Héctor Javier Pijeira Díaz

ELECTRODERMAL ACTIVITY AND SYMPATHETIC AROUSAL DURING COLLABORATIVE LEARNING
HÉCTOR JAVIER PIJEIRA DÍAZ

ELECTRODERMAL ACTIVITY AND SYMPATHETIC AROUSAL DURING COLLABORATIVE LEARNING

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Human Sciences of the University of Oulu for public defence in the OP auditorium (L10), Linnanmaa, on 3 May 2019, at 12 noon

UNIVERSITY OF OULU, OULU 2019
Abstract

This dissertation investigates high school students’ individual and interpersonal physiology of electrodermal activity (EDA) during collaborative learning in naturalistic settings. EDA is an index of sympathetic arousal, which is concomitant with cognitive and affective processes.

Two data collections were organized with students working collaboratively in triads. The first one took place during the performance of a science task, and the second during two runs of a six-week advanced physics course. The data collected included EDA (measured unobtrusively using Empatica® E3 and E4 wristbands), performance measures (pre- and post-tests, task solutions, and course exam), and questionnaires on cognitive, affective, and collaborative aspects of learning. The work was reported in three articles.

The results indicate that, on average, students spent more than half (60%) of the class at a low arousal level, possibly signaling relaxation, disengagement, or boredom. Most of the time (≈60–95% of the lesson), triad members were at a different arousal level, which might indicate that students took turns (alternating task-doers) in executing the task or applied some division of labor rather than truly collaborating. In terms of achievement, sympathetic arousal during the exam was a predictor of the exam grades, and pairwise directional agreement of EDA was positively and highly correlated to the dual learning gain. Arousal contagion could have occurred in up to 41% of the high arousal intervals found. The possible arousal contagion cases took place mostly on a 1:1 basis (71.3%), indicating that interactions in a collaborative learning triad seem to occur mainly between two members rather than among the three.

The findings provide an ecologically-valid picture of the students’ EDA responses in the classroom, both individually and collaboratively, benefiting from the connection of arousal to cognitive and affective processes to increase the saliency of otherwise elusive phenomena. Methodologically, the study contributes to the exploration and exploitation of psychophysiological approaches for collaborative learning research. On a practical level, it provides physiological indices that could be incorporated into learning analytics dashboards to support students’ awareness and reflection, and teachers’ pedagogical practices.

Keywords: collaborative learning, electrodermal activity, interpersonal physiology, psychophysiology, sympathetic arousal
Tässä väitöstutkimuksessa tarkastellaan elektrodermaalista aktiivisuutta (EDA) ja tästä johdetuua sympaattista vireystilaa ja fysiologisia indeksjä, samanaikaisesti yksilöiden ja yksilöiden välisten kognitiivisten ja affektiivisten prosessien kanssa.

Tutkimusaineisto kerättiin yhteisöllisen oppimisen tilanteista, joissa oppilaita työskentelivät kolmen hengen ryhmissä. Ensimmäinen osa aineistosta kerättiin oppilaiden suorittamalla luonnontieteiden alan tehtävällä ja toinen kahden fysiikan syventävän kurssin aikana. Aineistoon sisältyi EDA (Empatica® E3- ja E4-erästä), oppimisen mittauksia (alku- ja lopputesti, tehtävien ratkaisut ja kurssikokeet) sekä kyselylomakkeet oppimisen kognitiivisista, affektiivisista ja yhteisöllisen työskentelyn näkökulmista. Tutkimus on raportoitu kolmessa artikkelissa.

Tulokset osoittavat, että opiskelijoiden sympaattisen hermoston vireystila oli keskimäärin yli puolet (60 %) luokkatyöskentelystä alhainen, mikä viittaa mahdolliseen renoutumiseen, osallistumisen puutteeseen tai tylsistymiseen. Ryhmänjäsenet olivat suurimman osan aikasta (=60-95 %) eri vireystilan tasolla, mikä voi tarkoittaa, että he suorittivat tehtävää vuorotellen (tehtävän suorittajaa vaihdellen) tai jonkinlaisista työnjakoa käyttäen, yhteisöllisen työskentelyn sijaan. Sympaattinen vireystila kurssikokeissa ennusti kokeen arvosanoja. Lisäksi oppilasparien EDA:n samansuuntaisuus korreloitu vahvasti oppimistulosten kanssa. Yksilöiden välillä tapahtuva sympaattinen vireystilan "tarttumista" on voinut esiintyä jopa 41 prosentissa todetuista korkean vireystilan intervalleista. Mahdolliset "tarttumiset" ilmenivät enimmäkseen (71,3 %) 1:1 suhteessa, mikä viittaa siihen, että vuorovaikutus yhteisöllisessä oppimisessa näyttäisi tapahtuvan pääasiallisesti kaikkien kolmen sijaan.

Tulokset tarjoavat ekologisesti validin kuvan opiskelijoiden EDA-reaktiosta luokkahuoneessa sekä yksilöllisesti että yhteisöllisesti tarkasteltuna, selventää samalla kuvaa sympaattisen vireystilan yhteydestä kognitiivisiin ja affektiivisiin prosesseihin. Menetelmällisesti tutkimus kartoittaa psykofysiologisen lähestymistavan mahdollisuuksia yhteisöllisen oppimisen tutkimuksessa. Se esittelee fysiologisia indeksjä, jotka voitaisiin visualisooida oppimisen analytiikan sovelluksissa opiskelijoiden tietoisuuden ja reflektion sekä opettajien pedagogisten käytäntöjen tukemiskeksi.

Asiakas: elektrodermaalinen aktiivisuus, ihmistenvälinen fysiologia, psykofysiologia, sympaattinen vireystila, yhteisöllinen oppiminen
To my lifetime supervisors and indefatigable supporters,  
my parents
Acknowledgements

It was the early spring of 2015 when I started this journey to the Ph.D. Along the way, I have felt privileged to count on the supervision of eminent professors such as Sanna Järvelä (University of Oulu), Paul A. Kirschner (Open University of the Netherlands), and Hendrik Drachsler (Goethe University Frankfurt), to whom I am profoundly grateful. I am especially indebted to Sanna Järvelä, my principal supervisor, for her support in so many ways throughout these years. This Ph.D. would simply not have been possible without Sanna’s involvement and leadership, and her innate capacity to motivate and inspire people with her kindness, result-orientation, and passion for excellence. Thank you, Sanna, for being always available when I needed it, no matter how busy you were; for all the expert guidance of a natural leader like you, and for embracing multidisciplinarity to advance the learning sciences. To Paul A. Kirschner I am grateful for the insightful scientific discussions, for his honest commentary on my work, and for his efforts in shaping my scientific writing skills with a focus on structure, clarity, line of argument, and coherence. From his keen remarks as an experienced editor, I have learned much of what I now about academic writing formalities and the APA style, for example. Thank you, Paul, also for always coming to my conference presentations despite your tight agenda as a sought-after expert in those events, and for the feedback afterwards. To Hendrik Drachsler I am thankful for his dedication to getting me started with the Ph.D. endeavor, for helping me to pave the way for independent work to advance learning analytics, for introducing me to valuable tools and research methods, and in general, for the much-appreciated guidance and opportunities along these four years.

I would like to express my gratitude to Professor Timo Tobias Ley (Tallinn University), an internationally recognized expert in learning analytics and educational innovation, who is highly experienced in multidisciplinary work, for accepting to act as opponent at the public defense of this dissertation. For their meticulous work as pre-examiners of this dissertation, through which they provided me with insightful, actionable, and critically constructive feedback, I want to thank Engineering Pedagogy Professor Petri Nokelainen (Tampere University of Technology) and Associate Professor Daniel Spikol (Malmö University). Similarly, I am deeply grateful to the Academy of Finland post-doctoral research fellows Piia Näykki and Sonja Lutovac, for their elaborate, highly-detailed, and thorough commentary on a previous version of this dissertation, as part of the corresponding pre-defense seminar at the Faculty of Education.
During my doctoral studies, annual meetings with my follow-up group helped me to evaluate my progress and to keep my research plan up-to-date. I would like to show my appreciation to the chair of this group, Adjunct Professor Arto Hautala (Cardiovascular Research Group, Oulu University Hospital), as well as to the other two members: Dr. Miki Kallio (Unit for Strategy and Science Policy) and Dr. Arttu Mykkänen (Learning and Educational Technology research group; LET). I am thankful to this multidisciplinary group for their diligent work in this role, constructive feedback, experiences shared, and useful suggestions.

The University of Oulu Graduate School (UniOGS) does a great job to, as indicated in their goal statement, “provide the framework and conditions for high-quality, research-driven doctoral education.” I would like to acknowledge their work in helping us doctoral candidates to acquire general and transferable skills through the several courses they organize on ethics, scientific communication, intellectual property rights, and university pedagogy, just to name a few. Special thanks to Anthony Heape and Titta Kallio-Seppä for their guidance on procedures and paperwork.

The final revision of this dissertation in terms of (APA) style, formatting, and conventions was carried out by the series editor, University Lecturer Veli-Matti Ulvinen. I would like to show my appreciation for his rigorousness, precision, guidance, quick reaction, and clear and didactic suggestions.

Although the work on a dissertation is independent, my Ph.D. studies have benefited from the frequent scientific activities organized at the LET research group, including, but not limited to, reading circles, writing Wednesdays, seminars, presentations by renowned international experts, workshops, strategy days, and theoretical discussions. For that, and for an unparalleled work atmosphere, I thank each and every one at the LET team. You have been welcoming and supportive. I am lucky to have you all as colleagues. Special thanks to Márta Sobocinski, Eetu Haataja, and my lunch partner in crime Muhterem Dindar. Together with Márta and Eetu, I started the immersion in the promising world of psychophysiology and interpersonal physiology. We became avid readers on the topic and were often sharing and discussing what we learned and what it meant for our respective research interests. Theories, methods, findings, unsolved questions, challenges... Our discussions were illuminating on both the possibilities and the difficulties. In addition, I thank Márta, the first person I talked to after my arrival at Oulu Airport, for her generosity and invaluable practical support for this adventurer to settle down in his new home and workplace. To Muhterem, I am grateful for countless affable discussions over lunch on the nature and practicalities of science and scientific
lifestyle, as well as on more mundane topics; and most importantly, for his friendship.

In the first year of my Ph.D. studies, I had a special collaborator from the Center of Wireless Communications, M.Sc. Abdul Moiz. Together, we set up our own instance of the Open edX platform on an Ubuntu Server to be used in the second data collection of this dissertation, explored possibilities for real-time biosensor-data collection, and designed and populated a MySQL database, among other engineering tasks. During the months we were closely working together, we faced technical challenges, made design decisions, and in general, worked hard to get the system up-and-running. It was a practical and hands-on work. I would like to thank Moiz for his commitment to the project, hard work, contagious motivation, for fruitful discussions in design meetings, and for so many productive disagreements which led to the best solution for the project. Beyond work, his dependable friendship made a difference and is one of the highlights of my stay in Oulu. Thank you, Moiz.

This dissertation has been supported by the Academy of Finland via the SLAM project (grant number 275440), and the Faculty of Education. I express my gratitude to both institutions. In addition, I consider myself fortunate to have been part of the Faculty of Education these years, a pleasant and supportive working environment, packed with friendly people. I want to thank the Faculty administration and technical staff for their work to support research. I would also like to acknowledge my two long-term office mates, Markus Packalen and Laura Palmgren-Neuvonen. It was a pleasure sharing office with them, and our short conversations on science, everyday life, sports, languages, Finnish culture, travel, and so on, provided an excellent break from periods of intense concentration. I definitely enjoyed my time at KTK328, my research home for most of the Ph.D. studies.

My first scientific research experience comes from the time when I was working on my master’s thesis under the supervision of Associate Professor Pedro J. Muñoz Merino (Universidad Carlos III de Madrid). The work on learning analytics visualizations from a computer science perspective paved the way for my discovery of the learning sciences and the research field of technology-enhanced learning. I am thankful to Pedro for his exemplary work ethics, which inspired me, and for introducing me to this opportunity for doing my Ph.D. studies in a multidisciplinary, international environment.

Scientific writing is a skill that develops over time, and (in)formative, professional commentary directly on our writing equip us with valuable knowledge
of formalities, formatting, conventions, word choice, and clarity, among other aspects. In this sense, I would like to show my appreciation for the work of Scribendi’s editor.

A Ph.D. journey is a roller-coaster ride spanning practically all the spectrum of mental states and emotions: excitement, confusion, enjoyment, frustration, engagement, anxiety, flow, disgust, hope, disappointment, pride, fear, surprise, gratitude, and the list goes on. Therefore, beyond the academic setting, friends and family play an indispensable role in helping us regulate those emotions, keeping the motivation high, putting things in perspective, and strengthening our resilience. In Oulu, I have found an amazing gang from the four corners of the earth, who, in one way or another, have contributed to this journey. There are so many of them that it is virtually impossible to name them all, but huge thanks to each and every one of them for their kindness, time, company, and generosity. In addition to those aforementioned, I would like to offer my special appreciation to my YOK18 partner in crime and close friend Uzair Khan, Laura Jalkanen (to whom I am also indebted for the initial version of the Finnish abstract of this dissertation), Sonia Saher, Lisi Herrera, Onel Alcaraz López, and Ilaria Gabbatore. Friendship blossoms anywhere. Naturally, due to geographical reasons, I have spent more time during the Ph.D. with my friends in Oulu. However, my gratitude also goes to my dear friends from Spain and Cuba, and family therein. Special mention to Alejandro Reyes Bascuñana, Marielena García López, Guillermo López Lagomasino, Carmen Gutiérrez Rodríguez, and Lucía Hidalgo Sánchez. In Oulu, volleyball has been my favorite hobby during these Ph.D. years. Playing volleyball several times a week has proved an effective strategy to maintaining a positive attitude and my signature smile on my face, especially on moments of setback and despair. Accordingly, I would like to acknowledge the host of people with whom I have played volleyball through these years. I am glad I discovered this healthy hobby early enough in my Ph.D. studies.

Last but not least, my whole life would not be enough to thank the two persons this dissertation is dedicated to, my parents. You have stimulated my passion for learning since my early days. You have supported me in every possible way, and you have given your all for my education, not only in terms of knowledge, but also in values and principles. I owe you everything. Thanks for the gift of life, the unconditional love, and for being there through thick and thin. This is for you both.

March 14th, 2019

Héctor J. Pijeira Diaz
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDA</td>
<td>electrodermal activity</td>
</tr>
<tr>
<td>GSR</td>
<td>galvanic skin response</td>
</tr>
<tr>
<td>MSLQ</td>
<td>motivated strategies for learning questionnaire</td>
</tr>
<tr>
<td>PCI</td>
<td>physiological coupling index</td>
</tr>
<tr>
<td>PGR</td>
<td>psychogalvanic response</td>
</tr>
<tr>
<td>RQ</td>
<td>research question</td>
</tr>
<tr>
<td>SCL</td>
<td>skin conductance level</td>
</tr>
<tr>
<td>SCR</td>
<td>skin conductance response</td>
</tr>
</tbody>
</table>
List of original publications

This thesis is based on the following publications, which are referred throughout the text by their Roman numerals:


The articles are co-authored with the three supervisors of this dissertation, who offered guidance during the process. The author of this dissertation, first author in all three publications, was responsible for empirical data collection, data analysis, theoretical grounding and writing.
# Contents

**Abstract**  
17  
**Tiivistelmä**  
9  
**Acknowledgements**  
13  
**Abbreviations**  
15  
**List of original publications**  
17  
**Contents**  
17  
1 **Introduction**  
19  
2 **Theoretical framework**  
27  
  2.1 Collaborative learning  
  
  2.2 Electrodermal activity  
  
  2.3 Arousal  
  
  2.3.1 Cognition, affect, and arousal  
  
  2.3.2 Arousal and performance  
  
  2.3.3 Arousal in the classroom  
  
  2.4 Interpersonal physiology  
  
  2.4.1 Previous work on interpersonal physiology in collaborative learning  
  
  2.4.2 Arousal contagion  
  
  2.4.3 Physiological coupling indices  
  
  2.5 Research gap  
  
3 **Aims**  
49  
4 **Methods**  
51  
  4.1 Participants  
  
  4.2 Materials  
  
  4.2.1 The Empatica® E3 and E4 wristbands  
  
  4.2.2 Performance instruments  
  
  4.2.3 Online learning environments  
  
  4.2.4 Questionnaires  
  
  4.3 Procedure  
  
  4.3.1 Data collection I  
  
  4.3.2 Data collection II  
  
  4.4 Data analysis  
  
  4.4.1 Electrodermal activity  
  
  4.4.2 Performance  
  
  4.4.3 Questionnaires  

17
4.5 Research evaluation ................................................................. 64
4.6 Ethics ......................................................................................... 67

5 Overview of the articles 69

5.1 Article I: Investigating collaborative learning success with physiological coupling indices based on electrodermal activity .......... 69
5.2 Article II: Profiling sympathetic arousal in a physics course: How active are students? ................................................................. 70
5.3 Article III: Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members? ................................................................. 71

6 Main findings and discussion 73

6.1 Triad arousal levels in collaborative learning ........................................ 73
6.2 Arousal during exams and achievement ................................................ 75
6.3 Pairwise directional agreement and dual learning gain ....................... 76
6.4 Within-triad high arousal contagion in collaborative learning ............. 77

7 Conclusions 79
List of references 83
Original publications 99
1 Introduction

Successful collaborative learning endeavors have long been promoted and pursued by educational institutions at national and international levels as a way to attend to students’ academic needs and future employability as well as to the practical consideration of sharing scarce resources (e.g., computers and laboratory equipment; M. Baker, 2015). Collaborative learning has been used in instruction to a greater or lesser extent since at least the second half of the 19th century (Johnson & Johnson, 2002), as it is widely believed that individual outcomes (e.g., learning gain) will be favored by mutual intellectual engagement (Kuhn, 2015). Collaborative learning has been actively and extensively studied since it entered the educational research agenda in the 1970s (M. Baker, 2015), spurred by awareness of societal relevance and embraced by a vibrant community of learning scientists (Hoadley, 2018). Apart from the desired higher order thinking, an early review (Slavin, 1980) reported other benefits of collaborative learning such as the development of prosocial behavior. Furthermore, efficient collaboration is among those skills beyond domain knowledge that are permanently in high demand in today’s extremely specialized labor market. Specialization translates into distributed expertise that, when synergistically combined, constitutes a competitive advantage for organizations. Thus, organizations need expert collaborators or team players to thrive in a highly competitive global ecosystem. Such orientation to collaboration is expected to have already been developed as students advance through their academic years. Therefore, the development of collaborative learning has been embraced and promoted by educational stakeholders including practitioners, organizational administrators, and policy-makers.

