Investigating collaborative learning success with physiological coupling indices based on electrodermal activity

Héctor J. Pijeira-Díaz
University of Oulu
Oulu, Finland
Hector.PijeiraDiaz@oulu.fi

Hendrik Drachsler
Welten Institute
Heerlen, The Netherlands
Hendrik.Drachsler@ou.nl

Sanna Järvelä
University of Oulu
Oulu, Finland
Sanna.Jarvela@oulu.fi

Paul A. Kirschner
Welten Institute
Heerlen, The Netherlands
Paul.Kirschner@ou.nl

ABSTRACT

Collaborative learning is considered a critical 21st century skill. Much is known about its contribution to learning, but still investigating a process of collaboration remains a challenge. This paper approaches the investigation on collaborative learning from a psychophysiological perspective. An experiment was set up to explore whether biosensors can play a role in analysing collaborative learning. On the one hand, we identified five physiological coupling indices (PCIs) found in the literature: 1) Signal Matching (SM), 2) Instantaneous Derivative Matching (IDM), 3) Directional Agreement (DA), 4) Pearson's correlation coefficient (PCC) and the 5) Fisher's z-transform (FZT) of the PCC. On the other hand, three collaborative learning measurements were used: 1) collaborative will (CW), 2) collaborative learning product (CLP) and 3) dual learning gain (DLG). Regression analyses showed that out of the five PCIs, IDM related the most to CW and was the best predictor of the CLP. Meanwhile, DA predicted DLG the best. These results play a role in determining informative collaboration measures for designing a learning analytics, biofeedback dashboard.

CCS Concepts

•**Information systems** → **Data analytics;** *Data mining;* •**Applied computing** → **Collaborative learning;** •**General and reference** → *Empirical studies; Metrics;*

Keywords

learning analytics, biosensors, electrodermal activity, collaborative learning, physiological coupling indices

1. INTRODUCTION

Collaboration is regarded as a critical 21st century skill [53]. Labor market demands team players as today's challenges require professionals to work with each other. The earlier students start developing collaboration skills the sooner they will be able to incorporate them to everyday life and work. The interest in collaboration therefore concerns the students of today and workforce of tomorrow, teachers and researchers as the most direct stakeholders. Studies on collaborative learning have shown that collaboration promotes important cognitive activities, such as question asking, elaborating own understanding, argumentation of various points of views and conclusion making [34]. Research has also shown that collaborative learning is not self-evident and that true collaboration is infrequent [3]. Reaching high-level cognitive processing is a demanding task [29], and cognitive challenges tend to cause tension and increase the potential for both socio-emotional conflicts and poor collaborative learning experiences [36, 2]. More is needed to know about these socio-emotional and cognitive interaction processes, and new methodological solutions may offer what the earlier have not achieved. The measurement of collaboration in general is a challenge as such, as there are not enough indicators available so far. Operationalizing collaboration is not a straightforward process and there is not a single way to do it [16].

This paper approaches the collaboration quantification through Learning Analytics (LA) [54] and physiological data from biosensors. LA is the measurement, collection, analysis and reporting of learning data. It leverages information visualization, learning sciences, software engineering, statistics and data mining methods for providing feedback and awareness of the learning process. The interest in LA concerns a broad variety of roles: "researchers in education, leaders and policy-makers, educational practitioners, organisational administrators, instructional designers, product vendors, and [...] the learners themselves" [7].

The LA field started to develop from the bulk of digital traces that came with the Massive Open Online Courses (MOOCs). Together with the open access to high quality educational resources, the MOOCs' user-transparent tracking capabilities started to produce learning datasets of unprecedented dimensions. To tackle the necessity of turning the

raw data into useful information and present it in friendly, relevant ways, the multidisciplinary field of LA emerged.

LA techniques are a powerful tool for sense-making of educational data [49]. This study on collaborative learning features applies LA to an increasingly available [51] data modality: physiological data. Activity from the cardiovascular and electrodermal systems — just to name the two most popular physiological data sources in psychological studies [30] — has been linked to diverse cognitive, emotional and behavioural processes at both within- and between-subject research. We believe that tracking physiological data in learning situations can enrich the LA field, which has been mainly focused so far on digital traces.

Concurrently, collaborative learning research has relied on analysing transcribed verbal interactions and using self-reporting questionnaires [16]. There is a potential for complementing subjective reports with objective physiological data [11]. Wearable biosensors are gaining popularity in everyday life mostly for sports and fitness activity tracking. Despite not being their primary purpose to date, these biosensors can also be used to study the cognitive and affective domains of learning [46]. Physiological measures can be informative both at individual and group levels, each having a set of usages [11]. Therefore, we envision an application to monitor the learning process with wearables to provide timely feedback to the teacher and/or the learner. Before that promising future, extensive research is needed in three dimensions: 1) the physiological responses relevant for the learning process, 2) the significant features extractable from them, and 3) the learning-related psychological states they can describe. The field has been called Wearable Enhanced Learning[1] (WELL) and the architecture has been referred to as sensor-based platforms [46].

There is a number of physiological responses. The multi-sensor wristband used in the experiment allowed for the recording of Blood Volume Pulse (BVP), electrodermal activity (EDA) and skin temperature. This study considers EDA. It is an easy measurable, sensitive response [12] that previous studies have linked to arousal [11, 38], attention [55], task engagement [40], and cognitive load [48], among others.

The notion of a relationship between the physiological responses of two individuals interacting with each other has been long scientifically explored [31]. Little or no agreement on how to call this physiological relationship has been reached. The diverse nomenclature found includes (chronologically sorted): physiological linkage [31], physiological compliance [50], physiological synchronization [24], physiological correlation [15], joint changes in the physiological signals [25], physiological coupling [11] and physiological markers of togetherness [39]. From now on, this paper uses *physiological coupling* (PC) for the idea of influence it suggests. PC has been proposed for the assessment of social interactions and as a direct feedback [11].

To measure PC several indices have been used [35], six to our knowledge. [15] presented Signal Matching (SM), Instantaneous Derivative Matching (IDM) and Directional Agreement (DA). Pearson's correlation coefficient (PCC) has also been employed [23] as well as a variation: its Fisher's z-transform (FZT) [9] (i.e. the inverse hyperbolic tangent of the PCC). Weighted coherence (WC) was developed by [44]

---

[1] http://ea-tel.eu/special-interest-groups/well (Last accessed: 20/10/2015)

for frequency domain analysis.

