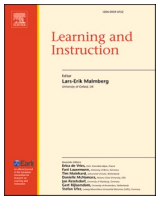
Learning and Instruction 107 (2027) 102474

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc



Emotions in flipped and lecture-based math courses

Susanna Hartikainen [a,*,1], Petri Nokelainen [a,1], Ilmari Puhakka [a], Riikka Kangaslampi [b]

[a] Faculty of Education and Culture, Tampere University, P.O.Box 700, FI-33014, Finland
[b] Faculty of Information Technology and Communication Sciences, Tampere University, FI-33014, Finland

ARTICLE INFO

Keywords:
Control-value theory
Emotion
Physiological arousal
Higher education
Flipped learning

ABSTRACT

Background: Emotions play an important role in learning. Activating teaching methods, such as flipped learning, can provide positive learning experiences but also pose emotional challenges to students.
Aims: This intensive longitudinal study investigated 1) how students' emotional valence before a learning event, learning mode (individual, peer, teacher-led), teaching implementation (lecture-based vs. flipped), and physiological arousal (electrodermal activity, EDA) during the learning event related to emotional valence after a learning event, and 2) how EDA varies across within-event segments and teaching implementations during learning events.
Sample: Participants were 69 students enrolled in an engineering mathematics course implemented either as lecture-based ($n = 34$) or flipped learning ($n = 35$).
Methods: Students reported their emotional states via a smartphone application and wore smart rings to measure EDA. A series of multilevel binomial logistic regression and linear mixed-effects models were analyzed across 546 learning events, totaling 827 h of data.
Results: Positive pre-event emotional valence predicted positive post-event emotional valence. Teacher-led events were associated with more positive post-event emotional valence than individual events, particularly in the lecture-based group. The main effect of the teaching implementation was not significant. EDA did not show robust associations with post-event emotional valence; however, lecture-based events showed a clearer decline in EDA across event segments than flipped learning events.
Conclusions: The results highlight the importance of supporting social and individual learning processes and examining physiological arousal as a temporal learning-event process rather than a simple proxy for post-event emotional valence.

1. Introduction

In current higher education (HE), the use of diverse activating and student-centered teaching methods is highlighted. This emphasis is rooted in developments in learning theories (Driessen et al., 2020) and supported by a body of empirical evidence demonstrating the positive impact of these methods on various learning outcomes (Freeman et al., 2014; Kozanitis & Nenciovici, 2023). One such activating teaching method is flipped learning, which has received growing scholarly attention in recent years, for example in engineering education (Karabulut-Ilgu et al., 2018). Talbert (2017, p. 20) defines *flipped learning* as a practice where students' first encounter with new topics moves from the group learning space to their individual learning space, and the resulting group space is turned into an interactive learning environment where students apply concepts and engage creatively with the topic.

Previous research has linked flipped designs to improved outcomes and positive emotions (Baytiyeh, 2017; Cao et al., 2025; Jeong et al., 2016; Karabulut-Ilgu et al., 2018; Lo & Hew, 2019; cf. Wagner & Urhahne, 2021). Nevertheless, from students' viewpoint, flipped designs have also induced resistance and negative emotions (Kapur et al., 2022; O'Flaherty et al., 2015), which are common when students face new teaching methods (Kilgour et al., 2016). These findings underscore the need to better understand students' emotions in flipped learning.

Emotions are receiving growing attention in educational research due to their strong links with learning, achievement, and mental and physical well-being (Pekrun, 2024). Beyond discrete emotions (Scherer, 2005),

* Corresponding author. Faculty of Education and Culture, Tampere University, PO Box 700, 33014, Tampere, Finland.
E-mail addresses: susanna.hartikainen@tuni.fi (S. Hartikainen), petri.nokelainen@tuni.fi (P. Nokelainen), ilmari.puhakka@tuni.fi (I. Puhakka), riikka.kangaslampi@tuni.fi (R. Kangaslampi).
[1] The first two authors contributed equally to this study.

https://doi.org/10.1016/j.learninstruc.2026.102474
Received 30 June 2025; Received in revised form 4 September 2026; Accepted 6 September 2026
Available online 16 September 2026
0959-4752/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

emotions can also be examined dimensionally – for instance, by valence (positive or negative) or arousal (activating/deactivating) (Pekrun & Linnenbrink-Garcia, 2014). Emotional arousal reflects the physiological component of emotions (Pekrun et al., 2023), which can be examined by categorizing perceived emotions as activating or deactivating, as well as via physiological measurement (e.g., electrodermal activity, EDA; Boucsein, 2012).