Shaped by theoretical and/or technological advances, as efforts to capitalize on the affordances of collaborative learning, a variety of methods has been applied across the decades in schools such as jigsaw (Aronson, Blaney, Sikes, & Snapp, 1978; Aronson & Patnoe, 2011), structured academic controversy, reciprocal teaching, student teams achievement division, and others cited by Koschmann (1996), such as expeditionary learning, group investigation, problem-based learning, and project-based learning, as well as several forms of small-group learning (A. L. Brown, 1992; Slavin, 1980; Webb, 1991). Based on societal needs, it is no surprise not only that collaborative learning pedagogies have thrived, but also that they are increasingly emphasized in educational systems through the design of curricula and instruction (OECD, 2017).
Before going any further, it is important to acknowledge that the terms collaborative and cooperative learning have sometimes been used interchangeably, while at other times an explicit distinction is made between them (M. Baker, 2015; Dillenbourg, Baker, Blaye, & O’Malley, 1996). This dissertation follows the distinction posed by Dillenbourg (1999), in which cooperation is seen as labor division (i.e., partners split the work, solve sub-tasks individually, and then assemble the partial results into the final output), while in collaboration, partners do the work together. Aligned with that distinction, this dissertation adopts the most widely used definition of collaboration (Sanna Järvelä et al., 2015), that of Roschelle and Teasley (1995, p. 70, emphasis added), which states that “collaboration is a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem.” It is acknowledged also that in real-life scenarios, collaboration and cooperation are hardly separable, given the nature of group work, which often requires a combination of the two (Jeong & Hmelo-Silver, 2016). Nonetheless, the focus here is on the joint work of learners and, therefore, on collaborative learning, understood as previously defined.

In terms of learning, the process of collaboration is as important as the outcome, since the aim is often not only that students reach a correct problem solution, but also that they acquire the skills to tackle such problems more efficiently in the future, whether together or alone, in virtue of the appropriation of a co-elaborated, deeper task-domain conceptual understanding (M. Baker, 2015). The success of the collaborative endeavor, however, cannot be taken for granted. Neutral and negative effects of collaborative learning have also been reported in the scientific literature (P. A. Kirschner, Sweller, Kirschner, & Zambrano, 2018; Kuhn, 2015; Mullins, Rummel, & Spada, 2011). On the negative side, for example, it has been found that low achievers progressively become passive when collaborating with high achievers (Dillenbourg et al., 1996), that groups sometimes engage in destructive interactions (Cohen, 1994), including interpersonal aggression (M. Baker, 2015) and blaming each other for making mistakes (Miyake & Kirschner, 2014), and that collaboration may even lead to a decline in thinking quality due to overconfidence produced by the group interaction (Kuhn, 2015). The complexity of the outcomes of collaborative learning, ranging from the positive and highly desirable to the negative and detrimental for learning, has naturally attracted the interest of educational researchers, who want to understand the phenomenon so that, ultimately, collaboration can be properly structured and adequate interventions can be designed.
Three paradigms have been used to study collaborative learning: the effect paradigm, the conditions paradigm, and the interactions paradigm (see Dillenbourg et al., 1996). In the **effect paradigm**, the question is whether collaborative learning is more efficient than learning alone. Positive outcomes have largely dominated the results, but some of the negative effects previously mentioned are stable and well documented (Dillenbourg et al., 1996). The variability in the findings led researchers to investigate under what conditions collaborative learning is efficient, which is the question the conditions paradigm tries to answer. When it comes to the quality of problem solutions and learning outcomes, whether working together is more effective than working alone depends indeed on factors such as the task-complexity, group size, differences in prior knowledge, age, team-member familiarity, and social skills (M. Baker, 2015; P. A. Kirschner et al., 2018). The many independent variables in this paradigm (i.e., the **conditions**) were shown not to have simple effects on the learning outcomes but instead to interact with each other in a complex, virtually inextricable way (Dillenbourg et al., 1996). Research then moved to explore intermediate variables that describe the interactions among group members to understand sources of variability in collaborative outcomes (Barron, 2000). The **interactions paradigm**, with higher granularity than its predecessors in terms of unit of analysis, is concerned with the questions of which interactions occur under which conditions and what effects these interactions do have (Dillenbourg et al., 1996). It was found that the simple frequency of interaction does not predict individual achievement in collaborative learning (Cohen, 1994), pointing to the need of characterizing the interactions.

Theories of learning (e.g., socio-cultural and situated cognition theories) relocating the attention from the individual to the activity structures in which learning occurs, highlighted the role of discourse in conceptual development (Barron, 2000). As an object of study, communication provides a window into the cognitive and social processes related to collaborative skills, such as grounding, mutual goal establishment, progress toward goals, and sharing perspectives (Kuhn, 2015). Types of verbal interactions of interest as learning mechanisms were, for example, negotiation, considered an indicator of joint involvement in task solutions, and argumentation, seen as a possible means for resolving socio-cognitive conflict (Beers, Boshuizen, Kirschner, & Gijselaers, 2007; Dillenbourg et al., 1996). The study of productive verbal communication has made extensive use of conversation analysis (Sacks, 1972), a methodology from the sociological research tradition which continues to be an important tool to analyze and make detailed meaning of interactions in collaborative learning (Stahl, Koschmann, & Suthers, 2014). Audio
and video recordings, together with the development of specialized software, have enabled code and count analysis of content data, which accounts for much of the field’s analytic efforts (Jeong, Hmelo-Silver, & Yu, 2014). However, those analyses are time-consuming, very expensive microgenetic methods and do not scale up well (Wise & Schwarz, 2017). Steps toward automation of content analysis have been taken using natural language processing technologies (Blikstein & Worsley, 2016; Kelly, Thompson, & Yeoman, 2015), but analyses of verbal data are often complex (Chi, 1997) for both human coders and machines. In addition to the cost and limitations of content analysis, and although verbal communication has understandably taken the lion’s share of process analysis focus in collaborative learning, researchers acknowledge that, to properly understand the collaborative process, it is also necessary to take into account other features integral to the collaboration process (e.g., nonverbal communication; M. Baker, 2015). Conversational processes alone do not explain the effects observed (Dillenbourg et al., 1996).

The complex processes of collaboration present a challenge for consistent, accurate, and reliable measurement across individuals and across user populations (OECD, 2017). A wide range of data sources has been used to different extents to study collaborative learning, with increased reliance on questionnaires, interviews, and observation (Jeong et al., 2014). However, process-oriented approaches, encouraged by the interactions paradigm, have made patent the need for new tools for analyzing and modelling interactions (Stahl et al., 2014). Collaborative learning is a process that occurs over time, but quantitative research has traditionally focused mostly on relationships between relatively static input and outcome variables (Wise & Schwarz, 2017). There is room for development, so the field needs to explore a larger repertoire of analytic strategies and data modalities (Jeong et al., 2014; Vogel & Weinberger, 2018), especially to address the need for real-time assessment measures (S. Dawson & Siemens, 2014).

In this context, novel technologies, innovative developments, and applications of established technologies are opening up attractive research opportunities, with wearable sensors being a prominent example. Wearable sensors are increasingly part of everyday life as standalone devices or are embedded into accessories, such as smartwatches, mostly for sports and fitness activity tracking (Swan, 2012). But the affordances of wearable sensors for continuous and unobtrusive measurement of physiological responses and human behavior have found research applications in such areas as healthcare (e.g., Koskimäki et al., 2017; Pentland, 2004) and marketing (e.g., Bettiga, Lamberti, & Noci, 2017), among other fields. In learning
research, the affordances to study cognitive-affective dimensions have also been recognized (Schneider, Börner, van Rosmalen, & Specht, 2015), as well as the “unprecedented insight into the minute-by-minute development” of social interactions (Blikstein & Worsley, 2016, p. 222). Wearable sensors, such as necklace microphones, can be used as sociometric badges (Kim, Mcfee, Olguin, Weber, & Pentland, 2012) to track the structural features of verbal communication, including speech overlapping, distribution, and contributions from individual team members. The biosensor type of wearables provide access to physiological responses, whose capability to make salient psychological cognitive and affective processes is the target of study of the field of psychophysiology (Cacioppo & Tassinary, 1990b).

Fig. 1. Nervous system simplification.

One of the most widely used physiological responses in psychophysiology is electrodermal activity (EDA), which refers to the variation in the electrical conductivity of the skin (M. E. Dawson, Schell, & Filion, 2017; Kreibig, 2010). The electrodermal system has a particularity. This is shown in the simplified scheme of the different branches of the nervous system displayed in Fig. 1, representing how the electrodermal system is innervated only by the sympathetic nervous system, as compared to other systems, such as the cardiovascular and the respiratory systems, which have both sympathetic and parasympathetic innervation. Furthermore, the electrodermal system is the only one in the human body with this feature of exclusive sympathetic innervation, which is why EDA is considered a pure marker of sympathetic activation (i.e., sympathetic arousal; Boucsein, 2012). Arousal refers to the degree of physiological activation and responsiveness triggered by an event, object, or situation during a person’s interaction with the
The psychophysiological value of arousal lies in the indirect access it provides to cognitive and affective processes (Mandler, 1984). Arousal increases with the cognitive process of attention (Raskin, 1973; Sharot & Phelps, 2004), which has been explained through evidence pointing to arousal and attention as sharing the same neural mechanisms (Critchley, 2002). Arousal is also the activation dimension of emotions according to the circumplex model of affect (see Fig. 2 in Sub-chapter 2.3.1), which conceives emotions as states characterized by arousal and valence (i.e., the degree of [un]pleasantness) components (Russell, 1980).

The opportunities are obvious, but wearable biosensors have not yet become popular in learning research (Blikstein & Worsley, 2016). Reasons that might prevent the use of physiological data can be related, first, to the technological equipment needed to collect them (e.g., the equipment can be expensive, or sometimes intrusive), and second, to the substantial domain-specific knowledge base needed to analyze and interpret them (Palumbo et al., 2017).

This dissertation utilizes the Empatica® E3 and E4 multi-sensor wristbands (Empatica Inc., Cambridge, MA, USA), devices specifically designed for research as opposed to consumer-oriented devices in the market with lower accuracy (Garbarino, Lai, Tognetti, Picard, & Bender, 2014), to measure EDA and calculate a number of derived indices of sympathetic arousal and interpersonal physiology, in collaborative learning situations in classroom and classroom-like environments. The coordination and synchronicity characteristics of collaborative learning processes (Roschelle & Teasley, 1995) are also hypothesized to be reflected in mutual influence and shared physiological states among collaborators (Palumbo et al., 2017). Accordingly, there is a need to investigate the characteristics of physiological group responses, such as homogenous activation levels and simultaneous or sequential physiological changes (Knierim, Rissler, Dorner, Maedche, & Weinhardt, 2018). Moreover, physiological metrics of collaboration can be considered universal, as they are objective indices (Chanel, Kivikangas, & Ravaja, 2012; Henning, Armstead, & Ferris, 2009; Montague, Xu, & Chiou, 2014; Picard, Fedor, & Ayzenberg, 2016; Swan, 2012) that are independent of language and culture.

Through three scientific articles, this dissertation explores EDA and derivative measures, such as sympathetic arousal and physiological coupling indices (PCIs), concomitant with individual and interpersonal cognitive and affective processes, respectively, during collaborative learning in a naturalistic setting. Article I studies
PCIs (i.e., group measures attending to different parameters) based on EDA, in relation to subjective and performance-related measures of collaboration. Article II provides a profile of the degree of physiological activation (i.e., arousal) in the classroom, together with the relationship of arousal to achievement on the exam. Finally, Article III explores common arousal levels and possible arousal contagion in classroom groups.

The structure of this dissertation follows. After this introduction, Chapter 2 lays out the theoretical background for the work, which involves collaborative learning research, EDA, arousal, and interpersonal physiology. The theory chapter ends with a sub-chapter reflecting on the research gap from different fields, and on the combination of need and opportunity motivating this work. Chapter 3 concisely presents the aims of the dissertation, which shape the methodology detailed in Chapter 4. The latter concludes with two sub-chapters addressing the research evaluation and ethical considerations. Chapter 5 provides an overview of the three articles on which this dissertation is based, connecting how each article contributes to the overall aims. Also aligned with the aims, the main findings are presented and discussed in Chapter 6. Finally, concluding remarks including significance of the work, implications, applications, and future work, are presented in Chapter 7.
2 Theoretical framework

With the integration of different theoretical frameworks, namely, the theories behind collaborative learning, arousal theory from general psychology, and EDA from the field of physiology, in addition to computer science approaches to physiological data processing, this work shows the commitment and will to keep advancing the interdisciplinarity that has characterized the learning sciences since their early days (Hoadley, 2018), which is a staple of collaborative learning research as well (Stahl et al., 2014).

2.1 Collaborative learning

Three leading theoretical frameworks for collaborative learning have been the socio-constructivist approach, which builds on the work of Jean Piaget and his followers; the socio-cultural approach, rooted in the work of Vygotsky; and the shared or situated cognition theory (see Dillenbourg et al., 1996). The socio-constructivist approach considers that learning is the result of disagreement between two points of view in social interaction, when such a socio-cognitive conflict leads to the coordination of points of view, resulting in an enhanced understanding at the individual level (Doise & Mugny, 1984). In the socio-cultural approach, communication rather than disagreement is seen as the mechanism of learning, with inter-psychological processes happening in the interaction and then being assimilated at the intra-psychological level (Wertsch, 1991). According to the socio-cultural tradition, communication goes iteratively from a dialogue with peers to an internal dialogue with oneself, the reflection part, when the new knowledge is then internalized (Vygotsky, 1978). The two theories acknowledge the role of both the group and the individual in the process of learning and the bidirectional influence between both roles (Butterworth, 1982).

Collaboration is considered to be conducive to higher order thinking skills through effective communication (Cohen, 1994). Verbal interactions activate several cognitive processes, including perception, comprehension, information processing, representation, and anticipation, among others (Ahn et al., 2018). According to Bruner (1994), dialogue might lead to deeper cognition because it involves both the comprehension and production of language, the latter more cognitively demanding than the former. Bruner sees dialogue and discussion as vehicles of learning, which not only lead to higher order processing, but also allow students to share and distribute the cognitive load so that no individual working
memory is required to hold all the information needed to solve a problem (F. Kirschner, Paas, & Kirschner, 2011). Apart from alleviating cognitive load, Bruer reasons that collective working memory renders encoded information more accessible, since the same knowledge may have a different representation in each participant’s long-term memory, and a wider repertoire of cues are then available to the group to trigger information retrieval from individual memories. In other words, collective working memory reduces cognitive load, facilitates positive interdependence, and increases information availability through redundancy (van Bruggen, Kirschner, & Jochems, 2002).

Collaborative processes thought to be beneficial for learning include, but are not limited to, knowledge co-construction (Jeong & Hmelo-Silver, 2016), transactivity or building on each other’s reasoning (Weinberger, Stegmann, & Fischer, 2007), and argumentation (M. Baker, 2015). Other support processes are, for example, the identification of shared knowledge (Roschelle & Teasley, 1995), the establishment of common ground (Reiter-Palmon, Sinha, Gevers, Odobez, & Volpe, 2017), mutual modelling of the collaborative partner’s knowledge state (Dillenbourg, 1999), and coordinating joint efforts (Järvenoja, Järvelä, & Malmberg, 2015). All the previous processes implicitly highlight the notion of synchronicity and simultaneity presumed in collaborative learning, as opposed to cooperative learning, which is characterized by individually solving sub-tasks through division of labor (Dillenbourg, 1999). Joint and/or together are the keywords that prominent researchers in the field consistently use to define collaborative learning or associated processes. Such descriptions include (emphasis added in all cases) “working together” (Cohen, 1994), “mutual engagement of participants in a coordinated effort to solve the problem together” (Roschelle & Teasley, 1995), the notion of “joint attention as social cognition” (Tomasello, 1995), partners doing the work together (Dillenbourg, 1999), “a jointly produced activity” (Enyedy & Stevens, 2014), “an active and joint process” (M. Baker, 2015), and “working together toward a shared learning goal” (Jeong & Hmelo-Silver, 2016), among others. Accordingly, shared cognition is regarded as a critical team state (Reiter-Palmon et al., 2017).