This study obtained five of those PCIs — SM, IDM, DA, PCC and FZT — from pairwise EDA time series to explore for the first time their ability to predict three collaborative learning features: 1) collaborative will, 2) collaborative learning product and 3) dual learning gain. The best predictor for each was determined among the five PCIs.

Our near future goal is the design and development of a biofeedback dashboard for monitoring collaborative learning with physiological data. Thus, apart from its value in the quantification of three collaboration features, this exploration is useful as a first step in determining those PCIs more informative for the dashboard and what they may signal.

The remainder of this paper is as follows. Section 2 motivates the physiological signals application to learning. Section 3 details the experimental design, data collection and processing. The results coming out of the analysis are discussed in section 4. Finally, conclusions and agenda points for further research are stated in section 5.

2. RATIONALE

Biosensors are in a process of democratization. General-purpose, low-cost, biosensors are increasingly available and becoming part of everyday life [51]. Technology has made possible the integration of a variety of sensors in a single device thanks to the progress in nanotechnology, low-power electronics and biosensor design [21]. Final products in this category include, but are not limited to, smartwatches, wristband sensors, wearable sensor patches, and smartclothes. The multi-sensor approach is a trend in the industry and the combination of cardiovascular activity, temperature, motion and EDA sensors has been regarded as the next-generation multi-sensor standard [51].

The physiological responses that biosensors record are originated in the nervous system, of which a simplified view is presented in figure 1. There are two big groups of physiological responses depending on whether they are controlled by the Central Nervous System (CNS) or the Autonomic Nervous System (ANS). The CNS controls the brain and the spinal cord. The ANS innervates the electrodermal, cardiovascular and respiratory systems. ANS responses offer a great potential for research and physiological computing since their measurement mechanisms are cheaper, faster and more unobtrusive than those of the CNS responses [38]. However, their easy measurement contrasts with the challenge that their intricate interpretation represents. Interpretation is inextricably linked to context as several stimuli can produce the same physiological responses [8].
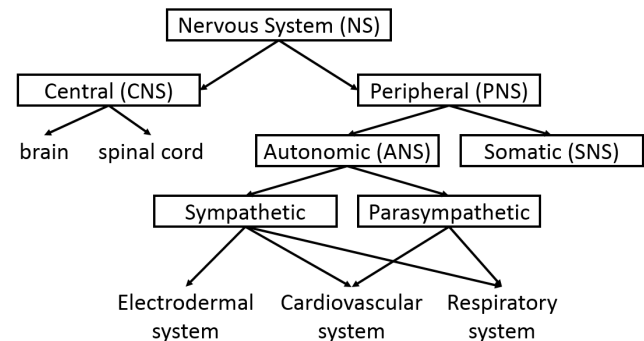

Figure 1: Nervous system simplification.

This paper considers the electrodermal system responses during a collaborative learning task. EDA phenomenon and measurement were first proposed in 1888 by Féré, who discovered a change in skin electric conductance elicited by external stimuli [18]. The increase in skin conductivity is caused by a sweat increment. Sweat production is induced by thermoregulatory and psychological mechanisms as well [12]. Therefore, a constant room temperature of 23 °C is recommended [5] to measure EDA produced by psychological sweat, the one of interest in psychophysiology.

Psychophysiology is the discipline that studies the inference of psychological states from physiological responses. The psychological states comprise cognitive, emotional, and behavioural phenomena [8]. The first systematic work in psychophysiology is considered to be the 1915 publication "Bodily changes in pain, hunger, fear and rage" by the American physiologist Walter Bradford Cannon, a pioneer in the study of the physiology of emotions [6]. Science is since looking for an improved understanding of the physiological footprints that psychological mechanisms leave.

The psychophysiological process is depicted in figure 2. The human body produces physiological responses as a result of its own functioning or upon external stimuli. The physiological responses can be recorded into physiological signals by means of biosensors. Different sampling rates (i.e. number of measurements per time unit) are used depending on the response and the application (see table 4 for the ones used in the experiment). Examples of physiological signals are EDA, electrocardiogram (ECG), electrooculogram (EOG), electromiogram (EMG) and electroencephalogram (EEG), among others. The same physiological response can be recorded as different physiological signals depending on the biosensor used. Thus for example, cardiovascular activity can be recorded as ECG or as BVP signals depending on whether an electrocardiograph or photoplethysmograph sensor are used respectively.

Every signal in turn has a number of physiological features. The features can be descriptive statistics such as the mean and the standard deviation, or signal-specific such as the heart rate extracted from either the ECG or BVP signals, or the Skin Conductance Level (SCL) and Skin Conductance Response (SCR) extracted from the EDA signal. Statistical features are clearly common to all the signals, while the number of signal-specific features varies from signal to signal. Typically, studies use from a couple to less than a dozen features, but at least 128 features from 4 signals have been used [19]. This provides a rough idea on how many features are extractable from physiological signals. Physiological features can be extracted both at individual and group levels. The physiological features are the final numbers to be interpreted so as to determine the psychological state. Either the number or the interpretation could be provided as biofeedback in applications.

A scientific review of research on the physiology of emotion [30] reported that the most common physiological responses used in the field are heart rate and EDA in that order.

There is widespread documentation on EDA proportionality to arousal (psychological activation) [38]. Significant differences in EDA have been observed between pleasant and unpleasant stimuli, that is, it can indicate valence [11]. EDA is highly sensitive to the novelty, intensity and perceived significance of stimulus [55]. An increase in cognitive load has also been reported to have a match in EDA rise. These are
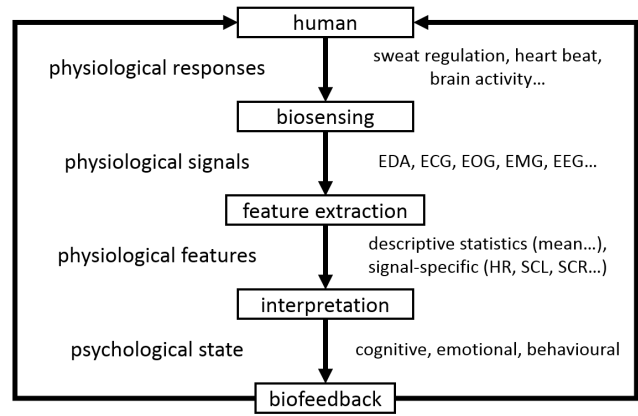

Figure 2: Biofeedback process (adapted from [38]).

just examples of EDA expressiveness at the individual level.

