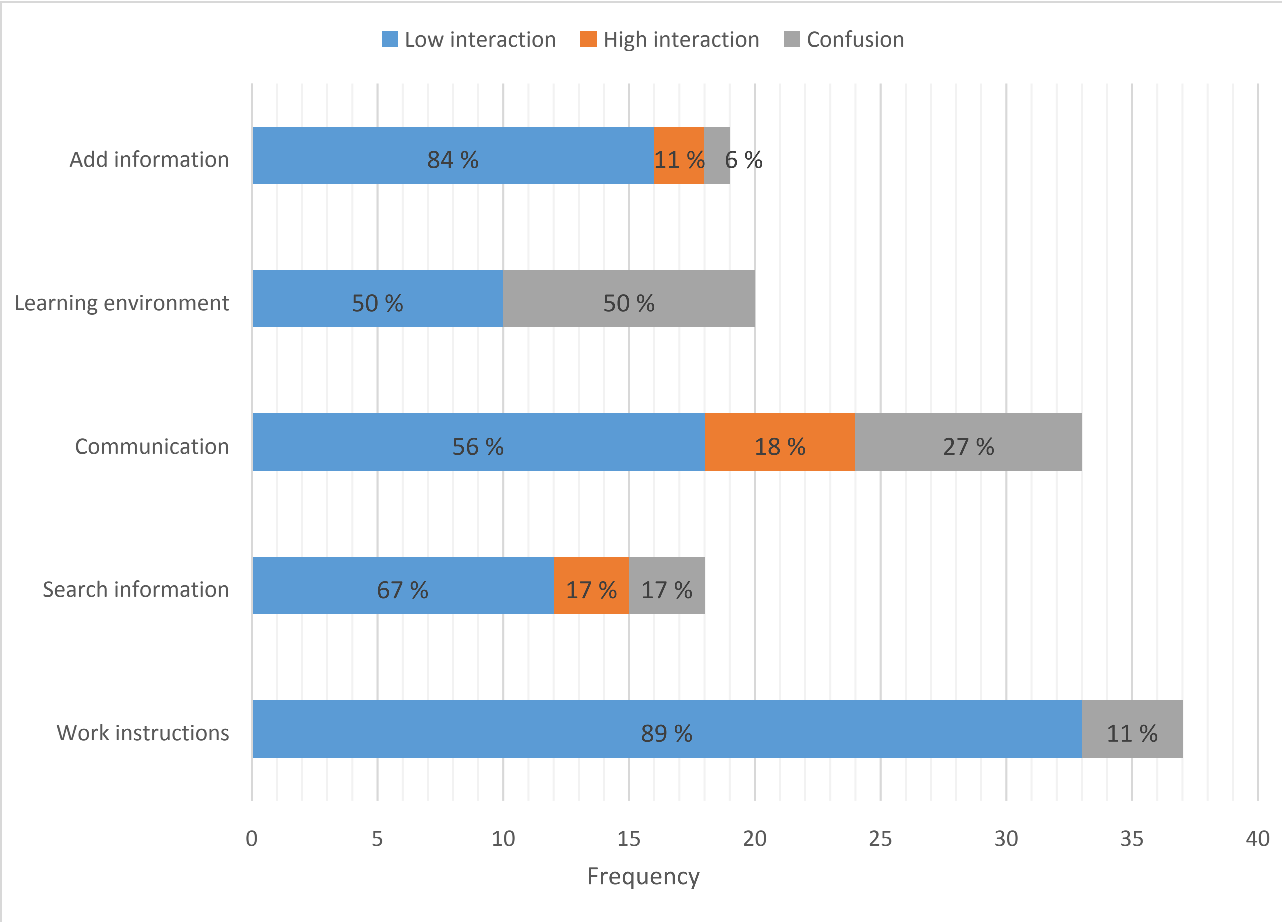


Figure 2. Overview of association between phase of working and quality of interaction.