As students learn collaboratively, they often create a joint product, artifact, or solution to the task at hand, and upon a justified and successful process, individual domain knowledge is expected to be enhanced (i.e., learning gains; Rummel, 2018). In this dissertation, the joint product and the pairwise learning gain (i.e., aggregation of the individual learning gains) are referred to as the collaborative learning product and dual learning gain, respectively.
Collaboration can only occur if the group members strive for building and maintaining a shared understanding of the task and its solutions (Barron, 2000). In other words, the students’ attitudes toward collaboration, that is, their motivation to constructively engage in collaborative work, become important factors in predicting the process success (Cohen, 1994). In this dissertation, the term collaborative will is used to designate the students’ attitudes toward collaboration.

2.2 Electrodermal activity

EDA refers to “electrical changes across the skin in areas of the body that are psychologically responsive” (Roy, Boucsein, Fowles, & Gruzelier, 1993, p. v). With a history of more than a century, a number of features have contributed to make EDA one of the most widely used response systems in the trajectory of psychophysicsology (M. E. Dawson et al., 2017; Kreibig, 2010), the discipline concerned with the measurement of physiological changes in order to increase understanding of psychologically relevant phenomena (Cacioppo & Tassinary, 1990b; Lytinen, 1984). A listing of such features follows:

1. EDA is easy to measure with a properly spaced pair of electrodes (Boucsein, 2012). So much so, that in its early decades, it was common for workers in this field to use equipment built in their own laboratories, until a commercially manufactured apparatus became more available (Venables & Christie, 1973).
2. The techniques used to record EDA are completely harmless and risk-free (M. E. Dawson et al., 2017).
3. Electrodermal changes can be easily detected by eye (G. E. Schwartz & Shapiro, 1973). Today, advanced sophisticated algorithms based on the mathematical operation of deconvolution (Benedek & Kaernbach, 2010b) allow for the offline biosignal processing of continuous EDA recordings, but before the invention of the computer, the simple quantification of responses in stimulus-response experimental paradigms proved convenient (G. E. Schwartz & Shapiro, 1973).
4. EDA is considered a remarkably sensitive measure because it changes with low levels of psychological and affective states and may vary with the intensity of the state (Thorson, West, & Mendes, 2018).
5. EDA is relatively inexpensive in terms of the hardware and software required, especially when compared to eye tracking, brain-computer interfaces and facial
expression recognition (Gonzalez-Sanchez, Baydogan, Chavez-Echeagaray, Atkinson, & Burleson, 2017).

6. The exclusive innervation of the sweat glands by the sympathetic nervous system provides a relatively direct and undiluted representation of sympathetic activity (M. E. Dawson et al., 2017). The sympathetic nervous system is the control system for the so-called fight-or-flight response, which reflects a state of activation, thus, the use of EDA to index arousal. Conversely, the parasympathetic nervous system is responsible for bringing the body back to homeostasis, the state of equilibrium corresponding to the so-called rest-and-digest response, thus counteracting arousal. In other words, EDA is a clean measure of sympathetic arousal without parasympathetic interference.

These features of EDA, together with the vast amount of research that EDA has been the object of, have led to its long-term consideration as a well-validated, widely accepted, and perhaps the most readily accessible and simplest indicator of sympathetic arousal (Boucsein, 2012; Critchley, 2002; Wang, 1957). Altogether, EDA has been regarded, not surprisingly, as a valuable addition to the toolkit of scientists, especially for those tackling psychophysiological problems (G. E. Schwartz & Shapiro, 1973).

EDA history dates back to the 19th century. By 1840, it was generally accepted that the body’s electrical characteristics provided a basis for diagnosis and therapy, as part of the view of vital processes as electrical (E. Neumann & Blanton, 1970). Vigouroux (1879), a French electrotherapist and electro-diagnostician, was the first to describe the changes in the electrical resistance of the skin. Féré (1888), usually credited with the discovery of the electrodermal reflex, demonstrated that rapid fluctuations of skin resistance could occur in response to emotional stimulation (Edelberg, 1972b). The report by Féré appears to be the first attempt to use EDA to index a psychological construct (Raskin, 1973). Since then, the amount of work on EDA has been extensively reviewed to summarize and update the understanding of the electrodermal mechanisms, as such, and their applications to psychophysiological research (Boucsein, 1992, 2012; Critchley, 2002; M. E. Dawson, Schell, & Filion, 1990, 2007; M. E. Dawson et al., 2017; Edelberg, 1967; E. Neumann & Blanton, 1970; Venables & Christie, 1980; Venables & Martin, 1967). There have been periods of development and periods of stagnation, the latter owing, at least in part, to the use of inadequate or poorly understood techniques (Venables & Christie, 1973).
It is important to be aware of different terminologies that have been used to refer to the EDA phenomenon. These terminologies not only can be found in today’s scientific literature, but also in classic and still highly relevant work on the topic. Although the formal history of EDA dates back to 1879 as previously indicated, according to Boucsein (2012, p. 2), the term EDA was first introduced by Johnson and Lubin in 1966 “as a common term for all electrical phenomena in skin, including all active and passive electrical properties which can be traced back to the skin and its appendages.” Other well used terms, at the time, were psychogalvanic response (PGR) and galvanic skin response or reflex (GSR). PGR was coined by Veraguth in 1908 when, oblivious to the three previous decades of work on EDA, he believed that he had discovered a new reflex (E. Neumann & Blanton, 1970). As for the GSR, its utilization has been long advised against due to ambiguity in its early use on the one hand, and to the findings that multiplicity and complexity of the EDA phenomena (e.g., spontaneous responses and psychologically elicited changes) cannot be explained simply as a galvanic element, on the other hand (Boucsein, 2012; Venables & Christie, 1973). Accordingly, EDA is now the preferred term over GSR and PGR, for variations in electrical conductance of the skin, including phasic changes that result from sympathetic neuronal activity (Critchley, 2002; M. E. Dawson et al., 2017). EDA is therefore the term used in this dissertation. Notwithstanding, it has to be mentioned that some authors, though acknowledging that GSR is now less often used, consider it “still accepted” (M. S. Schwartz, Collura, Kamiya, & Schwartz, 2017).

The physiological basis of EDA has been ascribed to muscular, vascular, and secretory mechanisms, causing a debate which mounting evidence settled several decades ago in favor of the secretory theory (Venables & Christie, 1973). Today, it is known that the eccrine sweat glands, uniquely innervated by the sympathetic nervous system, are the primary determinants of EDA (Boucsein, 2012; Critchley, 2002). The eccrine sweat glands are coiled, tubular structures that extend from the epidermis to the lower dermis and secrete water, electrolytes, and mucin (i.e., sweat, basically; Critchley, 2002). Eccrine and apocrine are the two forms of sweat glands in the human body, but the former are of primary interest in psychophysiology because, although their primary function is thermoregulation, all eccrine glands are believed to be involved in psychological sweating (M. E. Dawson et al., 2017). The distribution of eccrine glands, however, is not homogeneous over the entire skin. Their highest density (approximately 400/mm²) occurs in palmar and plantar (or sole) regions (Critchley, 2002). This means, on the one hand, that EDA measures in different body parts are not comparable, and on the other hand, that the palm and
the sole are the best areas to record EDA. However, in ecologically valid studies, the placing of electrodes on such regions interferes with the subjects’ activities and is therefore not practical. Consequently, alternatives have been explored and more practical locations for daily life measurements such as the wrist have been found viable (Poh, Swenson, & Picard, 2010; van Dooren, de Vries, & Janssen, 2012). This dissertation employs non-dominant wrist measurements of EDA via an unobtrusive watch-like research-quality off-the-shelf wristband (see Garbarino et al., 2014).

The EDA signal is the combination of a tonic, slow-varying component and a phasic, fast-changing component superimposed on the tonic level (Boucsein, 2012). The tonic component accounts for the level of skin conductance in the absence of any particular discrete environmental event or external stimuli, and it is denominated as the skin conductance level (SCL), after a suggestion by Venables and Martin (1967). The phasic component, of wider use in this dissertation, resembles peaks, any of which are referred to as skin conductance response (SCR), a term coined by Burch and Greiner in 1958 according to Edelberg (1967, p. 44) and endorsed by the Society for Psychophysiological Research in 1967 (Venables & Christie, 1973). SCRs rise steeply to their peak and decline slowly, following an exponential-like decay, to the baseline (Benedek & Kaernbach, 2010b). The succession of SCRs usually results in a superposition of subsequent SCRs, as one SCR arises on top of the declining trail of the preceding one (Boucsein, 2012). Today, advanced deconvolution-based algorithms are available to decompose superimposed SCRs in a way that enhances their quantification as compared to previous historical approaches (Benedek & Kaernbach, 2010b).

It was not an interest in the phenomenon itself which led to the high volume and rapid development of research on EDA, but the possibility of applications to medical diagnosis and to studies of mental processes of interest in a wide variety of fields (E. Neumann & Blanton, 1970). For example, Poh et al. (2010, pp. 1243–1244) cite potential clinical applications of EDA which have been researched, such as screening for cystic fibrosis, classification of depressive illnesses, prediction of functional outcome in schizophrenia, discrimination between healthy and psychotic patients, characterization of sympathetic arousal in autism, early diagnosis of diabetic neuropathy, and providing biofeedback in treating epileptic seizures. Edelberg (1967) cites clinical explorations for dermatological research and neurological diagnosis. Although the clinical applications are many, those outside medical settings (i.e., of interest in psychophysiological research) are not fewer. EDA has been largely explored in relation to cognitive and affective processes,
including, but not limited to, attention (Critchley, 2002; Thorson et al., 2018),
cognitive load measurement (Boucsein & Backs, 2000; Shi, Ruiz, Taib, Choi, &
Chen, 2007), memory retention (Critchley, Eccles, & Garfinkel, 2013), stress
detection (Visnovcova, Mestanik, Gala, Mestanikova, & Tonhajzerova, 2016),
engagement (Gillies et al., 2016), decision-making and judgment (Finger &
Murphy, 2011), self-efficacy (McQuiggan, Mott, & Lester, 2008), goal orientation
(Edelberg, 1972a), the state of flow (Knierim et al., 2018), and the study of the
social psychology of human trust (G. E. Schwartz & Shapiro, 1973). Some of those
processes in the cognitive and affective domains have led to incorporating EDA
into intelligent tutoring systems (Cooper et al., 2009; Strauss et al., 2005) and
intelligent interruption management or interruption notification mechanisms
(Goyal & Fussell, 2017; Knierim et al., 2018).

Yet, as diverse as the interpretations might seem in the cognitive and affective
domains, they all share a common denominator: arousal. Whether explicit or
implicit, the notion of arousal, in particular sympathetic arousal, is present
whenever EDA is being used for psychophysiological purposes, since EDA and
arousal are inextricably linked. A reminder of the underlying rationale follows. The
skin, an organ whose activity is electrically traceable (i.e., EDA), is the only organ
of the entire human body exclusively innervated by the sympathetic nervous system,
which is the controller of the so-called “fight-or-flight” response (i.e., preparing the
body for action, whether physical or intellectual). Consequently, EDA enables
access to sympathetic activity. In other words, in psychophysiological research,
above all, EDA is an index of sympathetic arousal (Boucsein, 2012; Critchley, 2002;
Wang, 1957).

2.3 Arousal

The interest in the study of arousal arises from its role in cognitive and affective
processes (see Sub-chapter 2.3.1) to the extent that it has an impact on performance
(see Sub-chapter 2.3.2). The construct of arousal is at the center of the connection
of body and mind (i.e., cognitive-affective function; Hanoch & Vitouch, 2004;
Mandler, 1984). The notion of arousal was developed hand in hand with the early
work on EDA. According to Neumann and Blanton (1970), Févé (1888) initially
studied the EDA phenomenon as an arousal indicator. He demonstrated that rapid
fluctuations of skin electrical resistance could occur in response to emotional
stimulation (Edelberg, 1967). Thereafter came “a long period of neglect” (Raskin,
1973, p. 126), until further work developed the arousal theory (e.g., Duffy, 1962;
Hebb, 1955). Although arousal is a term of widespread use in psychology, other terms have been used as synonyms in the scientific literature, including activation, energy mobilization, excitation, energy, tension, and activity (Duffy, 1962; Russell & Barrett, 1999).

While arousal measures the degree of physiological activation, and it would therefore be recorded more objectively and precisely through physiological responses, it has to be noted that different scales have been developed to measure it (Bettiga et al., 2017). Such scales might be convenient when biosensors are not available, but research has shown that there is often a mismatch between the two measures (i.e., self-reported scales vs. biosensors). Accordingly, Reisenzein (1983) proposes making a distinction between physiological arousal per se and perceived arousal. Arousal in this dissertation is measured physiologically, as indexed through EDA.

Physiological arousal, in turn, is a complex response which manifests through a variety of central (i.e., electrocortical), motor (i.e., muscle tension), autonomic (e.g., sweating, heart rate, and blood pressure), and endocrine (i.e., hormonal) responses (American Psychiatric Association, 2000). For example, an increase in muscle tension usually goes hand in hand with an increased SCL (Frijda, 1986). However, it should not be assumed that the arousal response in each of the different participating systems is necessarily equivalent, or even related. No correlation has been found between electrocortical and autonomic arousal (Frijda, 1986), and the heart has been reported to slow down even as other indices (e.g., pupil size) suggest an increase of arousal (Kahneman, 1973). In general, the different indices of physiological arousal, whether electrocortical, biochemical, electrodermal, cardiovascular, or skeletal motor, have been shown to correlate poorly, and the term directional fractionation has been chosen to refer to the different patterning (Lacey, 1967). Basically, directional fractionation means that different systems involved in the arousal response show a different pattern, which might even be in different directions (i.e., one response increasing while the other is decreasing). Accordingly, separate types of arousal have been proposed (e.g., sympathetic and electrocortical), which should not be mixed (Frijda, 1986). Sympathetic arousal as indexed by EDA is the type that is the focus of this work.

2.3.1 Cognition, affect, and arousal

Arousal has a bidirectional relationship with cognitive and affective processes (Mandler, 1982). Cognitive and affective drivers have a reflection in arousal levels
(Critchley, 2002; Poh et al., 2010), but, in turn, changes in arousal inform cognition (Critchley et al., 2013; Critchley & Garfinkel, 2018) and affect (Reisenzein, 1983; Schachter & Singer, 1962). It is known, for example, that people rely partly on information from their arousal states in judging their capabilities (Bandura, 1982), and that arousal has an effect on human behavior (Damasio, Tranel, & Damasio, 1991). Thus, the psychology and physiology of arousal have a reciprocal influence on one another. Parenthetically, arousal is also influenced by physical activity and by the intake of stimulant and depressant substances (e.g., caffeine; Hanoch & Vitouch, 2004; Russell & Barrett, 1999), but, obviously, the cognitive and affective drivers are the main interest in psychophysiology, as well as to this work where the focus is on classrooms and classroom-like spaces as students learn science.

For the cognitive part, there is a close interdependence between arousal and attention (Kahneman, 1973; Raskin, 1973; Robbins, 1997; Sharot & Phelps, 2004), which naturally extends to other cognitive processes that presuppose attention, such as memory and decision-making (Garfinkel, Critchley, & Pollatos, 2017). Arousal assists in the mental sharpening that enhances attention to and the processing of potentially important information (Critchley et al., 2013). Empirically, arousal has been shown to increase with attention in relation to eye-catching, engaging stimuli, and attention-demanding tasks (Poh et al., 2010). Such a relationship might be explained by indications from clinical studies that a common neural substrate may be shared by arousal and attention mechanisms (Critchley, 2002). By modulating the selectivity of attention and then increasing attentional time on arousing stimuli, arousal can delay disengagement and influence memory encoding (Fox, Russo, Bowles, & Dutton, 2001). In addition, arousal at the time of memory encoding is theorized to facilitate memory retrieval, more so for individuals with enhanced capabilities to be aware of their arousal and internal bodily states in general, known as interoceptive awareness (Garfinkel et al., 2017). Emotionally arousing incidents and stimuli have been shown to be much better remembered than those without emotional relevance (Marchewka et al., 2016). The representation of bodily changes in arousal may mediate the feeling of knowing ascribed to some familiar memories (Garfinkel et al., 2017). In terms of decision-making, interoceptive signals are said to shape rapid decision-making in stressful contexts (Critchley et al., 2013; Critchley & Garfinkel, 2018). The role of arousal in emotions, discussed next, has been at the center of its consideration as a well-suited tool in decision-making and judgment research (Finger & Murphy, 2011).

Arousal and valence are the two dimensions of emotions according to the classic circumplex model of affect (Russell, 1980). The model is represented
graphically in rectangular coordinate axes as shown in Fig. 2. The arousal axis (vertical) represents a continuum from deactivation to activation (bottom-up), whereas the valence axis (horizontal) represents a continuum from negative/unpleasant to positive/pleasant (left-right). The notion of arousal as an essential component of emotions, however, is much older than the model. The construct of arousal was born together with EDA research through the realization that fluctuations of EDA could occur in response to emotional stimulation (Edelberg, 1967; Féré, 1888; E. Neumann & Blanton, 1970). Later, when the theory of arousal or activation (Duffy, 1962) was further developed, the “general level of energy mobilization” was included as “a major aspect of what has been historically labeled as emotion” (Raskin, 1973, p. 126). Around the same time, Schachter and Singer (1962) formulated their classical cognition-arousal theory of emotion, also known as cognitive-physiological theory of emotion, two-factor theory of emotion, or simply the Schachter theory of emotional states (Reisenzein, 1983). According to the theory, an emotional state is the result of the cognitive appraisal (i.e., labelling) of the state of arousal caused by the emotional stimulus. In addition, arousal feedback can have an intensifying effect on emotional states, and this arousal-emotion relationship is mediated, in part, by causal attributions regarding the source of arousal (Reisenzein, 1983; Schachter & Singer, 1962).