The computer science field has been attracted to psychophysiology. *Affective computing* [41] is devoted to the variety of applications of machines able to recognize and properly respond to human emotions. It is based on the principle that no intelligent system is possible without emotions. *Physiological computing* [17] is a generalization of the *affective computing* field to include also cognitive computing, aimed at increasing user performance in contrast to the improved user satisfaction goal of *affective computing*. Although the very high complexity of psychophysiological inference and validation remains a hurdle, it has been suggested that even low accuracy psychophysiological inferences may be informative [11]. Potential applications would be ground-breaking and paradigm-changing [38], and learning might not be an exception.

Applications to the learning area include, but are not limited to, interaction support [11], feedback [46], intelligent tutors [1], the current educational challenge of formative assessment [45], intervention [14], teacher training [22], emotion recognition in e-learning [47] and self-regulated learning [37]. A recent review on sensor-based platforms [46] did not identify prototypes purposely focused on learning applications. Nevertheless, most prototypes were found to have capabilities relevant to at least one of the three learning domains [4]: cognitive, affective and psychomotor. Therefore, there is a need for and a lack of research on how can biosensors leverage the learning process [46, 47] at both individual and group scales.

PCIs and collaboration

At group level the physiological features are the PCIs, different indices to measure the relationship between the physiological signals of individuals within the group. Six PCIs have been used in research to our knowledge: SM, IDM, DA, PCC, FZT and WC. Signal Matching (SM) accounts for the pairwise difference between the signals once they have been normalized. The need for normalization comes from the fact that physiological signal levels are strongly dependent on individual characteristics [12]. Instantaneous Derivative Matching (IDM) compares the rate of change of physiological signals from a dyad by means of the derivative. Directional Agreement (DA) represents the per cent of data points going in the same direction across individuals (i.e. going up, down or staying constant at the same time). It is the most basic PCI. The Pearson's correlation coefficient

(PCC) determines the strength of the linear relationship between the physiological signals from two individuals. The Fisher's z-transform (FZT) is a transformation of PCC — by applying the inverse hyperbolic tangent — so as to obtain a normally distributed index. FZT is basically the same as PCC for values between -0.5 and 0.5. Weighted Coherence (WC) is a frequency domain PCI and therefore more useful for periodic physiological responses such as the heart beat and the respiration rate. Since EDA is the signal this study is focused on and it is not a periodical signal, WC has not been considered.

According to each PCI definition, the higher the values the higher or lower the PC. This direct or inverse relationship is summarised in table 1, together with the property that each PCI measures. By definition, PCIs are mostly computed pairwise, being DA the exception. DA can be calculated for as many individuals as desired. PCIs can be obtained on either an instantaneous or aggregated basis. Aggregated can be either for time windows, say for example 65s as in [15], or for the whole measurement session. They can all be aggregated but not all can be computed on an instantaneous basis. PCC and FZT are meaningless calculated for a single instant. Although in this study aggregated values for the whole collaborative learning task are used for the prediction of collaboration features, instantaneous values might have the power to detect particular points of interest in the study of collaboration. It has been claimed that understanding the underlying mechanisms of collaborative learning requires research to zoom in the collaborative interactions [13].

Table 1: Summary of PCIs meaning

PCI	Interpretation	Value	PC
SM	difference	lower	higher
IDM	rate of change	lower	higher
DA	direction	higher	higher
PCC	linear relationship	higher	higher
FZT	weighted mean PCC	higher	higher

PCIs can play a role in studying the social interactions [11] by providing an objective measure, although research in this direction remain scarce. A relatively small number of studies has focused on the PC phenomenon [25], as most of the research carried out in psychophysiology and applications has targeted the individual rather than the group level [11]. Although insufficient, psychophysiological research for collaboration has proved promising.

Evidence has been found for a connection between PC and team performance [35], with PCC and DA of cardiovascular activity being the most sensitive indices to differences between low and high performers [15]. PCC of heart rate in dyads has been reported to predict task completion time [23]. Also in dyads, PC has been associated to interaction and self-reported social presence [25, 10]. Conflicting interactions have been reflected in a significantly higher PC increment than that of collaborative interactions by means of the PCC index [10]. The directionality of the PC might point to who is the group leader [11]. Some of the collaboration features studied from the PC phenomenon perspective are listed in table 2 together with the PCIs employed.

Physiological signals are valuable social cues for the study of collaborative learning. Different collaboration features have been previously studied. This paper enriches the field

Table 2: Some collaboration features studied through PCIs

Collaboration feature	Study	PCIs
Team performance	[15, 35]	SM, IDM, DA, PCC
Interaction	[25]	PCC, WC
Task completion time	[23]	PCC, WC
Conflicting interactions	[10]	PCC
Team work	[22]	PCC, WC

with three collaborative learning specific features that have not been approached through PCIs before.

3. METHOD

Based on the exposed rationale, we post the research questions:

1. Which PCI reflects better the collaborative will?
2. Which PCI predicts the collaborative learning product and the dual learning gain the best?

At this point it is convenient to remember that the PCIs are calculated on a certain physiological response — EDA in this study. Therefore, it has to be kept in mind that here the answer to the research questions is linked to and valid only for the EDA signal. Further studies are needed for the comparison of PCIs across different physiological responses.

3.1 Setting

An experiment was conducted in the University of Oulu's Learning and Interaction Observation Forum (LeaForum[2]). This state-of-the-art research infrastructure was designed as a roomy, cosy space for up to 30 people. It is a convertible facility with flexible fittings to allow provision for different group situations. LeaForum is equipped with:

- a proprietary observation and recording system [27];
- trapezoidal tables convenient for group work;
- 14 tablets connected to the Internet via Wi-Fi;
- 6 Empatica S3 biosensors [20];
- 2 eye tracking glasses; and
- a backstage room prepared to follow the experiments in real time through the live video signals.

The video recording system is able to collect 360° video and high-quality audio from multiple microphones. The trapezoidal tables allow for a suitable and comfortable group setting, as they form an heptagon when seven of them are sided consecutively. In front and close to each team's tables, there was a table supplied with a variety of food items (cardboard boxes, plastic bags, fruits...) relevant for the task.

The tablets enabled the students to take pictures, search information on the web, respond to online questionnaires and access *weSPOT* [33], the virtual learning environment (VLE) used. WeSPOT is a working environment with social and personal open tools for Inquiry Based Learning (IBL).