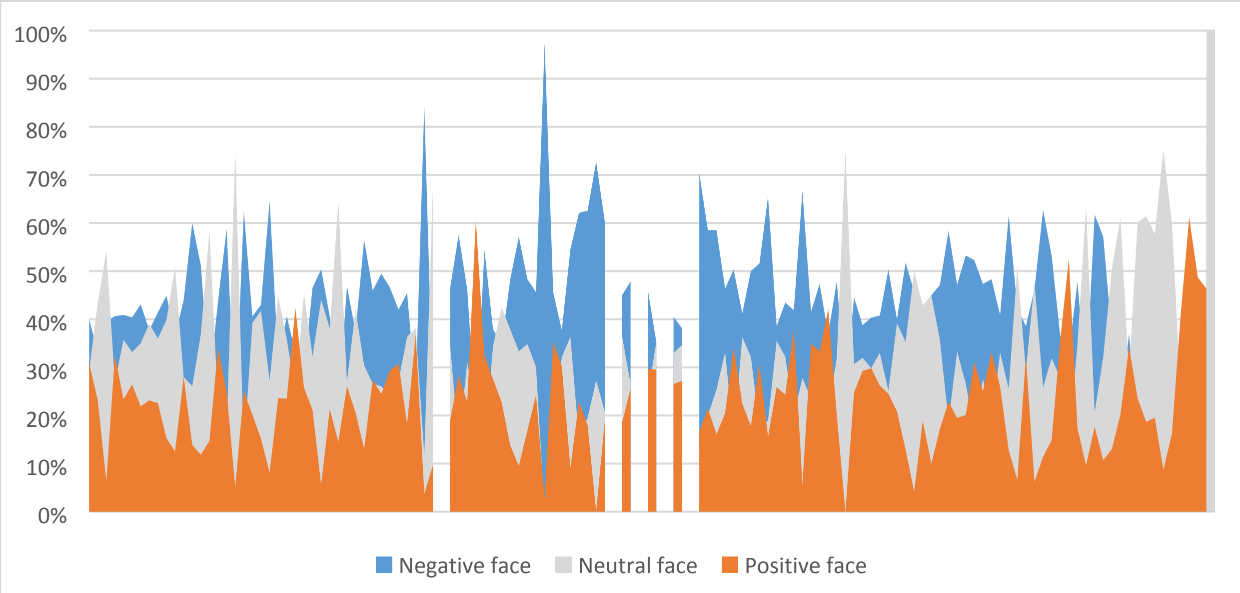


Figure 3. Occurrence of negative, neutral, and positive facial expressions during simultaneous arousal episodes.

- Simultaneous arousal episodes occurred throughout phases of collaborative learning
- Interaction was high, low or confused during simultaneous arousal episodes
- Negative faces occurred the most during the simultaneous arousal episodes
- Regulated learning can be found from simultaneos arousal episodes

Table 1. *Phases of working*

Phase	Description
Work instructions	Students listen to instructions about how they should proceed with the task.
Learning environment	Students use or discuss aspects of the weSpot learning environment and how to use the learning environment, reading and interpreting the task assignments according to the script.
Searching for information	Students search the Internet for information relevant to task completion (e.g., websites or the resources provided).
Adding information	Students have found the information relevant to the task's completion and add it to the Excel sheet.
Communication	Students discuss the contents of the task and the criteria for accomplishing the task.
Off-topic discussions	Students discuss irrelevant issues rather than working with the task.

Table 2. *Descriptions of qualitative modifiers*

Interaction type	Description of interaction type
Low-level interaction	Reading out loud or silently, presenting what has just been read, agreeing with group members' thoughts. Passive activity with low interaction. Regulation of learning is not visible.
High-level interaction	Discussing and sharing ideas, asking questions or asking for justification. Active interaction between group members. Regulation of learning is visible through interaction
Confusion	Markers of confusion. Not sure what to do, group members hesitate, constantly express uncertainty or say they do not know what to do. Markers of metacognitive monitoring.

Table 3. *Proportional distributions of recognized facial expressions during one-minute episodes*

Facial expressions	Min	Max	Mean
Negative facial expressions	0	97%	43%
Neutral facial expressions	0	75%	32%
Positive facial expressions	0	61%	21%