Building on the circumplex model of affect and particularizing for academic emotions, the quality of students’ information processing (i.e., cognition) and the types of strategies unfolded by them have been mapped to the four quadrants of the rectangular coordinate axes formed by arousal and valence (Pekrun, Goetz, Titz, & Perry, 2002), as shown in Fig. 2. The quadrants have been labelled according to the two arousal regions divided by the valence axis ("activating" for the top and "deactivating" for the bottom), and the two valence regions divided by the arousal axis ("positive" for the left and "negative" for the right). Thus, starting from the top right and moving clockwise, the four quadrants are positive activating, positive deactivating, negative deactivating, and negative activating. In the activating region, positive activating emotions have been shown to elicit flexible, creative learning strategies, such as elaboration, organization, and critical evaluation, whereas negative activating emotions, on the other hand, seem to evoke more rigid strategies, such as simple rehearsal and use of algorithmic procedures (Pekrun et al., 2002). In the deactivating region, a shallower, superficial processing of information has been found to predominate for deactivating emotions, whether positive or negative.
Such a model of characterizing emotions with arousal and valence dimensions might prove more useful than distinguishing specific emotions, since the latter might be affected by insufficient discriminant validity owing to the overlapping traits in different emotions (Russell & Barrett, 1999). Moreover, the arousal dimension might be more relevant than the valence dimension when it comes to learning. Activating states, whether positive (e.g., enjoyment of learning and pride of success) or negative (e.g., anxiety, frustration, and confusion) seem preferable. For example, it has been argued that negative activating states, which tend to receive more attention than negative deactivating states (e.g., boredom), could be productive for learning if they are properly regulated, and that they might not even need remediation (R. Baker, D’Mello, Rodrigo, & Graesser, 2010). Conversely, in the negative deactivating area, for example, boredom may impair learning by leading to behavioral or mental avoidance strategies when students face tasks that surpass their capabilities or are not engaging (Pekrun et al., 2002). The case of boredom is of particular interest because it has been reported as more persistent than other negative states, including both activating and deactivating ones (R. Baker et al., 2010).
The cognitive processes and affective states closely related to arousal here discussed, obviously, have an impact on performance too, leading to a transitive relation between arousal and performance.

### 2.3.2 Arousal and performance

Known for over a century, the relationship between arousal and performance is dictated by the so-called Yerkes-Dodson law (Yerkes & Dodson, 1908) and described by an inverted-U function elaborated by Hebb (1955), as illustrated in Fig. 3. When the arousal level is low, so is performance. As arousal rises due to increased attention, interest, and engagement, performance increases up to a level of optimal arousal for performance at the top of the inverted-U. If arousal keeps rising to too high of a level because of excessive stress or strong anxiety, performance is impaired. Such impairment has been explained through a reduction of the processing capacity available for cognitive evaluation (Mandler, 1982; Sanbonmatsu & Kardes, 1988). In the same vein, students with too high levels of test anxiety perform worse on tests, and their overall academic achievement is lower (Aritzeta et al., 2017). Test anxiety, located in the high arousal, negative valence quadrant of the circumplex model of affect (see Fig. 2), can reduce working memory resources and has been shown as the academic emotion most often reported by the students (Pekrun et al., 2002).

The relationship between performance and arousal has been taken advantage of, for example, to explore the development of interruption management systems. Sometimes, external agents (e.g., a teacher) need to give a team either feedback or additional information required to complete a task, while the task is being executed. Such an interruption, although relevant, may result in being disruptive to the work. However, a significant increase has been found in both task-performance and reported collaboration experience when the information was presented during arousal acceleration, compared with random interruption (Goyal & Fussell, 2017).

Studies of arousal in relation to performance and to the cognitive processes and affective states behind performance have been carried out in a variety of settings (Hanoch & Vitouch, 2004). Next, some related work is presented in the particular setting of the classroom, including during exams.
2.3.3 Arousal in the classroom

Though studies of arousal during classroom instruction are rare, some exceptions follow. Arroyo et al. (2009) used raw EDA from high school students taking their regular mathematics class over a period of five days, to determine their affective states in combination with self-reports, a camera for facial expression recognition, a pressure mouse, and a posture analysis seat. More recently, Gillies et al. (2016) reported a case study on the in-class arousal of sixth graders during a unit of their science curriculum, in relation to their beliefs and engagement. They determined arousal through EDA by averaging the signal within time bins. Calderón (2016) tracked cardiovascular measures of arousal during two science education course presentations from two student teachers. However, to the best of our knowledge, arousal has not been measured throughout an entire course. Furthermore, studies from some course sections which employed EDA to determine arousal have not used the frequency of SCRs to operationalize the level of arousal.

When it comes to formal learning situations, exams have received much more attention than everyday classroom lessons, given the special interest on the relation between arousal and performance. Exams are challenging situations often involving stress and anxiety because they have not only psychological implications.
for the students but also practical consequences for their progress in academic life, as well as their future. Real exams may produce different responses than challenging laboratory situations because individuals evaluate exams more seriously and are under more pressure to do well (Spangler, 1997). Learning scientists have long been interested in studying students’ responses to exams (particularly test anxiety) and the implication they have for their performance (Pekrun et al., 2002). The physiological responses in exam situations have been studied from the cardiovascular system, the endocrine system, and the immune system perspectives (Spangler, 1997). However, no exam studies have included pure measures of sympathetic activity (Spangler, 1997), as would be the case in EDA measures. Moreover, wristband EDA measures are less invasive and unobtrusive than the endocrine and immune system measurements involving saliva. Furthermore, the analysis of saliva to assess those physiological responses requires the use of medical laboratory equipment, while, for the wristband EDA measures, once the recording is done, the analysis is carried out using computer software.

As discussed in Chapter 1 and Sub-chapter 2.1, collaborative learning pedagogies are increasingly encouraged and used in the classroom for academic (e.g., stimulating higher order cognitive processes) and practical (e.g., sharing of resources) reasons, among others. It was also discussed that collaboration is not always effective, which has driven research to explore a variety of measures in an attempt to quantify and qualify collaborative processes. Arousal refers to an individual degree of physiological activation and has been a central construct in psychophysiology, the inference of psychological states from physiological responses, at the individual level. But group processes obviously require group approaches. The socio-cultural theory of learning postulates that inter-psychological processes at the group level are then assimilated by the individual at the intra-psychological level (Dillenbourg et al., 1996). Although the field of psychophysiology targets the individual, inter-psychological processes already provided a motivation several decades ago to also explore related inter-psychological processes, giving birth to the sister discipline of social psychophysiology or interpersonal physiology.

2.4 Interpersonal physiology

Once physiological technology developed to a point where empirical studies were easy to carry out, physiological approaches to the analysis of small-group interaction naturally became of particular interest to social psychologists and
psychophysicists (G. E. Schwartz & Shapiro, 1973). From the marriage of social psychology and psychophysiology, the field of interpersonal physiology (DiMascio, Boyd, & Greenblatt, 1957) was born. Beginning in the 1950s, social scientists started collecting physiological data from two or more people in interpersonal interactions to explore commonalities and interdependence between their physiological states (Thorson et al., 2018). Among the array of physiological signals, EDA (and by extension sympathetic arousal) enjoyed privileged attention as “a prime reflector of the social-biological nature of human behavior” (G. E. Schwartz & Shapiro, 1973, p. 412). Ascribed connections of interpersonal physiology to early development, language learning, group engagement, social connection, and group membership might have been responsible for a continuous interest in such an approach and a body of related literature which has kept growing (Delaherche et al., 2012; Palumbo et al., 2017).

Interpersonal physiology refers to “the study of the reciprocal influence of human physiological systems and social systems” (Kaplan & Bloom, 1960, p. 128). In dyads, the stability and influence model (Thorson et al., 2018) considers how a person’s physiology at a certain point in time is predicted by both his/her own physiology at a prior time point (i.e., the stability effect) and by the other dyad member’s physiology at the prior time point (i.e., the influence effect). This model easily scales up to groups of any size. The influence is not explained by physiology itself, as such, but is associated with the psychosocial processes which occur between or among group members (e.g., one person’s heart rate influencing a partner’s heart rate via a verbal outburst of anger; Thorson et al., 2018).

Whether driven by a shared environment, coordinated behaviors, matched responses to a third variable, interpersonal processes, or other factors, the mechanisms behind interpersonal physiology are still unclear (Field, 2012; Palumbo et al., 2017). Evidence from the past decades seems to point, at least in part, to the mirror system, composed of mirror neurons in several brain regions (Rizzolatti & Fabbri-Destro, 2008). The first demonstration of a mirror system in humans dates to 1995, when a study using magnetic stimulation found that “there is a system matching action observation and execution” (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995, p. 2608). Currently, the mirror system is discussed as the evolutionary basis for observational learning (Field, 2012) and might play a role in interpersonal physiology (Karvonen, Kykyri, Kaartinen, Penttonen, & Seikkula, 2016; Rizzolatti & Arbib, 1998).

The variety of methodologies and approaches from different disciplines are reflected in the number of conceptualizations used in the study of interpersonal
physiology, including but not limited to attunement (Field, 2012; Karvonen et al., 2016), compliance (Chanel et al., 2012; Elkins et al., 2009; Henning, Boucsein, & Gil, 2001), concordance (Slovák, Tennent, Reeves, & Fitzpatrick, 2014), co-regulation (Sbarra & Hazan, 2008), coupling (Chanel, Bétrancourt, Pun, Cereghetti, & Molinari, 2013; Singer et al., 2016), covariation (Kaplan, Burch, Bloom, & Edelberg, 1963), entrainment (Dikker et al., 2017), influence (Thorson et al., 2018), linkage (Simo Järvelä, Kivikangas, Katsyri, & Ravaja, 2014), physiological markers of togetherness (Noy, Levit-Binun, & Golland, 2015), and synchrony (Dich, Reilly, & Schneider, 2018; Gillies et al., 2016; Liu, Zhou, Palumbo, & Wang, 2016). Occasionally, these terms imply conceptual differences in what is being measured or are used to refer to specific analytic techniques or theoretical approaches, but a systematic review of the literature shows that this is inconsistent (Palumbo et al., 2017).

Different types of interpersonal relationships have been the focus of studies on interpersonal physiology, such as parent-child, therapist-client, individuals in a couple, artist-audience, conductor-choir, friends, teammates, groups of strangers, and so forth (Palumbo et al., 2017). Physiological interdependence has been found to be greater, for example, between individuals who are in a romantic relationship (Sbarra & Hazan, 2008), in psychotherapists who are more empathetic to their clients (Delaherche et al., 2012), and in game players who report greater rapport with each other (Noy et al., 2015). However, when it comes to students in a collaborative learning setting, the use of physiological data has received little attention, as evidenced by only two papers found in a recent systematic review of the literature including all publication dates through November 2015 (Palumbo et al., 2017).

2.4.1 Previous work on interpersonal physiology in collaborative learning

Early examples are found in the work of Kaplan and associates in the 1960s, which explored sympathetic arousal in relation to the affective orientation of group members in either dyads or four-person groups of medical and nursing students (Kaplan, 1967; Kaplan, Burch, Bedner, & Trenda, 1965; Kaplan et al., 1963). Four decades later, sympathetic arousal commonalities were investigated in dyads of college students in terms of nonlinear dynamics and linkage effects (Guastello, Pincus, & Gunderson, 2006). Physiological commonalities in emotion management and convergence during collaborative processes have been studied at the dyad level.
using a variety of physiological responses and measures, such as electrocardiography, respiration, EDA, temperature, and eye-tracking (Chanel et al., 2013). More recently, physiological commonalities in relation to cognitive and affective states (e.g., mental workload and emotional valence) during a pair-programming task in a classroom environment have been studied using electrocardiography and EDA (Ahonen et al., 2016; Ahonen, Cowley, Hellas, & Puolamäki, 2018). All these previous studies focused on physiological responses from the autonomic nervous system, which can be measured more cheaply, quickly, and unobtrusively than those of the central nervous system (see Fig. 1 in Chapter 1) and are more easily interpreted (Novak, Mihelj, & Munih, 2012). Nonetheless, using portable technology in a high school classroom during 11 lessons, commonalities in electroencephalogram signals have been suggested to predict group engagement and social dynamics (Dikker et al., 2017). None of the studies focused on triads as the unit of analysis.

The dyadic nature of some of the relationships studied (e.g., therapist-client and couples), together with the greater availability of pairwise physiological indices (discussed in Sub-chapter 2.4.3), might be among the reasons why interpersonal physiology has predominantly focused on dyads as units of analysis (Delaherche et al., 2012). However, the general dynamics of triads and larger groups are different than those of dyads. For example, dyads cannot form coalitions and have newcomers or old-timers, nor can there be majority/minority influence in dyads. Also, negotiation and conflict (i.e., argument) are more complicated in triads and larger groups (Reiter-Palmon et al., 2017). When it comes to learning, dyads are often considered to be peer learning, with real collaborative learning beginning with three or more participants. Therefore, the focus of this work is on triads.

2.4.2 Arousal contagion

It is generally recognized in social psychology that the emotional responses of one person may elicit emotional responses from another (Berger, 1962; Reeck, Ames, & Ochsner, 2016). The effect, which has been conceptualized as emotional contagion, is driven by feedback from facial and vocal expressions, postures, and behaviors (Hatfield, Cacioppo, & Rapson, 1994). Emotional contagion can also occur when changes in one’s visceral state during emotion may be mirrored in an observer, inducing a corresponding representation (Critchley et al., 2013). The emotion “caught” by the influenced person can be either similar or complementary, the latter sometimes called counter-contagion (Hatfield et al., 1994). Based on
whether the emotional contagion leads to a similar or complementary valence (i.e., pleasant or unpleasant), for example, Berger (1962) provided “a simple, yet convenient definitional framework” for the social emotions of empathy, envy, and sadism (G. E. Schwartz & Shapiro, 1973, p. 391). According to the framework, empathy is described as the social emotion where the valence is similar (whether pleasant or unpleasant) in influencer and influenced, while in envy, a pleasant state in the influencer causes an unpleasant state in the influenced person, and the other way around in sadism, which elicits a pleasant affect in the influenced based on an unpleasant state of the influencer. Rudimentary or primitive emotional contagion is described as relatively automatic, unintentional, uncontrollable, and largely inaccessible to awareness (Hatfield et al., 1994).

A particular case of interpersonal physiology explored in this dissertation is that of arousal contagion, inspired by the notion of emotional contagion, given that arousal is the activation dimension of emotion according to the circumplex model of affect (Russell, 1980). Arousal contagion helps studying the same phenomena in groups (e.g., to assess collaboration or leadership), without the need to distinguish specific emotions, which could result in insufficient discriminant validity, as there are overlapping traits in different emotions (Pekrun et al., 2002). The motivating affordance is that, as a degree of physiological activation, arousal is objectively measurable and theoretically relevant for the introspection of cognitive and affective states.

2.4.3 Physiological coupling indices

The theoretical relevance of interpersonal physiology for the study of group processes, including collaborative learning, faces the innate challenge of defining and delineating variables of theoretical significance and practical relevance (G. E. Schwartz & Shapiro, 1973), given the variety of physiological signals, the number of features extractable from each at the individual level, and the panoply of aggregation approaches for group level analyses (Delaherce et al., 2012; Kreibig, 2010; Palumbo et al., 2017). The individual features of physiological signals can be descriptive statistics (e.g., mean, median, variance, standard deviation, different percentiles, and interquartile range), which can be applied to all signals, or signal-specific, which depend on the particular signal at hand. For example, studied features from EDA include, but are not limited to, SCL, SCR amplitude, SCR latency, SCR rise time, SCR half-recovery time, SCR peaks per minute, and SCR area under the curve (Boucsein, 2012; M. E. Dawson et al., 2017). As an example
of other bodily systems, tens of signal-specific features have been extracted from respiratory and cardiovascular and signals (Kreibig, 2010). In an attempt to estimate cognitive load, Ferreira and associates (2014) employed up to 128 features from four signals. These examples illustrate the variety of options available at the individual level.

The aggregation of individual features at the group level, or the direct computation of group features from particular parameters (e.g., difference in amplitude, rate of change, direction, and linear relationship) of the individual signals, translates into a variety of indices which are called PCIs in this dissertation. Basically, PCIs are indices of interpersonal physiology.

Building on the indices proposed by Elkins et al. (2009), this dissertation explores as PCIs in relation to some collaborative learning aspects (see Article I), the indices of signal matching, instantaneous derivative matching, directional agreement, Pearson’s correlation coefficient, and a transformation of it. Except from directional agreement, those indices can only be computed on a pairwise basis. Signal matching measures the difference in the amplitude of two signals. Instantaneous derivative matching accounts for the difference in the rate of change, as determined by the first derivative of two signals. Directional agreement measures the percent of time that an arbitrary number of signals is changing in the same direction (i.e., upwards, constant, or downwards), making it a simple yet convenient index for the study of the interpersonal physiology for more than two persons.