The Empatica S3 wristband is a multi-sensor device especially designed for research purposes. Four sensors come embedded: 1) a photoplethysmograph (for cardiovascular

---

[2]http://leaforum.fi (Last accessed: 20/10/2015)

activity), 2) a thermometer, 3) an accelerometer, and 4) an EDA sensor. Two modes of operation are available: 1) real time Bluetooth streaming and 2) internal memory recording. The latter was used. Ambient temperature in LeaForum is usually controlled by means of a thermostat to be 23°C, the recommended temperature to measure EDA [5].

3.2 Participants

Participants in the experiment were 48 high-school students from the University of Oulu's Teacher Training School. Students came from six different classes and were aged 16 to 19. Gender distribution was 27 females and 21 males. No previous knowledge was required. The task was coordinated with their science teachers so that it was included within their course work, meaning it did not supposed extra effort for the students. Participation in the study was voluntary.

3.3 Task description

The task design in close collaboration with the students' science teachers ensured that it was aligned with their cognitive level. It also made it meaningful for their current studies. The experiment was run four times. The task was the same but two levels of scripting were used: guided and unguided — two runs each.

The task consisted in the design of a healthy, appropriate breakfast for an athlete training for a marathon. The specification of parameters included age, height, weight, daily caloric intake, number of weekly trainings and session length. Students were provided with two Google-Docs® files: a document and a spreadsheet. For both guided and unguided runs, the document contained a general description of marathon runners' nutritional needs. The spreadsheet had a template with rows for food items and columns for the weight (in grams) of the different nutrients contained by the food item. Several examples were provided.

The unguided task required students to search on the Internet which and how much of each nutrient a marathon runner needs for breakfast. They were asked to write their most relevant findings on the nutrients and their function.

In the guided runs of the experiment the provided document already contained the recommendation for the composition of the athletes' breakfast in per cent of each nutrient, energy consumption of marathon runners as well as the description and organic role of nutrient categories such as carbohydrates, proteins, minerals and fat. Students were prompted the steps to follow through the VLE. The steps according to the IBL paradigm were 1) plan the design method, 2) set the criteria, 3) collect the data, 4) discuss the findings, and 5) communicate the results.

3.4 Procedure

A pre-test and a Motivated Strategies for Learning Questionnaire (MSLQ) [43] were applied to the participants two weeks in advance to the laboratory experiment. The aims were on the one side to have a preliminary evaluation of students previous knowledge on the task subject, motivation and attitudes towards learning; and on the other side to have a group formation criterion. The MSLQ questionnaire is considered a social-cognitive view of motivation and self-regulation of learning. Students' scores in each of the two tests were categorized in low, middle and high. These categories were then used as the basis to form the groups as heterogeneously as possible.

Students were organized in 16 triads for the collaborative task. Due to LeaForum capacity and the number of available biosensors, the experiment was run four times in two consecutive days, a morning and afternoon run each day. The unguided task was used in the first day's sessions. On the contrary, second day's sessions were carried out with the guided task.

Every session lasted for two hours (on task) and fifteen minutes (preparation). At the beginning of each session, the students were introduced to LeaForum and its equipment, especially the sensors they were to wear throughout the task. Each student was provided with a tablet and a microphone for the videotaping. Two students in separate groups were wearing eye-tracking glasses each run. The six biosensors available at LeaForum allowed for the physiological tracking of two groups each session. After the sensor familiarization, the collaborative learning task together with the VLE were explained. Having completed the task understanding phase, students started to work with their groups. During the whole experiment, two researchers from the team were around to handle possible issues with the task or the equipment. The rest of the research team was following the experiment from LeaForum's backstage room, prepared for observation without causing disturbance or distraction to students.

Upon task completion, students delivered a report with their solution via the VLE. Finally, they were asked to take the post-test.

3.5 Collected data

The number of sensors available formed a limitation for tracking every student. In addition, as the task was collaborative, not all the students interacted with the VLE (task distribution). From the capability to track the physiological responses of 24 students (6 biosensors ∗ 4 runs), the real number was reduced to 20 due to four cases of device misplacement. Those 20 were the students we finally considered for this study from the experiment data focused on EDA. The total number of students in each data modality collected is specified in table 3.

Table 3: Students tracked in each data modality

Data modality	Students
Pre-, post- and MSLQ questionnaires	48/48
Videotaping	48/48
VLE activity logs	33/48
Wristband biosensors	20/48
Eye-tracking	8/48

At the group level, data from the biosensors is distributed as follows: 5 groups with all three members tracked, 2 groups with two out of the three members tracked and 1 group with a single member tracked. The latter cannot be used therefore to measure collaboration, but it can be compared to students working in other groups during the same session for contrasting the PCIs of people working together with those of people working simultaneously but in other teams.

The dimensions of the data appear in table 4. The sample frequency specifies the number of measurements the biosensors collect every second. Thus for example, EDA data for each participant adds up to around 28,800 samples (taking four samples a second during approximately two hours).

Considering the 20 participants tracked with this data modality, accounts for over half a million records.

Table 4: Physiological data dimensions

Data	Units	Sample frequency (Hz)
Temperature	°C	4
EDA	μS	4
BVP	nW	64
Accelerometer	1/64g	32
IBI	s	—
Pupil diameter	mm	64

BVP is the direct measure coming from the photoplethysmograph sensor. BVP peaks indicate a heart beat. Using the BVP signal an Empatica proprietary algorithm extracts the interbeat interval (IBI) (i.e. the time in between two consecutive heart beats), from which in turn heart rate is calculated.