A prominent dimensional framework of emotions in education is the control-value theory (CVT) of achievement emotions, which views individual appraisals of control and value as antecedents of emotions (Pekrun et al., 2023). In CVT, learners' appraisals can be shaped by various environmental factors. For example, interaction with teachers and peers influences students' emotional experiences (Quinlan, 2016; Schutz et al., 2006). Teachers' instructional choices matter, as well. For example, activating teaching methods require greater emotional and cognitive engagement from students, which may enhance negative emotions (Wright et al., 2022).

In educational emotion research, there have been calls to complement subjective self-report methods of emotions with objective physiological data because self-reports can only access conscious processes and face risks of memory biases and tendencies to answer in socially desirable ways (e.g., Pekrun et al., 2023). Additionally, longitudinal, within-person approaches are needed to complement cross-sectional and between-person research of emotions (Pekrun, 2019). This research implements a longitudinal design and physiological measurements to examine how factors related to learning events (students' perceived emotions, different teaching implementations, learning modes, and physiological arousal) influence students' self-reported emotional valence after a learning event. In addition, because physiological arousal is a distinct component of achievement emotions, we examined EDA explicitly as a during-event arousal outcome. Thus, the present study addresses two complementary phases of the learning event. Post-event emotional valence captures students' subjective emotional evaluation after the learning event, whereas EDA captures physiological arousal while the learning event is unfolding. This distinction is important because physiological arousal does not indicate whether an experience is positive or negative; instead, it provides process-level information about the intensity and temporal dynamics of physiological activation during learning.

2. Emotional valence and arousal

Achievement emotions are related to various achievement activities and outcomes which are judged according to competence-based standards of quality (Pekrun et al., 2023; Pekrun & Perry, 2014). According to the control-value theory (CVT) of achievement emotions, emotions are seen to be shaped by their appraisals of control and value (Pekrun, 2024; Pekrun et al., 2023; Pekrun & Perry, 2014). Value consists of how important the object is for oneself and whether the event can support the attainment of one's goals. Control includes prospective perceived control over actions and outcomes, the perceived likelihood to reach the desired outcomes, as well as the perceived (internal or external) reasons for successes and failures. For example, when students perceive control regarding the outcome of studying, positive activating emotions, such as anticipatory hope, are likely to occur. In turn, uncertainty about control and focus on anticipated failure may enhance anxiety. The outcome-related emotions after a learning event are shaped by one's perceived control of success or failure of the outcome. For example, attributing success to one's own abilities or effort can trigger pride, while attributing the cause of failure on oneself enhances shame.

CVT categorizes achievement emotions based on their valence (positive/negative or pleasant/unpleasant) and arousal (activating/deactivating) in addition to object focus, which refers to emotions focusing either in the activity at hand or in the prospective or retrospective outcome (Pekrun et al., 2023). According to CVT, positive activating emotions are most beneficial to learning, while negative deactivating emotions may hinder learning by reducing cognitive resources and motivation (Camacho-Morles et al., 2021; Huang, 2011; Pekrun & Perry, 2014). Negative activating emotions, such as anxiety, may also impair learning, although their effects can be more context-dependent because they may sometimes mobilize effort. Positive deactivating emotions may be pleasant but are not necessarily as supportive of active learning engagement as positive activating emotions.

Emotional arousal can be examined by categorizing perceived emotions into activating or deactivating emotions, as well as via physiological measurements, which have been facilitated by recent technological developments (Horvers et al., 2021). One of the most used indicators of emotional psychophysical reactions is EDA (Boucsein, 2012), which reflects the activation of the sympathetic branch in the autonomic nervous system and can be measured from the secretion of eccrine sweat glands (i.e., sweating from palmar or plantar regions of body; Boucsein, 2012; Dawson et al., 2017). A unique attribute of EDA is that it reflects exclusively sympathetic nervous system reactivity and is thus associated with arousal, emotion, and attention (Dawson et al., 2017).