Signal matching, instantaneous derivative matching, and directional agreement, applied to cardiovascular and EDA measures, have been explored in connection to team performance, joint activity, and shared trust in technology (Elkins et al., 2009; Montague et al., 2014). Much more often employed in interpersonal physiology has been Pearson’s correlation, a standard and popular statistical tool, which in this case determines the strength of the linear relationship between the physiological signals from two individuals. Although no conclusive results have been obtained, cross correlation has been found to predict task completion time (Henning et al., 2001) to be higher in a conflicting situation than when playing cooperatively in digital gaming (Chanel et al., 2012), to indicate team performance after controlling for task/technology conditions (Montague et al., 2014), and to relate to self-reported social presence in digital gaming (Simo Järvelä et al., 2014), among others. Even half a century ago, correlations of the frequency of electrodermal SCRs over time between two individuals were presumed to relate to rapport (G. E. Schwartz & Shapiro, 1973). However, correlations, although simple, powerful, and easy to
interpret, might not be well-suited for continuously measured physiological signals because data typically show sequential dependency (i.e., autocorrelations) and non-stationarity over time (i.e., changing mean and variance), which violates general linear model assumptions of data independence and stationarity (Palumbo et al., 2017). In an attempt to overcome such limitations, some transformations, such as the Fisher’s z-transform, and sophisticated statistical techniques, such as dynamical correlation, have been proposed. The Fisher’s z-transform consists of applying the inverse hyperbolic tangent to the correlation coefficient, in order to stabilize the variance and obtain a normally distributed index (Chanel et al., 2013; Simo Järvelä et al., 2014). Dynamic correlation is a nonparametric approach that places no assumption on homogeneity of the sample, although it has limitations of its own and complicates interpretation (Liu et al., 2016).

Although not applicable to the EDA signal, other PCIs found in the interpersonal physiology literature are, for example, dynamic common activation and weighted coherence. Dynamic common activation reflects both the strength and variation of functional magnetic resonance imaging activation across subjects (Singer et al., 2016). Weighted coherence (Porges et al., 1980) is a frequency domain PCI which quantifies the similarities of two individuals’ physiological responses on a specified frequency band regardless of phase differences (Montague et al., 2014). Weighted coherence is calculated over a range of frequencies rather than at only a specific frequency, which makes it suitable to capture the inherent variability in frequency of interpersonal physiological indices during active behaviors (Henning et al., 2009). Naturally, being a frequency domain index, it is mostly useful for periodic physiological signals, such as heart beat and respiration rate (Chanel et al., 2013). EDA, the subject of this dissertation, is not a periodic signal. Therefore, weighted coherence was not considered among the PCIs.

In sum, PCIs can play a role in studying social interactions (e.g., as in collaborative learning and team processes) by providing an objective measure, but research in this direction needs further explorations to keep enhancing our understanding on the affordances of interpersonal physiological processes to index interpersonal psychological processes related to cognition and affect and, ultimately, to learning.

### 2.5 Research gap

Collaborative learning is extensively researched (M. Baker, 2015; Jeong & Hmelo-Silver, 2016). Yet, the field is missing continuous measures with real-time potential,
as collaboration is a process whose evolution in time is of interest and a determinant of success (Dillenbourg et al., 1996; Jeong et al., 2014). These measures have been also referred to as online measures, meaning that they are taken as the process unfolds, and not before or after. In addition, the field is on the lookout for measures complementing subjective data as coming from self-reports (Wise & Schwarz, 2017).

EDA has a long history of research, over a century old (Boucsein, 2012). However, it has not been until recently that it has become easily and unobtrusively measurable in the wild, outside laboratory settings, thanks to the increasing availability and technological quality of wearable biosensors (Garbarino et al., 2014; Torniainen, Cowley, Henelius, Lukander, & Pakarinen, 2015). Yet, students remain an understudied population, as clinical studies take the lion share of EDA research (M. E. Dawson et al., 2017; Poh et al., 2010).

In psychology, arousal is a commonly studied construct due to its connection to cognition and affect, and the natural psychological interest in these domains (Kuppens, Tuerlinckx, Russell, & Barrett, 2013). Yet, we know little about arousal in the classroom and its patterns. In the few studies tackling arousal in the formal academic setting of the classroom (e.g., Arroyo et al., 2009; Gillies et al., 2016), the affective component steals the show, while it is rare to find interpretations or foci on the realm of cognitive processes such as attention. A little more studied have been arousal or physiological measures in exams situations (Spangler, 1997; Spangler, Pekrun, Kramer, & Hofmann, 2002). However, measures of arousal based on the frequency of SCRs have not yet been tested during exams in connection to students’ performance.

Interpersonal physiology has long been exploring the commonalities and influence between subjects with certain relationships, predominantly parent-child, therapist-client, and individuals in a couple (Palumbo et al., 2017). When it comes to individuals collaborating, workplace performance is a more frequent target than collaborative learning in schools (Elkins et al., 2009; Henning et al., 2001). In terms of group size, dyads have received much of the attention, while other common group sizes in collaborative learning such as triads have been neglected (Reiter-Palmon et al., 2017; Thorson et al., 2018).

In this context, this dissertation has found a research gap to address, matching multidisciplinarity, interest, opportunity, and relevance. This work contributes to the particular case of collaborative learning by investigating PCIs, arousal levels, and arousal contagion, as such, and in relation to outcomes as learning gain and academic achievement.
3 Aims

This dissertation pursues the exploration of EDA and derivative measures, such as sympathetic arousal and PCIs, concomitant with individual and interpersonal cognitive and affective processes, during collaborative learning in a naturalistic setting. To that extent, the main aims of this dissertation are to:

1. investigate the interpersonal physiology of EDA in connection to collaborative learning processes (addressed in Article I: research question [RQ] 1, RQ2; Article III: RQ1, RQ2),
2. profile sympathetic arousal in the classroom at both the individual and interpersonal levels in collaborative learning (addressed in Article II: RQ1, RQ2, RQ3; Article III),
3. examine both the individual and interpersonal physiology of EDA in relation to academic achievement (addressed in Article I: RQ2; Article II: RQ4), and
4. explore sympathetic arousal contagion among group members as a result of the cognitive-affective social interactions in collaborative learning (addressed in Article III: RQ3, RQ4).

In order to meet the aims, two data collections were organized, one in a classroom-like research environment of the University of Oulu (see Article I) and another during two consecutive runs of a regular course in a real classroom (see Articles II and III).
4 Methods

Two research designs were conceived for this dissertation, having ecological validity in mind. To each design, a data collection process was associated. There were similarities and differences between the designs. The participants in the studies were Finnish high school students. A classroom-like environment and a real classroom were used for collecting the data in naturalistic locations. The topics students were dealing with during the data collections were part of their curriculum and the tasks were either embedded in or regular part of their coursework. The materials involved a biosensor wristband (Empatica®) to record EDA, performance instruments (e.g., knowledge tests, course exams), online learning environments, and questionnaires. Specialized computer software was used for the analysis of the relevant variables according to the aims. In the light of the reproducibility principle, this chapter presents the research methodology in detail, an overview of which is listed in Table 1.

Table 1. Method overview.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Data collection I</th>
<th>Data collection II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>N = 24; high school students</td>
<td>N = 24; high school students</td>
</tr>
<tr>
<td>Location</td>
<td>Classroom-like</td>
<td>Classroom, classroom-like</td>
</tr>
<tr>
<td>Topic</td>
<td>Nutrition</td>
<td>Advanced Physics</td>
</tr>
<tr>
<td>Duration</td>
<td>2 h 15 min</td>
<td>2 x (75 min, 3 x week, 6 weeks)</td>
</tr>
<tr>
<td>EDA sensor</td>
<td>Empatica® E3</td>
<td>Empatica® E4</td>
</tr>
<tr>
<td>Performance measures</td>
<td>Pre- &amp; post-test, joint solution</td>
<td>Course exam</td>
</tr>
<tr>
<td>Online learning environment</td>
<td>weSPOT</td>
<td>Open edX</td>
</tr>
<tr>
<td>Questionnaire(s)</td>
<td>Motivated Strategies for Learning Questionnaire (MSLQ)</td>
<td>MSLQ and a questionnaire adapted from Schmitz and Skinner (1993)</td>
</tr>
<tr>
<td>EDA analysis software</td>
<td>EDA Explorer</td>
<td>Ledalab</td>
</tr>
<tr>
<td>Variables</td>
<td>PCIs, collaborative will, collaborative learning product, dual learning gain</td>
<td>Sympathetic arousal: level, direction, and contagion; exam grades</td>
</tr>
</tbody>
</table>

4.1 Participants

The participants in the two data collections for this research were high school students from the University of Oulu’s Teacher Training School.
Participants ($N = 24$) in the first data collection (see Article I) were 13 males (54%) and 11 females (46%) aged 16 to 19 years ($M = 17.3; SD = 0.62$).

Participants ($N = 24$) in the second data collection (see Articles II and III) were randomly selected from those enrolled in the regular, elective Advanced Physics course, which runs twice during the spring term. Twelve students were recruited in each course run as participants. The gender distribution was six females (25%) and 18 males (75%) aged 16 to 17 years ($M = 16.5; SD = 0.5$). Their attendance average was $10.9/12$ ($SD = 1.3$) for the first course run and $10.5/12$ ($SD = 1.3$) for the second. They were high achievers in the preceding physics course, obtaining an average grade of 9.0 out of 10 ($SD = 0.6$).

4.2 Materials

The materials for this study can be divided into four groups: the EDA sensors, performance instruments, online learning environments, and questionnaires.

4.2.1 The Empatica® E3 and E4 wristbands

Essential tools in the data collection for this dissertation were the Empatica® E3 and E4 (Empatica Inc., Cambridge, MA, USA) biosensor wristbands (see Fig. 4 for pictures of the E4; the E3 is similar), since physiological data, particularly EDA, is at the core of this research. The Empatica® E3 was used for the data collection I (see Article I), and the E4 for the data collection II (see Articles II and III). The Empatica® E3 and E4 wristbands are equipped with four sensors: a photoplethysmograph to measure blood volume pulse, an EDA sensor, a 3-axis accelerometer, and an infrared thermopile to measure peripheral skin temperature. The EDA sensor, the one used in the studies of this dissertation, has a sampling frequency of 4 Hz (i.e., it takes four measurements per second). The Empatica® E3 and E4 EDA sensor utilizes the exosomatic method to measure EDA, which consists of applying a small external current by means of two electrodes and determining the skin electrical conductance through Ohm’s law (Edelberg, 1967). The exosomatic method is usually credited to an 1888 study by the French neurologist Charles Féré, but was already described by the French electrotherapist and electrodagnostician Romain Vigouroux in an 1879 study (Edelberg, 1967). The units for the skin electrical conductance are microSiemens ($\mu$S). On a historical note, resistance, the electrical inverse of conductance, was also used in the early days of EDA research. However, in 1970, the editor of the journal
Psychophysiology proposed the elimination of resistance measures and the use of conductance only when measuring exosomatic activity, based on the peripheral mechanisms of the skin (Venables & Christie, 1973, p. 6).

The Empatica® E3 and E4 have two modes of operation: recording in internal memory and real-time streaming via Bluetooth®. The former was deemed more suitable for the studies of this dissertation, which used offline analysis of the data. Data recorded in internal memory are stored in a comma-separated values file for each sensor and are accessible via a USB connection. The EDA file consists of a first row with the recording starting time, a second row with the sampling frequency (4 Hz), and the continuous EDA measures in the subsequent rows. The Bluetooth® streaming affords applications for learning beyond research, such as for learning analytic dashboards, intelligent tutoring systems, and implementation of alerts in critical moments, among others. Such potential applications are facilitated by a software development kit, including a mobile programming interface that enables the developing of apps to leverage real-time data from the multisensor wristband.

![Fig. 4. Front and back view of the Empatica® E4 watch-like biosensor wristband (photos by courtesy of Empatica Inc.).](image)

Although the two data collections involved a larger number of participants, for the studies of this dissertation, participants were limited by the biosensor wristbands available. At the time of the first data collection, our laboratory was equipped with six Empatica® E3 wristbands, and the number rose to 12 Empatica® E4s for the second data collection, in a purchase deal that involved returning the E3s. The Empatica® E3 and E4 are research devices, more accurate and precise than
consumer-oriented biosensor wristbands available in the market, but also several times more expensive, which limited the number that our laboratory was able to acquire, as has happened in other studies also (e.g., Koskimäki et al., 2017).

4.2.2 Performance instruments

The performance measures in the first data collection included a pre-test, a post-test, and a collaborative learning product (i.e., the groups’ joint solution to the task), while in the second data collection, a real-life advanced physics course exam was used. The pre- and post-tests were identical: multiple-choice tests consisting of ten questions with four options each. An example of a question is: “Fiber intake is important because it: a) allows for consistent energy supply, b) maintains bowel health, c) lowers cholesterol levels, and d) contains calcium.” In total, out of the 40 choices, 22 were correct. Shared spreadsheets using Google Sheets technology were provided for students to collaboratively report on their joint solution to the task. The physics course exam consisted of three sections (A, B, and C) with three questions each. Section A was theoretical and focused on the definition, connection, and explanation of concepts and phenomena (e.g., resonance, interference, and Doppler effect). Sections B and C had a more practical character and focused on calculations and graphical representations. The students had to choose and answer two questions from each section, except from section C, where only one question was required to be answered. Thus, students had to answer in total five questions on the exam.

4.2.3 Online learning environments

Two online learning environments were used in this study: weSPOT (Mikroyannidis et al., 2013) and Open edX (https://openedx.org/) for the first and second data collections, respectively. WeSPOT stands for working environment with social and personal open tools for inquiry-based learning. Accordingly, weSPOT scripts collaborative learning in compliance with an inquiry-based learning paradigm (Edelson, Gordin, & Pea, 1999) consisting of five steps: 1) plan the design method, 2) set the criteria, 3) collect the data, 4) discuss the findings, and 5) communicate the results. Open edX is an open source software which powers the edX platform (https://www.edx.org/) for so-called massive open online courses (MOOCs). It was chosen because of its customization affordances since it is open source and utilized to create a dedicated online course following the
traditional course structure as provided by the teachers (e.g., course content, additional support learning material, simulations, and videos).

### 4.2.4 Questionnaires

The MSLQ (Pintrich, Smith, Garcia, & McKeachie, 1991, 1993) was used in both data collections to assess the learning regulation profiles of the students, which served as criterion for the formation of groups as heterogeneously as possible. The MSLQ is one of the most frequently used inventories for that purpose (Bjork, Dunlosky, & Kornell, 2013; Panadero, 2017). The questionnaire consists of 81 items organized in two sections, motivation and learning strategies, which contain 6 and 9 scales, respectively. Each item is rated on a Likert scale from 1 (not at all true of me) to 7 (very true of me).

The peer learning scale of the MSLQ—comprising items 34, 45, and 50—was also utilized to assess the student’s predisposition to collaborate in a learning task (i.e., collaborative will; see Article I). The items of this scale account for the students’ propensity to explain learning material to group members, work together with others to complete an assignment, and discuss the task in the group.

A questionnaire (see Table 1 of Article II) to assess students’ perception of cognitive, affective, motivational, and collaborative aspects on a lesson basis was adapted from the one employed by Schnitz and Skinner (1993) for an experiment in a similar setting, a naturalistic study with data collected in the classroom over four months. The questionnaire consisted of four items asking students to rate their task domain knowledge, mood, motivation, and group functioning, on a scale of 1 to 10. The simplicity principle prevailed in choosing and tailoring this questionnaire given the limited time available for these measurements in the naturalistic setting of our study: a real high school course, packed with content and coursework, including both paper-and-pencil tasks and hands-on experiments. Conveniently, the questionnaire was embedded in the dedicated course created on the Open edX platform. In this way (i.e., simplicity and easiness), and since the students were expected to complete the questionnaire in every single lesson, we minimized the impact on the measurement environment (Winne & Perry, 2000). While larger inventories are considered to be more reliable, rating scales based on fewer items are acknowledged to be economical and useful under specific circumstances (Pekrun et al., 2002).
4.3 Procedure

Two data collections were organized targeting EDA specifically, but also performance measures and questionnaires as data sources. Two important issues involving EDA recording were considered in both data collections: room temperature and biosensor wristband adjustment. Room temperature for EDA measures is recommended to be 23 °C (Boucsein, 2012). No adjustment was required in this regard since the temperature in the locations for both data collections was controlled by a thermostat, whose standard setting coincided with the recommended temperature. Also, according to the good practices of recording EDA, students were instructed to wear the biosensor wristband on their non-dominant hand in order to reduce the effect of movement artifacts and to adjust the wristband strap neither loosely nor tightly so that the electrodes made proper but not excessive contact with the skin, which would result in a pressure artifact (Edelberg, 1967).

In both data collections, students worked in triads that were formed based on the principles of gender diversity (Bear & Woolley, 2011) and heterogeneity of learning regulation profiles (i.e., according to the MSLQ scores), for the sake of between-triad comparability.

4.3.1 Data collection I

Two weeks before the collaborative task, the students were asked to answer the pre-test and the MSLQ at their school. The collaborative task was carried out in a classroom-like research infrastructure of the University of Oulu (i.e., the Learning and Interaction Observation Forum; http://www.oulu.fi/leaf-eng/). The experiment was organized in four sessions with 12 different participants each \( (N = 48) \), although, for the purpose of this dissertation, only participants wearing the biosensor wristband were considered \( (N = 24) \), that is, six participants in each session. The sessions took place in the morning and the afternoon of two consecutive days for a total of four sessions. The duration of the sessions was 2 h 15 min. The first 20 min were for preparation and instructions, followed by 1 h 40 min to work on the actual task and 15 min for the post-test. In the preparation phase, students were introduced to the materials, particularly the biosensor wristband, the weSPOT environment, and the collaborative learning task.