From the data collected, figure 3 is very illustrative of the power of EDA data. There, the EDA signals have been synchronized first among themselves and then with the video. The figure shows shadowed an interval where there is an almost simultaneous peak in the three EDA responses. A video inspection reveals that the peak occurs at the time the team feels that something has gone wrong with the VLE, that is, a technical challenge. It is also seen in the image that the student in the middle has taken his hands to his head, signaling thereby that he is worried. The figure shows also the individual differences in the level of the EDA signal. As it has been stated before, EDA is strongly dependent on individual characteristics, and so are other physiological responses.
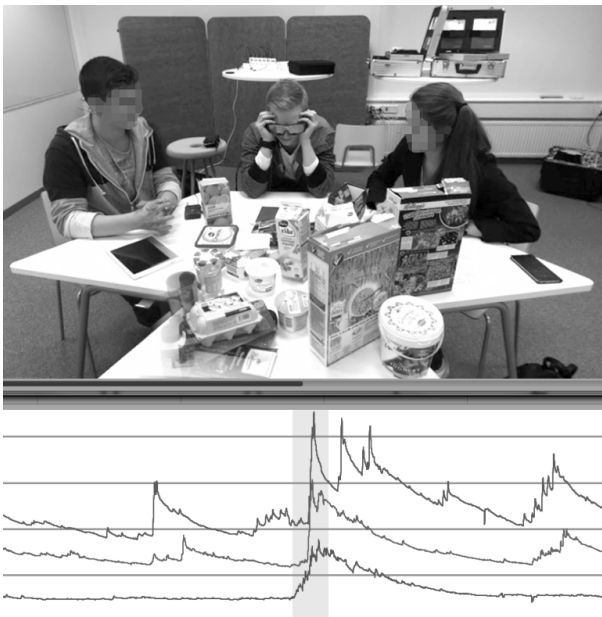

Figure 3: Challenge-driven team EDA peaks.

3.6 Data processing

Data processing steps are summarized in figure 4.

After the collection the data had a high degree of sparsity. Different sources usually imply different formats. Even raw
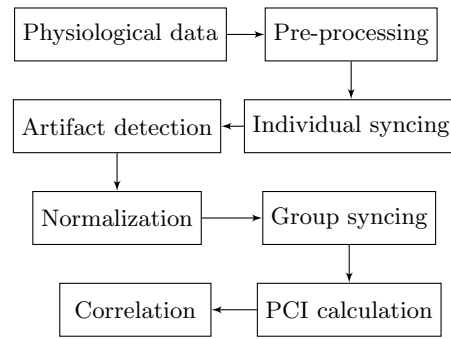

Figure 4: Physiological data processing.

data coming from the same source may be separated. This is the case of the biosensors. The five magnitudes they provide (first five rows in table 4) are each saved to a different file. File format is typically comma separated values (CSV) with a first row indicating the start recording time, a second row displaying the sample frequency and the remaining rows for the measured values. First step is therefore to have the data in such a way that every value is time stamped for each magnitude and participant (pre-processing).

The synchronization of multimodal data and even within the same data modality is a challenge [11, 28]. It is also a core process for the exploration of PCIs as they rely on responses happening at the very same instant. Otherwise the results would not be reliable. Mostly manual synchronization was performed in the data processing. Automated mechanisms are under study for the goal of a timely, learning-oriented, biofeedback dashboard.

We chose to begin the physiological signals synchronization within individuals with a magnitude sampled at 4Hz, which is the maximum common divisor of the sampling frequencies in table 4. Thus, starting with temperature, the physiological signals were synchronized using Microsoft Excel VLOOKUP function one participant and signal at a time (individual synchronization).

Biosensor recordings may include artifacts as a result of a variation in the contact between the skin and the sensor due to pressure, excessive movement, or adjustment of the device [52]. There is a need to detect those artifacts to prevent wrong data from distorting the analysis. Three free applications were found for the processing of EDA data: Ledalab [26], EDA Explorer [52] and PsPM[3]. Ledalab allows for the manual detection of artifacts. PsPM does not include this feature either in manual or automatic way. EDA Explorer automatically detects artifacts based on 5-second epoch classification. EDA Explorer was therefore used to categorize EDA data points into signal or noise (artifact detection). As a result, 90% of the EDA data (490,880 from 547,848 records) was classified as correct and the remaining 10% was taken out of the analysis.

The strong dependence of EDA values on individual characteristics makes a normalization process necessary for comparability — at least for the level-dependent PCIs. EDA signals are then brought to a baseline. Normalization was done by computing the t-scores (i.e. subtracting the sample mean from every value and then dividing by the sample standard deviation). Both sample mean and standard deviation were computed on an individual basis.

---

[3]http://pspm.sourceforge.net (Last accessed: 20/10/2015)

To explore the collaboration through physiological responses, a second synchronization has to be done to align the data across individuals. The group synchronization is carried out on a session basis, meaning for all the individuals tracked with biosensors within a session.

The indices SM, IDM and DA were computed according to [15]; PCC with lag 0 as in [23] and FZT following [9]. PCIs were obtained pairwise for all team member combinations (AB, AC and BC) to ensure a greater index reliability [22]. The instantaneous values were then aggregated as the arithmetic mean throughout the collaborative learning session. The mean over a certain time period is a commonly accepted physiological feature [38]. As a result of the whole process, five 21x21 matrices were obtained, one for each PCI.

3.7 Collaborative learning measures

Three collaborative learning measures (CLMs) were used in the study: 1) collaborative will (CW), 2) collaborative learning product (CLP) and 3) dual learning gain (DLG).

The predisposition to collaborate in a learning task, that is, the CW, was measured using the peer learning scale of the MSLQ questionnaire. MSLQ is a widely used self-report instrument with 81 items distributed in two sections — motivation and learning strategies — with 6 and 9 scales respectively. Students' rate ranges from 1 (not at all true of me) to 7 (very true of me). The peer learning scale comprises three items — numbers 34, 45 and 50 — in which the students rate their willingness to explain learning material to their team members, work together with others to complete a learning assignment and discuss task material with the team. Individual MSLQ scales are analysed by averaging the ratings of the different items the scale is comprised of. Therefore, the CW measured in this way ranges from 1 to 7, just as the individual items.

The CLP was measured by the score of the report delivered by the students after the task. The report included two files: 1) a document with their task notes and conclusions, and 2) a spreadsheet with their task solution. The assessment criteria involved the diversity of the solution (number of nutrients considered), depth of the answer, accuracy and focus on the problem. The scoring scale ranges from 4 to 15 points.

The individual learning gain was computed by subtracting pre-test from post-tests scores. DLG was then calculated as the sum of the individual learning gains.

4. ANALYSIS AND RESULTS

Figure 5 shows in a scatter chart with smooth lines the pairwise PCI values and the CLMs sorted from left to right by descendent CLP. The figure illustrates nearly the same behaviour between FZT and PCC. Therefore, they may be considered redundant PCIs and it is simpler to use PCC instead of FZT, as FZT is a derivative measure of PCC and thus requires further calculation, which in turn translates into no added benefit.

With the five PCIs computed on the one side, and the three CLMs on the other side, fifteen regression analyses were performed pairwise (see figure 6). The aim was to explore the power of every PCI to predict each CLM and thereby being able to answer the research questions. The results are summarized in table 5.