Students sat at individual tables clustered in four groups of three. Each table was supplied with a digital tablet to support the task execution. The students in two
groups were each provided with an Empatica® E3 biosensor wristband to measure EDA.

The task consisted of the design of adequate, healthy breakfasts for at least one of several cases corresponding to people with different nutritional requirements, such as a marathon runner in training, a teenager, a patient with a heart condition, a diabetic person, and someone wanting to lose weight. Relevant variables of the target hypothetical case, needed to carry out the task, were provided, such as age, height, weight, and lifestyle details or health issues with implications for the breakfast characteristics. The task was designed to fit in the students’ science curriculum by being aligned with their current knowledge and the content and goals of their science studies.

On the tablet provided to each student, a shortcut to two Google-Docs® files was available: a document and a spreadsheet. The document detailed the nutritional needs of marathon runners. The spreadsheet had a template with rows for food items and columns for the weight (in g) of the different nutrients contained by the food item. The template included several examples for illustration purposes. Upon task completion, students delivered their solution through the weSPOT environment. Finally, the students were asked to answer the post-test.

### 4.3.2 Data collection II

Ecological validity was at the center of the second data collection. Therefore, the classroom was the chosen setting. Using Empatica® E4 biosensor wristbands, EDA was measured in 12 students simultaneously in each of two consecutive runs of an elective, advanced physics course (including the exam), as the course is offered twice during the spring term. The course consists of 18 lessons, plus the exam. There are lessons three times a week, lasting 75 min each. The students put the wristband on at the beginning of each lesson and took it off at the end. Topics of the lessons include, among others, wave formation and interference, reflection and refraction, sound waves, and geometrical optics. In general, the lessons begin with a theoretical part (i.e., the teacher’s explanation of the topic) followed by a collaborative practical activity with paper-and-pencil problems, or hands-on experiments, and instructions and support material in the Open edX environment. In the practical activity, students put the learned physical theories, laws, and principles to the test in groups of three. The groups remained the same for the duration of the course. The students accessed Open edX during class on the tablets assigned to them by the school as a technological support to their studies. At the
end of each lesson, students were asked to answer the questionnaire adapted from Schmitz and Skinner (1993), which was embedded in the Open edX environment.

4.4 Data analysis

Aligned with the materials and data collected, the analysis targeted three data modalities. First, the EDA data collected by the biosensor wristbands were processed to extract different relevant indices in accordance with the aims. Second, the variety of performance instruments were assessed. Third, questionnaire data were prepared for correlation analysis. Next, the analysis of these three data modalities is detailed.

4.4.1 Electrodermal activity

Two approaches are used to analyze the EDA data, one utilizing the EDA signal as such (see Article I), and the other based on sympathetic arousal, a derivative measure from the EDA signal (see Articles II and III). For efficient use of the storage capacity in internal memory, the biosensor wristband by design records the data in a comma-separated values file containing a single column, where the first row displays the starting time in UNIX format (i.e., the number of seconds elapsed from 00:00:00 UTC on January 1st, 1970), the second row stores the sampling frequency in Hz, and the actual EDA signal values (in $\mu$S) start from the third row on. Using the initial time and the sampling frequency (number of samples taken each second), together with the position of each value, it is straightforward to obtain the time-value pairs needed for synchronization and analysis. Therefore, common to both approaches is an initial data pre-processing to transform the file provided by the biosensor wristband to a data format where each EDA value appears next to its corresponding timestamp. Next, the specificities of each approach are discussed.

Approach I: Electrodermal activity signal based

In this approach, used in Article I, the EDA signal constitutes the pillar of the analysis. The EDA signal is susceptible to pressure and movement artifacts (Edelberg, 1967). Pressure artifacts are linked to wristband maladjustment when it is put on too tight. To reduce pressure artifacts, students were instructed to wear the wristband not too tight. As for reducing movement artifacts, as explained earlier, students were asked to wear the biosensor wristband in their nondominant hand.
Notwithstanding, it is necessary to identify possible artifacts in the data to be able to discard them in the analysis phase for the sake of the reliability of the results. The web-based tool EDA Explorer (Taylor et al., 2015) was used to automatically detect artifacts. EDA Explorer employs a machine learning classifier to determine on a 5-second epoch basis whether it corresponds to noise or a clean signal. Reportedly, the accuracy of such a binary classifier surpasses 95% (Taylor et al., 2015). The outcome of applying EDA Explorer to the EDA signal recorded in the first data collection was that 490,880 from 547,848 data points (i.e., 90% of the signal) were classified as clean by the algorithm. Consequently, those were the data selected for the next analysis steps, while the other 10%, labelled as noise, was discarded at this point.

The PCIs to be computed from the clean EDA look at different parameters of the signals. Some of these PCIs are robust in respect to individual differences in the SCRs (e.g., directional agreement is immune to anything related to amplitude as it only considers the direction), while others are highly sensitive to it (e.g., the signal matching precisely accounts for amplitude differences). Therefore, especially for the sensitive cases, it is necessary to normalize the signal so as to make it comparable across group members. The individual differences are caused, for example, by the thickness of skin (Braithwaite, Watson, Jones, & Rowe, 2015), especially the thickness of the corneum (M. E. Dawson et al., 2017), which is the outermost layer of the epidermis. Normalization for comparability purposes was then accomplished by calculating the t-scores (i.e., subtracting the sample mean from every value and then dividing by the sample standard deviation). Both sample $M$ and $SD$ were computed on an individual basis.

The final step before computation of the PCIs is the synchronization in each group of the clean, normalized data for all group members. Since each individual signal has a timestamp for each data point, the synchronization process can be basically seen as the alignment of the timestamps at the group level. Although, the synchronization step was carried out after normalization, the reader should note that both processes are totally independent and, as such, can be executed in any order without affecting the final result.

The clean, normalized, and synchronized data are the base for the computation of the PCIs. Since, by definition, most of the PCIs are computed pairwise, with directional agreement being an exception as it easily extends to groups of any size, the PCIs were obtained pairwise for all triad member combinations (AB, AC, and BC). The PCIs used are signal matching, instantaneous derivative matching, directional agreement, Pearson’s correlation coefficient, and Fisher’s $z$-transform.
Signal matching consists first of a between signal differentiation and then of an aggregation. This is, instantaneous absolute differences in the two students’ signals were computed and then aggregated as the mean of the entire collaborative learning session.

Instantaneous derivative matching compares the slopes of the two students’ signals. Mathematically, this is expressed in Eq. (1) as:

$$\frac{1}{T} \sum_{t=0}^{T-1} |(a_{t+1} - a_t) - (b_{t+1} - b_t)|,$$

where $T$ is the time in number of samples of the signal, $a$ and $b$ refer to the signal of each student, respectively, $t$ refers to the current data point, and $t+1$ to the next data point.

Directional agreement refers to the percentage of the time in which the signals (EDA in this case) of the students are going in the same direction, that is, simultaneously rising, falling, or staying constant. For the computation of directional agreement, first it was determined whether there was instantaneous agreement or not, and then the percentage of agreement was obtained.

Pearson’s correlation coefficient needs no further explanation as it is a standard and widely used measure in statistical analysis. Fisher’s $z$-transform is a transformation of Pearson’s correlation coefficient—by applying the inverse hyperbolic tangent—so as to obtain a normally distributed index.

**Approach II: Sympathetic arousal based**

In this approach, followed in *Articles II and III*, EDA is not used as such, but as the means to index sympathetic arousal, the degree of physiological activation, in this case from the sympathetic nervous system. Since arousal was to be indexed by the frequency of SCRs, the MATLAB-based Ledalab (version 3.4.9; http://www.ledalab.de/) software was used for the detection of SCRs. It allows for the separation of superimposed responses using continuous decomposition analysis by means of nonnegative deconvolution (Benedek & Kaernbach, 2010b). Thus, Ledalab helps to address the traditional bias of counting the SCRs without accounting for those in a superimposed response, a classical method known as trough-to-peak (Boucsein, 2012). Ledalab is also the tool for the analysis of the EDA signal recommended by the E3 and E4 biosensor wristband manufacturer (Empatica Inc., 2018).
The Ledalab algorithm assumes raw, unfiltered data as the input for decomposition analysis and feature extraction of the EDA signal (Benedek & Kaernbach, 2010a). Therefore, no signal preprocessing, such as cleaning or filtering, was performed on the EDA signal as obtained from the wristband sensor.

In determining which change in conductance will be considered an SCR, an amplitude threshold needs to be specified as parameter. For this dissertation, 0.05 $\mu$S have been chosen for the threshold as it is a long-time standard owing to the maximum sensitivity of earlier EDA sensors (Braithwaite et al., 2015; Venables & Christie, 1973, 1980).

By means of Ledalab, the SCR onsets (i.e., the starting point of the peaks) were determined from the EDA recordings. The onsets were used to compute the SCR frequency on a minute basis, to be used as a measure of arousal in peaks/min (ppm). The calculation of SCR frequency followed a moving window (width 1 min; step 250 ms) approach, so that arousal at each instant would be the count of SCR onsets in the previous minute. Arousal was thus computed every 250 ms, the maximum temporal resolution available as it is the sampling interval of the Empatica® E3 and E4 EDA sensors. The frequency approach has the advantage of not having to standardize the data for comparison across individuals, as amplitude measures are not considered. The frequency on a minute basis and the total count of SCRs or arousal episodes were both used in the analysis to answer different RQs, according to their scope and target.

The SCR frequency was also used as the criterion to categorize arousal into low, medium, and high levels, based on the literature. Typically, a frequency of 1–3 ppm occurs at rest (M. E. Dawson et al., 2017, p. 225), and as frequency increases with the arousal level, values higher than 20 ppm are interpreted as high arousal (Boucsein, 2012, p. 222). Accordingly, frequencies of up to 3 ppm were labelled as low, from 20 ppm and up were labelled as high, and anything in between was labelled as medium. The incidence of each arousal level was determined for every student in terms of percentage relative to the time of each lesson. The persistence of each occurrence of an arousal level for every student was computed as the continuous duration of the occurrence (i.e., without switches of level in between).

Notwithstanding the precautions taken in instructing students on the correct placement of the biosensor wristband, its occasional maladjustment resulted in the measurement of basically noise rather than an actual EDA signal, due to inadequate electrode-skin contact. Therefore, an important part of the analysis consisted in the removal of those cases for the sake of the reliability of the results. A technique for automatic assistance to such a purpose was developed for this dissertation, building
on the typical frequency of SCRs at rest (1–3 ppm). Essentially, based on the minimum of 1 ppm as the worst case, the students’ EDA recordings where the number of SCR onsets was less than 1 ppm times the lesson duration in min were not considered in the analysis. That was the case for 137 (33%) of the EDA recording sessions during the course lessons. Due to students being absent for some lessons, 46 (11%) of the expected recording sessions were missing. Taking together the missing sessions and those discarded, 237 (56%) were available for further analysis at the individual level (see Article II). As for the group level (see Article III), it was necessary to identify the cases where valid EDA data were available for all members of a triad in a certain lesson, since the unit of analysis in this context was the triad. Eighteen such cases (corresponding to 18 x 3 = 54 recordings) were found, meaning that in the final sample for group level analysis, 7 of 8 triads and 15 of 35 lessons were represented. Fortunately, in the exam, there were no absent students, and no EDA data had to be discarded as noisy.

In the previous approach (see Approach I: Electrodermal activity signal based) one of the PCIs used was directional agreement, computed directly on the EDA data. For Article III, directional agreement was selected among the PCIs for its easy and direct calculation for groups of any size, since the other indices can only be computed pairwise, and also since it produced the better results in the analysis for Article I. In addition, since the focus of Article III was on sympathetic arousal in collaborative learning, directional agreement was applied to the arousal signal derived from the EDA signal, rather than to the EDA signal as such as in Approach I. Moreover, Approach I applied directional agreement to the whole session, as this approach uses a moving window allowing for a continuous measure of directional agreement for every instant during the lesson after the second initial minute. The latter is because a first minute is necessary for the initialization of the arousal measure, and a second minute is necessary for the first computation of directional agreement based on those initial arousal values. Previously, directional agreement has been calculated immediately from data points (e.g., Elkins et al., 2009; Montague et al., 2014), but it was noticed from the data that sometimes there are instantaneous fluctuations that do not actually represent the direction when the adjacent samples are considered. For example, there can be an instantaneous small increase in one value in respect to the previous one, but this could actually be a tendency to decrease if we look at the neighboring data. To overcome this issue, it was decided to base the direction calculation on one-minute trend lines rather than on two consecutive values which are 250 ms apart. Trend lines were determined for one-minute windows every 250 ms (moving window), and the sign slope of the
trend line was used as the measure of direction for that instant, namely, upward (+1), constant (0), or downward (−1). This means that the arousal direction at each instant was the sign of the slope of the trend line involving the previous minute of data. Trend lines were obtained using MATLAB’s built-in `polyfit` function. Finally, again using a moving window of 1 min with a step of 250 ms, the arousal directional agreement for each triad at every particular instant was computed as the percentage of the previous min in which all triad members’ arousal was going in the same direction.

Another phenomenon targeted in the analysis was high arousal contagion (see Article III). To that extent, the duration of each high arousal interval was marked at the beginning of the interval. Then, for each interval, it was determined if, within it, some other triad member(s) reached the level of activation of high arousal (influencer point of view or potential arousal contagion), and, if so, at which latency (i.e., time elapsed from the interval start to a triad peer’s high arousal interval start). It was also determined how many triad peers could have caused it via arousal contagion (influenced point of view), if any.

**4.4.2 Performance**

The pre- and post-tests were graded from 0 to 40 by adding the sum of options chosen or left unchosen correctly (scoring 1) or incorrectly (scoring 0). The individual learning gain was calculated by subtracting pre-test from post-tests scores. The dual learning gain was aggregated as the sum of the individual learning gains.

The collaborative learning product was measured as the grade of the groups’ joint solution to the task of designing appropriate breakfast menus for people with different needs. The solution was graded from 4 to 15 by a research assistant, according to the variety of the breakfast (up to 3 points; e.g., 1 if there were <3 ingredients), accuracy (up to 3 points; e.g., 2 if there was realistic information about the macronutrients and fiber, but the solution contained small mistakes in the calculations), adequacy in terms of matching the needs of macronutrients (up to 3 points) and energy (up to 3 points), attention to all the special needs of the particular person (up to 2 points) of the different cases, and number of breakfast menus designed (1 point for more than one).

The results of both the dual learning gain and the collaborative learning product were correlated (see Article I) to five PCIs: signal matching, instantaneous derivative matching, directional agreement, Pearson’s correlation coefficient, and
Fisher’s z-transform. The purpose was to investigate the predictive power of such indices in terms of the collaborative learning outcomes.

The exam of the advanced physics course (data collection II) was graded by the teacher. In this sense, it constituted an authentic measure as it is precisely how the students’ achievement is determined in real life. In light of the Yerkes-Dodson law, this performance measure was correlated to the students’ number of sympathetic arousal episodes (indicating sympathetic activation) during the exam (see Article II).

### 4.4.3 Questionnaires

The MSLQ scales are scored by averaging the ratings of the items belonging to them (Pintrich et al., 1991). Consequently, the collaborative will, as measured by the peer learning scale, ranges from 1 to 7, similar to the individual items. The collaborative will was correlated to the five PCIs (see Article I): signal matching, instantaneous derivative matching, directional agreement, Pearson’s correlation coefficient, and Fisher’s z-transform. The purpose was to study the relation between such objective indices and the students’ subjective ratings as to their propensity to engage in collaborative behavior.

The ratings of the questionnaire adapted from Schmitz and Skinner (1993) were used directly for correlation with the in-class arousal, as measured by the count of SCRs or sympathetic arousal episodes. The questionnaire measured students’ perception of cognitive, affective, motivational, and collaborative aspects on a lesson basis. Unfortunately, in some lessons, students did not complete the questionnaire as required. From the 237 cases where valid student EDA data were available, 153 had the corresponding answered questionnaires. Hence, correlation analyses between the questionnaires and EDA data were performed on this subset of 153 cases.

### 4.5 Research evaluation

Since EDA is at the core of this research, its evaluation as a physiological response and as a signal is paramount for reliability, together with the validity of its use as a proxy for sympathetic arousal.

Research utilizing EDA in a variety of disciplines has a solid tradition spanning over a century, with the large number of studies stimulating periodic and comprehensive reviews over the decades (Boucsein, 1992, 2012; Critchley, 2002;
To date, EDA remains one of the most common data sources in the field of psychophysiology (Boucsein, 2012; Cacioppo & Tassinary, 1990a; Kreibig, 2010), that is, the area of study of psychological processes based on physiological signals. The results of the extensive research EDA has been the subject of have enabled its consideration in the scientific community as a well-validated, widely accepted measure of sympathetic arousal (Critchley, 2002). In addition, it is regarded as perhaps the most readily accessible and simplest indicator of sympathetic arousal (E. Neumann & Blanton, 1970), in connection to the exclusive characteristic of the electrodermal system as being uniquely innervated by the sympathetic nervous system (Boucsein, 2012). In the early years of EDA, there were discrepancies as to whether the EDA phenomenon was produced by vasomotor activity or sweat gland activity, but its interpretation as a measure of arousal has undisputedly accompanied EDA since its very discovery (Edelberg, 1967). Building on such a rich and long tradition, it is therefore considered valid to use EDA in this dissertation to index sympathetic arousal.

No less important is the ecological validity (see A. L. Brown, 1992) of the data collections of this dissertation for the interpretation of the results. In this regard, it is a strength of this research that the study for Articles II and III took place during a real course in an actual classroom, while the study for Article I took place in a classroom-like research environment, with a task integrated in a high school science curriculum. Not only the locations contributed to the ecological validity, but also the unobtrusive, seamless nature of the comfortable, watch-like, and scientific-quality of the wearable biosensor used. Although it has to be acknowledged that the palms and soles constitute the preferred skin sections for recording EDA (Boucsein et al., 2012), as they contain the highest density of eccrine sweat glands (Critchley, 2002; Edelberg, 1967), studies have shown that the wrist is a viable alternative to them (Poh et al., 2010; van Dooren et al., 2012), while also being more practical for daily life measurements. The EDA data collections also complied with the recommendations of recording the signal from the nondominant hand to minimize motion artifacts (Boucsein et al., 2012; Edelberg, 1967), and under a constant temperature of 23 °C to minimize thermoregulatory interference (Boucsein, 2012).

Technological advances contribute to the accessibility and reliability of physiological measures of interest in psychological research (Cacioppo & Tassinary, 1990a; Swan, 2012), including applications to learning (Schneider et al., 2015). In this context, the Empatica® E3 and E4 biosensor wristbands have been designed
purposefully for scientific research, while being practically unobtrusive (Garbarino et al., 2014). Their EDA sensor has been validated (Empatica Inc., 2017) against the Procomp® Infiniti SA9309M EDA sensor (Thought Technology Ltd., Montreal, Canada). Successful validation has been carried out also for other wristband sensors (e.g., McCarthy, Pradhan, Redpath, & Adler, 2016). However, despite the quality of the sensor, part of the data were discarded (see Sub-chapter 4.4.1) as a result of concerns regarding noisy recordings possibly due to wristband maladjustment. The analysis for the first data collection used the EDA Explorer software to detect noise within the EDA signal, with a reported accuracy of 95% (Taylor et al., 2015). In the case of the second data collection, the algorithms powering the Ledalab software used are a recent development in EDA signal processing based on mathematical and physiological modeling (Boucsein, 2012). Ledalab is the tool to analyze EDA that is currently recommended by the Empatica® E3 and E4 manufacturer (Empatica Inc., 2018). The amount of data discarded is comparable to that of other studies also using biosensors in a science high school classroom (e.g., Arroyo et al., 2009). Further technological developments are needed to systematically maximize the amount of valid data. Nevertheless, the dimensions of the second data collection involving two entire runs of a physics course still allowed us to study the phenomenon of interest at the individual and interpersonal levels in collaborative learning. The several-month data collection together with the two course runs also helped to compensate for the number of participants that was limited by the number of biosensor wristbands available.

It is known that gender plays a role in collaboration (Bear & Woolley, 2011), and that same gender structure is preferable when comparing group measures based on EDA (Boucsein et al., 2012). Unfortunately, due to the predominantly male gender of the students enrolled in the course for the second data collection, it was not possible to compose all triads with the same gender structure, having six triads with one female and two males, and two triads composed of three males. Nonetheless, clear effects of gender in psychophysiological studies have been elusive (Cacioppo & Tassinary, 1990a). However, research using electrocardiography and/or EDA measures has reported largely comparable responses in males and females, resulting in gender having only a small effect (S. A. Neumann & Waldstein, 2001), and in other cases has deliberately excluded gender from analysis after initial examination of results, due to an inconsistent effect on the variables (Simo Järvelä et al., 2014). Therefore, the reportedly small and inconsistent effect of gender mitigates this limitation.
Finally, the fact that the second data collection was conducted with a student population of high achievers must be considered when it comes to external validity or generalizability of the results to a more heterogeneous sample.

4.6 Ethics

This dissertation follows the ethical guidelines and principles of research in the humanities and social and behavioral sciences of the Finnish Advisory Board on Research Integrity (2009; 2012). Such principles involve the three areas of respecting the autonomy of research subjects, avoiding harm, and insuring privacy and data protection.

Since physiological data were to be collected on the participants while learning, an ethical statement at the project level was requested to, and the corresponding approval obtained from, the Ethics Committee of Human Sciences of the University of Oulu.

Participation in the study was voluntary, and refusing to participate had no implication at all for the students’ grades. An informed written consent was obtained from the students, who could revoke it at any time during the respective studies.

Written information was provided to the students and their parents, including the principal researcher’s contact information, the research topic, the method of collecting data and the estimated time required, the purpose for which data would be collected, how it would be archived for secondary use, and the voluntary nature of participation (Finnish Advisory Board on Research Ethics, 2009, p. 7). In addition to the written information, the principal researcher of the project, together with several project members, detailed the information orally to the students prior to the data collection and answered their questions.

Since the first data collection took place in a classroom-like laboratory outside the school, the task was coordinated with the science teachers so that it was part of their course work, meaning it did not represent extra effort for the students. In the second data collection, the measurements were taken in their regular classroom following a minimum intervention method.

The personal data collected were those strictly necessary to analyze the data and report on the research. For the data analysis phase, data were anonymized by replacing students’ names with numeric identifiers. Such anonymization also allowed for protection of privacy in any potential research publications that would
report the findings. To ensure confidentiality, the digital library data were stored and made available only to authorized research team members.

During the studies, pictures were taken to support research dissemination by helping stakeholders to better understand the design, setting, environment, material, cognitive-affective states, and so on. Due to the sensitivity of this information, their informed written consent was also requested. However, even in those cases where consent was obtained, using pictures with faces was avoided unless it was necessary to make a point about a particular relevant cognitive-affective state.
5 Overview of the articles

This dissertation involves three published scientific articles, which address the four aims pursued. In investigating psychological processes associated with learning through the proxy of a concomitant physiological signal, the articles follow multidisciplinary approaches at the intersection of the learning sciences, psychophysiology, and computer science. Articles I and III focus on the interpersonal level within triads, and Article II examines sympathetic arousal in the classroom at the individual level. The connection of the articles to the dissertation aims and to the data collections is summarized in Table 2.

Table 2. Overview of relevant aspects about the articles.

<table>
<thead>
<tr>
<th>Article</th>
<th>Aim</th>
<th>Data collection</th>
<th>Individual/Interpersonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1, 3</td>
<td>I</td>
<td>Interpersonal</td>
</tr>
<tr>
<td>II</td>
<td>2, 3</td>
<td>II</td>
<td>Individual</td>
</tr>
<tr>
<td>III</td>
<td>1, 2, 4</td>
<td>II</td>
<td>Interpersonal</td>
</tr>
</tbody>
</table>

5.1 Article I: Investigating collaborative learning success with physiological coupling indices based on electrodermal activity

This article investigates the interpersonal physiology of EDA in connection to collaborative learning processes (Aim 1) and to students’ academic achievement (Aim 3). The article is based on the first data collection of the dissertation (see Subchapter 4.3.1). Participants (N = 24) were high school students tasked with designing a healthy breakfast for an athlete training for a marathon. The task was aligned with their science curriculum and fit into their coursework. Students worked in groups of three to solve the task.

Interpersonal physiology was operationalized through PCIs involving parameters such as the amplitude (signal matching), rate of change (instantaneous derivative matching), direction (directional agreement), linear relationship (Pearson’s correlation coefficient), and weighted correlation (Fisher’s z-transform). Three collaborative learning measures were explored in relation to the PCIs (Aim 1): 1) the collaborative will, that is, the predisposition to collaborate in a learning task, measured using the peer learning scale of the MSLQ (Pintrich et al., 1993); 2) the collaborative learning product, that is, the task solution produced by the students in collaboration; and 3) the dual learning gain, that is, the aggregate of the individual learning gains as determined by the difference between pre- and post-
test scores. The collaborative learning product and the dual learning gain were used as indicators of academic achievement.

Directional agreement, the percentage of the time that the EDA of the group members were going simultaneously in the same direction (i.e., upwards, downwards, or constant), was found to be the best predictor ($r = .7$) of the dual learning gain (Aim 3). In other words, the more the team was physiologically coupled in terms of the direction of their EDA, the more they learned together. Directional agreement was also the only physiological coupling index found in the literature that is readily applicable to teams of any size, while the other indices can only be applied pairwise by definition, meaning that for teams with more than two group members, the indices have to be computed for all the possible pairs and then aggregated. The latter is not only more laborious, but also complicates operationalization and interpretation, since there are several competing approaches for aggregation. As a physiological coupling index, directional agreement emerged, thus, as being simpler and more convenient with more explanatory power than signal matching, instantaneous derivative matching, Pearson’s correlation, and the Fisher’s $z$-transform, when considering measures for teams of more than two individuals.

Instantaneous derivative matching, which accounts for the similarities in the rate of change of the EDA signal, was found to have a moderate correlation to the collaborative learning product ($r = .59$; Aim 3) and to the collaborative will ($r = .5$; Aim 1). This is to say that teams for which the rate of change of the EDA signal of their members was closer produced a better task solution, and the teams with individuals more prone to collaborate showed a more similar rate of EDA change.

### 5.2 Article II: Profiling sympathetic arousal in a physics course: How active are students?

This study builds on the notion that learning is an active process (Bjork et al., 2013; Hebb, 1955) involving cognitive and emotional processes and that arousal is a physiological component of those processes (Critchley, 2002; Poh et al., 2010). Learning activations are examined from an EDA perspective. The approach chosen is necessarily exploratory due to the lack of studies characterizing sympathetic arousal in the naturalistic setting of an actual class, as opposed to in a laboratory. Sympathetic arousal is profiled in the classroom (Aim 2)—in terms of incidence and persistence of low, medium, and high arousal levels—during two consecutive runs of an elective advanced high school physics course, which is offered twice to
the students during the spring term. Incidence was described in terms of percentage relative to the time of each lesson, which was 75 min. Persistence was operationalized as the continuous duration of a certain arousal level (i.e., without switches of level in between). Participants in the data collection (see Sub-chapter 4.3.2) were high school students ($N = 24$) who were randomly selected from the course population. In the course exam, the relation between arousal and performance was also explored (Aim 3) in the light of the well-known Yerkes-Dodson law (Yerkes & Dodson, 1908) as theoretical background.

EDA was measured unobtrusively during the course lessons and exam via the Empatica® E4 wristband to index arousal by the frequency of SCRs, which resemble peaks in the signal. Based on the peaks/min (ppm), the level of arousal was categorized into low, medium and high (Boucsein, 2012, p. 222; M. E. Dawson et al., 2017, p. 225). There were at least six changes of arousal level within the lessons, with an average of 48, representing a change every 1.5 min. High arousal was the only level that did not occur for each student in every lesson. On average, low arousal was the level of highest incidence (60% of the lesson) and longest persistence, lasting on average three times longer than medium arousal and two times longer than high arousal level occurrences.

In general, this picture of pervasive low arousal in the classroom is aligned with a similar finding from a study involving three computer-based learning environments (R. Baker et al., 2010), where boredom (i.e., a low arousal state) was found to be the most persistent state, in contrast to others in the high arousal area, such as confusion, frustration, and engaged concentration. This might be considered, at best, to be neutral and, at worst, detrimental to learning.

During the course exam, arousal was positively and highly correlated ($r = .66$) with achievement (Aim 3) as measured by the students’ grades; the more active the students in terms of sympathetic arousal, the higher the grades and vice versa. The study adds to the extant literature by providing evidence supporting the left half of the Yerkes-Dodson curve (i.e., non-detrimental arousal) during an authentic exam situation, from a sympathetic arousal perspective as indexed by EDA.

5.3 **Article III: Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members?**

This paper explored sympathetic arousal during collaborative learning in the naturalistic setting of the classroom (Aims 1 & 2) by studying, first, the
commonalities in the direction and level (i.e., low, medium, high) of arousal in a triad, and, second, the phenomenon of arousal contagion (Aim 4) inspired by the notion of emotional contagion (Hatfield et al., 1994). The study benefited from the affordances of the Empatica® E4 biosensor wristband for unobtrusive and continuous measurement of EDA to index sympathetic arousal. EDA was measured in 24 high school students working in triads during the lessons of two runs of an elective advanced physics course (see Sub-chapter 4.3.2).

Most of the time (≈60–95% of the lesson) the triad members were at different arousal levels, and, when they were at the same level, it was mainly the low arousal (or deactivated) level. The low percentage of simultaneous medium or high arousal suggests that, when the triad is doing something, the work is being led or conducted by one or two of the triad members who have been termed task-doers or the people who are leading the problem-solving at the moment (Miyake & Kirschner, 2014). Meanwhile, the other member(s) is/are less involved or engaged. Given the synchronicity and coordination presupposed in collaborative learning as a jointly produced activity (Enyedy & Stevens, 2014; Roschelle & Teasley, 1995), where joint attention is paramount (Tomasello, 1995), and given the close relation between arousal and attention (Sharot & Phelps, 2004), the results provide evidence for the claim that only certain specific phases of group work are truly collaborative (M. Baker, 2015).

Possible within-triad arousal contagion cases (Aim 4) took place mostly on a 1:1 basis (71.3%) and with a latency from within a few seconds up to 10 min, but usually within 1 min. The predominance of arousal contagion on a 1:1 basis suggests that the strongest interactions, including influence, occur mostly at a pair level, even if the team is larger, a triad in this case. In addition, the study illustrates the potential of physiological measures to enrich and strengthen research on collaborative learning by exploring a new form of data using a computational approach with high granularity to characterize the process, which is significantly less labor-intensive than traditional approaches, such as conversation and video analyses, and offers affordances which are all welcome and needed in the field (Wise & Schwarz, 2017).
6 Main findings and discussion

In connection to the dissertation aims, the main findings are presented and discussed in this chapter. Results linked to Aims 1 and 2 are presented and discussed in Sub-chapter 6.1. The relation of academic achievement to the individual and interpersonal physiology of EDA (Aim 3) is examined in Sub-chapters 6.2 (individual) and 6.3 (interpersonal). Finally, the sympathetic arousal contagion among group members as a result of the cognitive-affective social interactions in collaborative learning (Aim 4) is discussed in Sub-chapter 6.4.

6.1 Triad arousal levels in collaborative learning

In the naturalistic space for formal learning, the classroom, low arousal had the highest incidence among activation levels (i.e., low, medium, and high). On average, students spent 60% of each lesson at a low activation level, accounting for more than half of the class. These results are conspicuous in the light of, on the one hand, the interest that is reasonable to assume in the students, as the advanced physics course subject of study is elective, and on the other hand, the intricate interconnection of interest, attention, engagement, and arousal. In addition, episodes of low arousal were the most persistent, averaging three times longer than those of medium arousal and two times longer than high arousal level occurrences. Taken together, these findings provide a profile of sympathetic arousal in the classroom at the individual level (Aim 1).

This picture of pervasive low arousal in the classroom, in general, is aligned with a similar finding from a study involving three computer-based learning environments (R. Baker et al., 2010), where boredom (i.e., a low arousal state) was found to be the most persistent state, in contrast to others in the high arousal area, such as confusion, frustration, and engaged concentration. Therefore, in both digital and physical learning environments, low arousal levels seem to be the standard. Unfortunately, students at a low arousal level are considered, in general, to be disengaged (Pekrun et al., 2002), and at best, to be accumulating shallow facts in comfortable learning environments without challenges, which does not yield deep learning (D’Mello & Graesser, 2012). It has to be acknowledged that, as much as moments of relaxation facilitate disengagement, low arousal might serve the recovery purposes necessary in between intense periods of cognitive efforts and could also reinforce motivation for the next stage of learning (Pekrun et al., 2002). Consequently, low arousal plays a role in learning, but, ideally, the proportion of
low arousal should be small in comparison with states of activation (i.e., medium and high arousal), as learning is most effective when it is active (Bjork et al., 2013; Hebb, 1955).

The predominance of low arousal (i.e., low activity or inactive students) in class, even when students themselves choose to enroll in the course, which presupposes interest, might support the view that, too often, school tasks and practices are inappropriate for the intended learning goals (Scardamalia & Bereiter, 2014), thus failing to engage even interested learners. Moreover, it supports the claim that more attention should be paid and resources dedicated to tackling the problem of bored students (or inactive ones, in general), than dealing with confused or frustrated students (R. Baker et al., 2010), because the latter both exist less often and have shorter persistence on average.

The previous result (i.e., predominance of low arousal in the classroom) has an individual level scope. At the group level, this dissertation explored the commonality in the arousal levels of triad members during collaborative learning (Aim 1 and group level part of Aim 2). The triad members were at different arousal levels most of the time (≈60–95% of the lesson). When they happened to be at the same level, it was mainly the low arousal (or deactivated) level, while <4% of the lesson the triad members were simultaneously at the high arousal level. As far as the synchronicity and coordination presupposed in collaborative learning, this result is aligned with findings of systematic inequalities in participation among group members (Cohen, 1994), infrequency of true collaboration (Barron, 2003), and claims that, most likely, only specific phases of group work are collaborative (M. Baker, 2015). In the same vein, Dillenbourg (1996) proposes to refer to precise categories of interactions rather than to collaboration.