In addition, a regression analysis was performed between pre- and post-test scores. High correlation (0.86) was found between the two tests, with the pre-test score being able to predict 74% of the post-test scores variance. The regression line obtained is displayed in figure 7. The line has a positive slope (0.70), indicating that in general there was a positive learning gain for participants in the experiment.

Table 5: Correlations in the regression analyses

CLM	SM	IDM	DA	PCC	FZT
CW	0.01	**0.50**	0.12	0.09	0.09
CLP	0.13	**0.59**	0.12	0.32	0.35
DLG	0.27	0.20	**0.70**	0.17	0.17

In general, weak correlation (below 0.3) was found among the PCIs and the CLM. IDM showed a moderate to strong correlation with the CW (0.5) and the CLP (0.59). IDM was therefore the best predictor for those CLMs. DA revealed a strong correlation to the DLG (0.7), meaning it explains almost half of DLG variation (determined by the square of the correlation coefficient). This might indicate that while the linearity (i.e. proportional changes) in the EDA of students learning collaboratively appears to have no predictive power, the direction in which the signals are changing could tell about the resulting learning gain.

Once again the simplest solution proved to be the better. DA is the simplest of the PCIs, yet it yielded the highest correlation in the regression analysis. This is in agreement with former studies finding DA as the most sensitive PCI to differences in team performance [15]. It is also the most powerful at group level by enabling the computation of one index for all the members and not only pairwise, which is a common limitation of the other PCIs. In addition, DA does not require normalization thanks to its level independence, being just driven by the direction of change as the name itself implies. All of this means that DA can get to a dashboard quicker than its counterparts, making it the most suitable PCI for real time applications.

It is also worth stating that this study is not without its limitations. The sample size was bounded by the number of biosensors available. Also, the experiment was conducted in a laboratory setting, which differs from the students' natural learning environments. The EDA signal was recorded from a single wrist — as in the overwhelming majority of studies using EDA, but research has shown that EDA is an asymmetrical response (i.e. it manifests different in left and right hemispheres) [32, 42]. Therefore, recording EDA in the two wrists might offer additional information.

5. CONCLUSIONS AND FUTURE WORK

A physiological approach to collaborative learning research has been presented. This study used the EDA records from the rich multimodal dataset produced in the experiment. EDA, a measure of psychological sweat if temperature is kept constant, has been linked in former studies to cognitive and emotional processes. Here, five PCIs were calculated from the EDA of students working collaboratively to investigate for the first time their possible connection to three collaborative learning features. The best predictor for each of the latter was obtained from the PCIs: IDM for the CW and the CLP; and DA for the DLG.

Further studies on the applications of EDA to the learning sciences should consider the laterality effect. Lateral measures need to be compared and correlated to learning
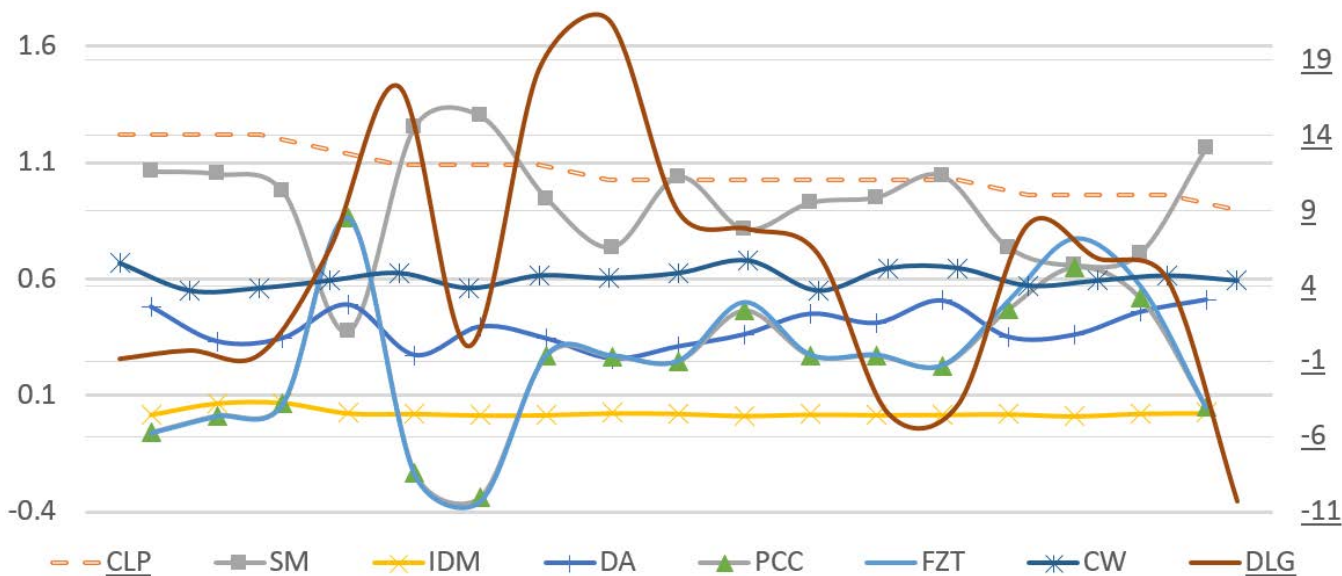
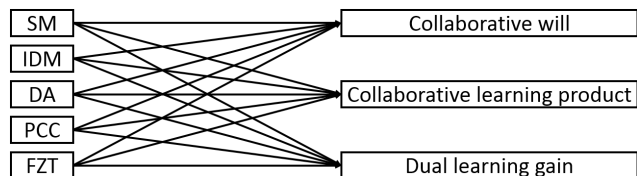**Figure 5: PCI and learning measures.**

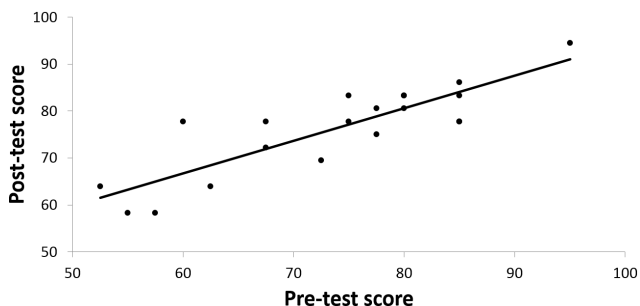**Figure 6: Pairwise regression analyses.**

Figure 7: Post- vs. pre-test regression line.

outcome indicators. It may result that measurements from one side are more significant or sensitive in a learning context. But it may also turn out that none of them could be safely ruled out.