The use of interactive activities, such as communication about and coordination of activities (e.g., transactive activities; see P. A. Kirschner et al., 2018) during the learning process, is ascribed to positively affect the learning outcome (Vogel & Weinberger, 2018). However, the difference in the triad members’ arousal levels might indicate a division of labor, or that the students are taking turns in doing the collaborative work (i.e., alternating task-doers). Some spontaneous division of labor may occur in collaboration, in which there are two roles, task-doer and observer, corresponding to task execution and task monitoring, respectively (Miyake, 1986). The distribution of roles depends on the nature of the task and may change frequently (Dillenbourg et al., 1996). In any case, division of labor is a feature associated with cooperation rather than collaboration (Dillenbourg, 1999). Although cooperation and collaboration are hardly separable in practice when it
comes to group work (Jeong & Hmelo-Silver, 2016), the finding of triad members being mostly at different arousal levels points to cooperation taking the lion’s share of group work.

When given a group task, it has been shown that students tend to approach it as individual work, unless there is some reason for the group to interact (Cohen, 1994). Hertz-Lazarowitz (1989) distinguishes between simple (low) and complex (high) collaborative group tasks. According to her, the former tends to be characterized by interactions among the means and the product, while in the latter, students interact about the process to do the work together, involving discussions on planning and decision-making, which she argues lead to more high-level elaboration and, consequently, to more effective learning. Collaboration needs to be structured beyond physical proximity to others, discussing material with other students, helping other students, or sharing materials with other students (Johnson & Johnson, 2002). The importance of coordinated and cohesive participation has been repeatedly emphasized by collaborative learning research (e.g., Barron, 2000; Fransen, Weinberger, & Kirschner, 2013; Kreijns, Kirschner, & Jochems, 2003).

### 6.2 Arousal during exams and achievement

The relation between individual physiology and academic achievement is part of the third aim of this dissertation. A high, positive, and statistically significant Pearson’s correlation ($r = .66$; $p = .02$) was found between the number of sympathetic arousal episodes of the students during the course exam, as indexed by the count of SCRs, and their achievement, as measured by their exam grades. The more active the students, the higher the grades and vice versa. This result is aligned with the well-known Yerkes-Dodson law (Hebb, 1955; Yerkes & Dodson, 1908) and places our observation space on the left half of the inverted-U curve that characterizes performance as a function of arousal according to the law. The inverted-U shape describes how performance improves with arousal but only up to a point from where any further increase in arousal (e.g., excessive stress or strong anxiety) impairs performance (see Fig. 3). Because the participants were high achievers in the previous physics course, it is reasonable to presuppose that they would have high physics self-efficacy (Bandura, 1982) or confidence in their capacity to succeed in the exam. Therefore, stress and anxiety at the level that hinders performance was neither expected nor observed. The Yerkes-Dodson law has been previously tested across a variety of cognitive tasks in laboratory settings (Sanbonmatsu & Kardes, 1988). Arousal during exams has been explored in
relation to academic emotional reactions (especially test anxiety), using a variety of physiological responses, namely, those from the cardiovascular, endocrine, and immune systems, but not from the perspective of pure sympathetic arousal (Spangler, 1997). This study adds to the extant literature by providing evidence supporting the left half of the Yerkes-Dodson curve (i.e., non-detrimental arousal) during an authentic exam situation, from a sympathetic arousal perspective as indexed by EDA.

6.3 Pairwise directional agreement and dual learning gain

The second part of aim three of this dissertation concerns the relation between interpersonal physiology and academic achievement. Directional agreement measures the percentage of the time that the signals of a number of individuals are going in the same direction (i.e., upwards, downwards, or constant). During the collaborative task of designing a healthy breakfast for a marathon athlete, EDA-based directional agreement was found to have a strong correlation ($r = .7$) to the dual learning gain, the aggregation of individual learning gains, which were determined by means of pre- and post-tests. In other words, at the pairwise level, students whose sympathetic activations as indexed by their EDA signals were more attuned in terms of direction learned better together. The commonalities in EDA direction might reflect a process of joint attention, as sympathetic activation (arousal) has been shown to increase with attention (Critchley, 2002). Joint attention, in turn, might be a result of the mutual engagement and coordinated effort to solve the problem together, which defines collaborative learning (Roschelle & Teasley, 1995).

Directional agreement explained almost half ($r^2 = .49$) of the variation in the dual learning gain. Of the PCIs explored (i.e., signal matching, instantaneous derivative matching, directional agreement, linear correlation, and Fisher’s z-transform), directional agreement proved to be the best predictor of this indicator, which is in agreement with a former study showing directional agreement as the most sensitive PCI to differences in team performance (Elkins et al., 2009).

This is a useful result, considering, on one hand, that directional agreement is simpler than the other PCIs studied and, on the other hand, that it is the only one which scales up best (i.e., it can be calculated directly for groups of any size, as compared to pairwise only calculations of the other PCIs).
6.4 Within-triad high arousal contagion in collaborative learning

The fourth aim of the dissertation pursues the exploration of sympathetic arousal contagion among group members during collaborative learning. The cognitive-affective social interactions which take place in collaborative learning act as a mechanism through which emotional contagion (Hatfield et al., 1994) might occur (Thorson et al., 2018). Being that arousal is one of the two dimensions of emotion in the circumplex model of affect (Russell, 1980), it motivated the notion of arousal contagion explored in this dissertation, derived from the emotional contagion phenomenon. The focus is on high arousal contagion as the most intense activation level.

At the individual level, 635 high arousal intervals occurred, of which 263 (41%) might have caused or have been caused by arousal contagion, meaning that they took place at the time that some other triad member was experiencing a high arousal interval. In other words, a triad member in high arousal causes a teammate (or two) to reach that level of activation in 4 of 10 intervals of high arousal. It is worth noting that this ratio serves as an upper boundary, since the other triad member(s) might have come to the high arousal level on his or her own through mechanisms other than arousal contagion. Nonetheless, to assume that, most likely, triad members might be influencing one another appears reasonable, given that, although not impossible, it is difficult to isolate oneself in a collaborative learning setting (Miyake & Kirschner, 2014).

The possible arousal contagion cases took place mostly (71.3%) on a 1:1 basis, and 1:2 contagion cases accounted for about 15%, while cases where two triad members in high arousal could have brought the third member to that activation level (i.e., 2:1) through arousal contagion represented almost 14%. The fact that the majority of arousal contagion cases happened on a 1:1 basis suggests that, although there are three students working (learning in this case) together, the interaction seems to be mostly between two of the members, not necessarily always the same two, but between two rather than among the three of them. Although a priori it would seem more probable that two triad members in high arousal provoke the third to be highly active, compared to one provoking the other two to reach a high arousal state, the results showed that arousal contagion happened quite similarly on a 1:2 basis as on a 2:1 basis, the former being only 1% higher than the latter.

In terms of latency, cases of possible arousal contagion were found to occur from within a few seconds up to close to 10 min. Most of the possible high arousal contagion happened within one min, and from then on, the chances of contagion
tended to decrease with time. Provided that arousal contagion occurred, its latency seemed to be, therefore, relatively short. This might suggest that arousal contagion is most likely to happen if the high arousal cues are made available early by the triad member(s) at that activation level. Longer latency values (e.g., from 5 to 10 min) might be explained in several ways. On one hand, no visible cue might be available to indicate that another triad member is at a high arousal level (Thorson et al., 2018). On the other hand, the cue might be visible, but the appraisal could differ at first (e.g., the to be influenced triad member does not see a need to become active at that moment according to the situation or his/her own motivation to participate), and the situation might be reappraised a few min later (e.g., due to a persistent cue or a situation change; Koole, 2009).

The directionality of physiological influence (i.e., who influences whom) has been used as indicator of group leadership (Chanel & Muhl, 2015). To that extent, the triad members leading the contagion were examined. Evidence of all possible profiles was found. Thus, in some cases, there was a predominant member of the triad from whom the arousal contagion arises or two predominant members, while in other cases, it is quite evenly originated from all three members. This result aligns with the reported distinct dynamics of collaborative learning unfolding in different triads (see Enyedy & Stevens, 2014; Miyake & Kirschner, 2014). In fact, the importance of developing ways to study interaction quality over time is underscored by reports that suggest large variations in the ways that group interactions can unfold (Barron, 2000).
7 Conclusions

Collaborative skills have been largely taken for granted, but research shows that they are not an automatic consequence of group work and that training is necessary for effective collaborative groups (Cohen, 1994). Collaborative learning has been shown to be ineffective and even counterproductive if it is not properly implemented (M. Baker, 2015; P. A. Kirschner et al., 2018; Kuhn, 2015; Mullins et al., 2011). Consequently, researchers on educational psychology, teachers, and policy-makers, have long been interested in the study of collaborative learning and the devising of metrics to assist in distinguishing effective from ineffective collaboration, so that these results could be used in training. Wise and Schwarz (2017) use the term collaborative learning analytics. Few instruments in educational psychology capture joint action, and there is a need to provide accounts of interactions that capture the dynamic interplay in collaborative learning (Barron, 2000). Although the extent to which the evaluation of learning can be based only on group processes is debatable, some phenomena may emerge only in learners’ interactions, such as co-constructed knowledge or social skills (Vogel & Weinberger, 2018).

Interactions are reflected in a variety of modalities, including verbal and non-verbal (Enedy & Stevens, 2014). In a dominantly descriptive design, studies of collaborative learning have mainly involved conversation analyses and surveys, including questionnaires (e.g., open-ended, multiple-choice, and Likert scales) and interviews (Jeong et al., 2014). However, those analyses are time-consuming, very expensive microgenetic methods, and do not scale up well (Wise & Schwarz, 2017), which has led to the consideration of alternative data modalities to support the characterization of interactions. One such modality is physiological data, which has a tradition in the study of cognitive-affective processes in psychophysiological research, although it is rarely employed in educational psychology. Physiological data might assist in the instantaneous characterization of interactions according to parameters, such as commonalities and influence among peers, and could help to identify fluctuations in attention and coordination, features of interaction that are not often described in research on collaborative learning processes (Barron, 2000).

One of the most popular and widely studied physiological signals in psychophysiology is EDA, regarded as one of the simplest and most readily accessible indicators of sympathetic arousal. In turn, sympathetic arousal is an important psychological construct at the intersection of mind and body (Mandler,
connected to cognitive processes, especially attention (Raskin, 1973), and to affective processes through the circumplex model of affect (Russell, 1980).

This dissertation was conducted in the light of the gaps identified in the literature, the need for scalable ways to characterize interactions in collaborative learning, the theoretical frameworks of collaborative learning research, EDA, sympathetic arousal and interpersonal physiology, and the affordances of a wristband biosensor for an unobtrusive, continuous recording of EDA in collaborative learning settings. The overall aim was the exploration of EDA and derivative measures, such as sympathetic arousal and PCIs, concomitant with individual and interpersonal cognitive and affective processes, respectively, during collaborative learning in a naturalistic setting.

At the individual level, students spent more than half of the class at a low arousal level, possibly signaling disengagement. At the group level, triad members were most of the time at a different arousal level, and when the levels coincided, it was mainly at the low arousal level. The difference might indicate that students took turns (alternating task-doers) in executing the task, or applied some division of labor, rather than truly collaborating. In terms of achievement, two physiological indices were found to relate to learning gains and exam grades. In the first data collection, pairwise EDA-indexed directional agreement was positively and highly correlated to dual learning gain. The more the EDA signals moved in the same direction, the better students learned together. In the second data collection, arousal during the exam, as indexed by the total count of SCRs (peaks in the EDA signal), was a predictor of the students’ exam grades. The more active they were, the better their achievement. Arousal is also known to impair performance when the level is too high, for example, due to excessive stress or strong anxiety (Yerkes-Dodson law), but those levels were not reached in our sample, probably because the students were high achievers in the preceding level course of the same subject. In terms of influence from a physiological perspective, 41% of the high arousal intervals that were found could have caused or have been caused by arousal contagion, a construct derived in this dissertation from the phenomenon of emotional contagion and the classical consideration of arousal as the activation dimension of emotion. Furthermore, the possible arousal contagion cases were found to take place mostly (71.3%) on a 1:1 basis, indicating that interactions in the collaborative learning triad seem to occur mostly between two of the three members, not necessarily always the same two, but between two rather than among the three of them.

From the methodological point of view, two significant contributions were made by this dissertation. First, the scheme for classifying sympathetic arousal
levels into low, medium and high was obtained from the combination of different literature sources and results. Second, and also based on results reported in the literature, the criteria were established to disregard as noise those EDA recordings where the number of peaks in the signal (SCRs) is lower than the minimum expected theoretically.

A panoply of data sources and their corresponding methodological approaches have been used in collaborative learning research. This dissertation has delved into EDA, drawing from and building on the psychophysiological research tradition, at the individual and interpersonal levels, and applied to the study of collaborative learning. Importantly, the data were collected in ecologically valid settings, either classrooms or classroom-like environments.

This work is part of the quest to advance learning research by enriching traditional subjective data sources (e.g., questionnaires and interviews) with objective online measures of learning processes (Azevedo, 2015; Boekaerts & Corno, 2005; Winne, 2010). The current state of the art makes it imperative to focus on the physiological signal at this stage as a first integration step, but it is my hope that the studies comprising this dissertation will serve as the basis for including physiological signals as a regular practice in data collection and experiment designs in the learning sciences. As many other researchers in the field (e.g., Azevedo, 2015; Boekaerts & Corno, 2005; Sanna Järvelä, Hadwin, Malmberg, & Miller, 2018; Panadero, Klug, & Järvelä, 2015), I believe that triangulation of different data modalities, including physiological data, has the potential to provide stronger evidence and lead us as a community to draw sounder conclusions.

Several directions could be considered for future work. First, more physiological signals can be explored as long as it makes theoretical sense. Second, other indices could be developed or reused from the literature. It is important to keep it simple for interpretability and practical usefulness. Previous research reported that the simplest measures used were shown to be the most sensitive ones, which has led to the claim that physiological indices, both at individual and interpersonal levels, should be straightforward and uncomplicated (Elkins et al., 2009), which extends to other measures as well. Third, the study of the combination of indices from different signals and/or data modalities (e.g., physiological, self-reported, and observation), since the eyes of science are on multimodal approaches, should be pursued so that the limitations of individual modalities can be overcome, leading to higher validity of the conclusions drawn. Physiological data, as opposed to self-reports, is objective, but cannot be interpreted without knowledge of the context. In this sense, qualitative data are needed to further characterize and
understand the variations in PCIs and sympathetic arousal at individual and interpersonal levels, in terms of, for example, why a triad reaches full directional agreement in one lesson but not in another, and what situations cause triad members to sustain full arousal directional agreement during several min. Fourth, although here a general profile of the class and relations of the physiological measures to academic achievement were pursued, the methodology can be applied to the study of specific events of interest (e.g., directional agreement or arousal levels within a response window).

The physiological indices used in this dissertation could be fed back to the students via a learning analytics dashboard to support their reflection upon and awareness of the learning process. Both retrospective and real-time applications (e.g., alarms when certain thresholds are surpassed) are possible. It is known that individuals vary in their ability to perceive their autonomic physiological state, known as interoceptive awareness, which they need to inform cognitions and behaviors (Critchley et al., 2013). For example, people rely partly on information from their physiological state in judging their capabilities (Bandura, 1982). At the collaboration level, it has been argued that groups should become aware of their interpersonal processes and take time to discuss how they are doing as a group (Cohen, 1994). For both individual and interpersonal level purposes, the incorporation of physiological measures would enrich and provide a new dimension to learning analytics dashboards, which to date largely focus on log data (Schwendimann et al., 2017). From the teacher’s perspective, such a dashboard could assist in group composition (Ahonen et al., 2018) and identify those activities which cause more students to be active and engaged. Similarly, it could help in distinguishing those tasks leading to actual group work rather than to alternating task doers or division of labor. In sum, the findings of this study have applications to learning dashboards, prompts (e.g., in intelligent tutoring systems or online learning environments), and in general, to the development of tools for better collaboration.
List of references


Critchley, H. D., & Garfinkel, S. N. (2018). The influence of physiological signals on


87


Schwartz, M. S., Collura, T. F., Kamiya, J., & Schwartz, N. M. (2017). The history and definitions of biofeedback and applied psychophysiology. In M. S. Schwartz & F.


Original publications


Copyright of Article I held by the author. Article II reprinted with permission from John Wiley and Sons. Permission is not required from the publisher of Article III (i.e., Elsevier), as the author retains the right to include the article in a thesis or dissertation, provided it is not published commercially.

Original publications are not included in the electronic version of the dissertation.


170. Suša, Rene (2016) Social cartographies of internationalization of higher education in Canada: a study of exceptionalist tendencies and articulations


172. Rönkä, Anna Reetta (2017) Experiences of loneliness from childhood to young adulthood: study of the Northern Finland Birth Cohort 1986

173. Irjala, Marja (2017) Osallinen, syrjässä, marginaalissa, onnellinen?: tutkimus oppisolmouskoulutuksen erityisopiskelijoista Suomessa ja Saksassa


177. Pihkala, Suvi (2018) Touchable matters: reconfiguring sustainable change through participatory design, education, and everyday engagement for non-violence

178. Mertala, Pekka (2018) Two worlds collide?: mapping the third space of ICT integration in early childhood education

179. Juutinen, Jaana (2018) Inside or outside?: small stories about the politics of belonging in preschools


181. Tuomisto, Timo (2018) Kansanopistopedagogiikka kolmessa kristillisessä kansanopistossa


Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/
Héctor Javier Pijeira Díaz

ELECTRODERMAL ACTIVITY AND SYMPATHETIC AROUSAL DURING COLLABORATIVE LEARNING