EDA is valuable as a highly sensitive index of psychological activation, but other physiological responses also deserve to be explored. The PCIs predictive power is likely to vary across physiological responses, and it is of interest to find out which physiological response maximizes the PCIs correlation to different learning features.

Research is needed with focus on whether there is a PC among people in a collaborative setting. An experiment design might be to track the physiological signals of a number of individuals working alone. Then, the individuals are asked to work in groups. The physiological responses in the two situations can be compared to find changes. The con-

trast of PCI values alone and in group can tell if the PC phenomenon occurred.

As biosensors are becoming more and more common in everyday life, we believe that they can increase the accessibility to a data source with potential to enrich the LA field. They can be used as input for a learning-oriented biofeedback dashboard in a computer supported collaborative learning (CSCL) context. Such a dashboard is the next milestone in our research agenda.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] I. Arroyo, D. G. Cooper, W. Burleson, B. P. Woolf, K. Muldner, and R. Christopherson. Emotion sensors go to school. In *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems that Care: from Knowledge Representation to Affective Modelling*, pages 17–24, Amsterdam, The Netherlands, The Netherlands, 2009. IOS Press. URL: http://dl.acm.org/citation.cfm?id=1659450.1659458.

[2] M. Baker, S. Järvelä, and J. Andriessen. *Affective Learning Together: social and emotional dimensions of collaborative learning*. Routledge, 2013.

[3] B. Barron. When smart groups fail. *Journal of the Learning Sciences*, 12(3):307–359, 2003. doi:10.1207/S15327809JLS1203_1.

[4] B. S. Bloom. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1-2*. Longmans: McKay, 1974.

[5] W. Boucsein. *Electrodermal Activity*. Springer, 2 edition, 2012. doi:10.1007/978-1-4614-1126-0.

[6] T. M. Brown and E. Fee. Walter Bradford Cannon: Pioneer physiologist of human emotions. *American Journal of Public Health*, 92(10):1594–1595, 2002. `doi:10.2105/AJPH.92.10.1594`.

[7] S. Buckingham-Shum. Learning Analytics. *UNESCO Institute for Information Technologies in Education (ITTE), Policy Brief*, 2012. URL: http://iite.unesco.org/pics/publications/en/files/3214711.

[8] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson. *Handbook of Psychophysiology*, chapter Psychophysiological Science: Interdisciplinary Approaches to Classic Questions About the Mind, pages 1–16. Cambridge University Press, 3 edition, 2007.

[9] G. Chanel, M. Bétrancourt, T. Pun, D. Cereghetti, and G. Molinari. Assessment of computer-supported collaborative processes using interpersonal physiological and eye-movement coupling. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, pages 116–122, 2013. `doi:10.1109/ACII.2013.26`.

[10] G. Chanel, J. M. Kivikangas, and N. Ravaja. Physiological compliance for social gaming analysis: Cooperative versus competitive play. *Interacting with Computers*, 24(4):306–316, July 2012. `doi:10.1016/j.intcom.2012.04.012`.

[11] G. Chanel and C. Muhl. Connecting Brains and Bodies: Applying Physiological Computing to Support Social Interaction. *Interacting with Computers*, 2015. `doi:10.1093/iwc/iwv013`.

[12] M. E. Dawson, A. M. Schell, and D. L. Filion. *Handbook of Psychophysiology*, chapter The Electrodermal System, pages 159–181. Cambridge University Press, 3 edition, 2007.

[13] P. Dillenbourg. What do you mean by collaborative learning? *Collaborative-learning: Cognitive and Computational Approaches*, pages 1–19, 1999.

[14] ECAR-ANALYTICS Working Group. The predictive learning analytics revolution: Leveraging learning data for student success. *ECAR working group paper*, 2015.

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

[16] N. Enyedy and R. Stevens. *The Cambridge Handbook of the Learning Sciences*, chapter Analyzing collaboration, pages 191–212. Cambridge University Press, 2014.

[17] S. H. Fairclough. Fundamentals of physiological computing. *Interacting with Computers*, 21(1-2):133–145, 2009. `doi:10.1016/j.intcom.2008.10.011`.

[18] C. Féré. Note sur les modifications de la résistance électrique sous l'influence des excitations sensorielles et des emotions. *Comptes Rendus des Séances de la Société de Biologie*, 5:217–219, 1888.

[19] E. Ferreira, D. Ferreira, S. Kim, P. Siirtola, J. Röning, J. F. Forlizzi, and A. K. Dey. Assessing real-time cognitive load based on psycho- physiological measures for younger and older adults. In *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2014 IEEE Symposium on*, pages 39–48, Orlando, FL, 2014. IEEE. `doi:10.1109/CCMB.2014.7020692`.

[20] M. Garbarino, M. Lai, S. Tognetti, R. Picard, and D. Bender. Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. *Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare - "Transforming healthcare through innovations in mobile and wireless technologies"*, pages 39–42, 2014. `doi:10.4108/icst.mobihealth.2014.257418`.

[21] M. K. Garg, D.-J. Kim, D. S. Turaga, and B. Prabhakaran. Multimodal analysis of body sensor network data streams for real-time healthcare. In *Proceedings of the International Conference on Multimedia Information Retrieval - MIR '10*, page 469, New York, New York, USA, Mar. 2010. ACM Press. `doi:10.1145/1743384.1743467`.

[22] R. A. Henning, A. G. Armstead, and J. K. Ferris. Social psychophysiological compliance in a four-person research team. *Applied Ergonomics*, 40(6):1004–1010, 2009. `doi:10.1016/j.apergo.2009.04.009`.

[23] R. A. Henning, W. Boucsein, and M. Claudia Gil. Social-physiological compliance as a determinant of team performance. *International Journal of Psychophysiology*, 40(3):221–232, Apr. 2001. `doi:10.1016/S0167-8760(00)00190-2`.

[24] R. A. Henning and S. L. Sauter. Work-physiological synchronization as a determinant of performance in repetitive computer work. *Biological Psychology*, 42(3):269–286, Feb. 1996. `doi:10.1016/0301-0511(95)05162-7`.

[25] S. Järvelä, J. M. Kivikangas, J. Kätsyri, and N. Ravaja. Physiological linkage of dyadic gaming experience. *Simulation & Gaming*, 45(1):24–40, 2014. `doi:10.1177/1046878113513080`.

[26] C. Karenbach. Ledalab — a software package for the analysis of phasic electrodermal activity. Technical report, Allgemeine Psychologie, Institut für Psychologie, 2005.

[27] A. Keskinarkaus, S. Huttunen, A. Siipo, J. Holappa, M. Laszlo, I. Juuso, E. Väyrynen, J. Heikkilä, M. Lehtihalmes, T. Seppänen, and S. Laukka. MORE — a multimodal observation and analysis system for social interaction research. *Multimedia Tools and Applications*, 2015. `doi:10.1007/s11042-015-2574-9`.

[28] J. Kim, M. Snodgrass, M. Pietrowicz, K. Karahalios, and J. Halle. BEDA: Visual analytics for behavioral and physiological data. *Proceedings of the Workshop on Visual Analytics in Healthcare*, 2013.

[29] P. A. Kirschner, J. Sweller, and R. E. Clark. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2):75–86, 2006. `doi:10.1207/s15326985ep4102\_1`.

[30] S. D. Kreibig. Autonomic Nervous System activity in emotion: a review. *Biological psychology*, 84(3):394–421, July 2010. `doi:10.1016/j.biopsycho.2010.03.010`.

[31] R. W. Levenson and J. M. Gottman. Marital

interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology*, 45(3):587–597, 1983. `doi:10.1037/0022-3514.45.3.587`.

[32] C. A. Mangina and J. Beuzeron-Mangina. Direct electrical stimulation of specific human brain structures and bilateral electrodermal activity. *International Journal of Psychophysiology*, 22(1-2):1–8, Apr. 1996. `doi:10.1016/0167-8760(96)00022-0`.

[33] A. Mikroyannidis, A. Okada, P. Scott, E. Rusman, M. Specht, K. Stefanov, P. Boytchev, A. Protopsaltis, P. Held, S. Hetzner, K. Kikis-Papadakis, and F. Chaimala. weSPOT: A personal and social approach to inquiry-based learning. *Journal of Universal Computer Science*, 19(14):2093–2111, 2013. `doi:10.3217/jucs-019-14-2093`.

[34] N. Miyake and P. A. Kirschner. The social and interactive dimensions of collaborative learning. In *The Cambridge handbook of the learning sciences*, pages 418–438, 2014.

[35] E. Montague, J. Xu, and E. Chiou. Shared experiences of technology and trust: An experimental study of physiological compliance between active and passive users in technology-mediated collaborative encounters. *IEEE Transactions on Human-Machine Systems*, 44(5):614–624, 2014. `doi:10.1109/THMS.2014.2325859`.

[36] P. Näykki, H. Järvenoja, S. Järvelä, and P. Kirschner. Monitoring makes a difference: quality and temporal variation in teacher education students' collaborative learning. *Scandinavian Journal of Educational Research*, pages 1–16, 2015.

[37] D. J. Nicol and D. Macfarlane-Dick. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in higher education*, 31(2):199–218, 2006. `doi:10.1080/03075070600572090`.

[38] D. Novak, M. Mihelj, and M. Munih. A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting with Computers*, 24(3):154–172, May 2012. `doi:10.1016/j.intcom.2012.04.003`.

[39] L. Noy, N. Levit Binnun, U. Alon, and Y. Golland. Being in the zone: physiological markers of togetherness in joint improvisation. *Frontiers in Human Neuroscience*, 9(187), 2015. `doi:10.3389/fnhum.2015.00187`.

[40] A. Pecchinenda. The Affective Significance of Skin Conductance Activity During a Difficult Problem-solving Task. *Cognition & Emotion*, 10(5):481–504, 1996. `doi:10.1080/026999396380123`.

[41] R. Picard. *Affective Computing*. MIT press, 1997.

[42] R. W. Picard, S. Fedor, and Y. Ayzenberg. Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review*, 2015. `doi:10.1177/1754073914565517`.

[43] P. R. Pintrich et al. *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. ERIC, 1991.

[44] S. W. Porges, R. E. Bohrer, M. N. Cheung, F. Drasgow, P. M. McCabe, and G. Keren. New time-series statistic for detecting rhythmic co-occurrence in the frequency domain: the weighted coherence and its application to psychophysiological research. *Psychological bulletin*, 88(3):580, 1980.

[45] M. Russel. *Improving Student Retention in Higher Education: The Role of Teaching and Learning*, chapter Leveraging student engagement with assessment: collecting intelligence to support teaching, student progress and retention, page 108. Taylor & Francis, 2008.

[46] J. Schneider, D. Börner, P. van Rosmalen, and M. Specht. Augmenting the Senses: A Review on Sensor-Based Learning Support. *Sensors*, 15(2):4097–4133, 2015. `doi:10.3390/s150204097`.

[47] L. Shen, M. Wang, and R. Shen. Affective e-Learning: "Using Emotional" Data to Improve Learning in Pervasive Learning Environment Related Work and the Pervasive e-Learning Platform. *Educational Technology & Society*, 12(2):176–189, 2009.

[48] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen. Galvanic skin response (gsr) as an index of cognitive load. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '07, pages 2651–2656, New York, NY, USA, 2007. ACM. `doi:10.1145/1240866.1241057`.

[49] G. Siemens and R. S. J. d. Baker. Learning analytics and educational data mining. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, page 252, New York, New York, USA, Apr. 2012. ACM Press. `doi:10.1145/2330601.2330661`.

[50] T. Smith and K. Smith. *Handbook of Human Factors*, chapter Feedback-control mechanisms of human behavior, pages 251–293. Wiley, 1987.

[51] M. Swan. Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. *Journal of Sensor and Actuator Networks*, 1(3):217–253, Nov. 2012. `doi:10.3390/jsan1030217`.

[52] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard. Automatic identification of artifacts in electrodermal activity data. *EMBC (to appear). IEEE*, 2015.

[53] B. Trilling and C. Fadel. *21st Century Skills: Learning for Life in Our Times*. Wiley, 2009.

[54] G. Wolfgang and H. Drachsler. Translating learning into numbers: a generic framework for Learning Analytics. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 15(3):42–57, 2012. `doi:10.1145/2330601.2330634`.

[55] R. S. Woodworth and H. Schlosberg. *Experimental psychology*. Oxford and IBH Publishing, 1954.