Research has generally indicated that activating emotions (i.e., high emotional arousal) relate to higher EDA. For instance, reviewing studies of physiological arousal in emotion, Kreibig (2010) found that low EDA was associated with emotions that do not urge for action (sadness, contentment, relief), while high EDA was connected to emotions that reflect an action tendency (e.g., anger, anxiety, joy, pride). Additionally, Roos et al. (2021) found that self-reported test anxiety, a negatively valenced activating emotion, has commonly been associated with elevated physiological arousal across studies using cardiovascular measures, EDA, cortisol sampling, or saliva pH. Furthermore, recent research indicates that positive and negative activating emotions can be associated with higher physiological arousal, whereas negative deactivating emotions can be associated with lower physiological arousal (Huber & Bannert, 2023; Ketonen et al., 2023; Stempfer et al., 2025). These studies thus support the principles of CVT and promote the role of EDA as a physiological indicator of activating and deactivating emotions (Pekrun et al., 2023).

Self-reported emotional arousal and valence, on the other hand, do not go hand in hand. Instead, Kuppens et al. (2013) showed that on a group level, both intense positive and intense negative emotions were weakly associated with emotional arousal, however, considerable variation occurred on an individual level. In addition, the relation between emotional arousal and valence might also vary across contexts and domains.

Similar asymmetry has been found also in relation to physiological arousal and valence. While higher EDA has been connected to negative emotions in some studies (e.g., Harley et al., 2019), positive high EDA situations have also been identified (e.g., successfully overcoming challenges; Muldner et al., 2010). In Kreibig's (2010) review, EDA was associated with both negative and positive emotions in studies. Consequently, as both positive and negative valence can enhance emotional and physiological arousal, the subjective perception of valence is needed to reflect how individuals experience their arousal in a certain context.

Research on situation-related achievement emotions has increased rapidly in recent years but studies have often focused on measuring trait-level rather than state-level emotions (see Camacho-Morles et al., 2021). While trait-level emotional valence is considered to be relatively stable over time in specific contexts (e.g., Pekrun & Linnenbrink-Garcia, 2014), state-level emotions have significant situational and temporal variability (Eaton & Funder, 2001; Stone et al., 1999). Nevertheless, recent research has indicated long-term stability of both aggregated state and self-reported trait emotions (Keller et al., 2025), although research in HE context is limited (Hartikainen et al., 2022).

2.1. Student emotions in learning events

According to CVT, students' perceived control and value can be enhanced, for example, by designing cognitively activating tasks, considering the motivational quality of materials and tasks by matching them with student's needs, providing students' autonomy and offering

cumulative successes (Pekrun & Perry, 2014). These aspects can be promoted, for example, with instructional designs that provide students an active role as a learner. In flipped learning, the active role of the learner is especially emphasized in the group learning space (Li et al., 2023), but can occur also in the individual learning space, especially when autonomy and responsibility of a student are emphasized when familiarizing the new content.

Previous studies have associated flipped designs with positive student emotions, such as fun and enthusiasm (e.g., Jeong et al., 2016), and enjoyment (Baytiyeh, 2017). Thus, they could aid in increasing students' control on activities and outcomes they value (Pekrun et al., 2023). However, student resistance is a problem often associated with flipped designs (Karabulut-Ilgu et al., 2018). Additionally, individual learning of the content outside of classroom and the need to take responsibility for one's learning have triggered negative emotions (O'Flaherty et al., 2015). Difficult tasks and lack of support have been connected to students' negative emotions in the contexts of various active learning activities (Chen et al., 2023; Shekhar et al., 2020). These experiences may reflect students' lack of control in these situations, which require them to take an active and responsible disposition to learn.

Importantly, the emotional meaning of individual, peer, and teacher-supported learning events may depend on the broader instructional arrangement in which they occur. In flipped learning, individual work often involves first engagement with new content and therefore requires students to regulate their own understanding before classroom activities. In lecture-based learning, individual work may more often involve practicing or consolidating content after teacher-led introduction. Similarly, peer and teacher-supported events may serve partly different functions across instructional arrangements: In flipped learning, they are typically designed to support application, discussion, and consolidation after preparatory work, whereas in lecture-based learning they may provide more direct instructional guidance and immediate clarification. Empirical findings suggest that these differences in learning activities can also have different emotional implications. Individual out-of-class learning and high responsibility for one's own learning in flipped or active-learning settings have been associated with negative emotional experiences, particularly when students perceive insufficient support (O'Flaherty et al., 2015; Shekhar et al., 2020), whereas collaborative learning can support positive emotions through shared responsibility and joint work on challenging tasks (Hartikainen et al., 2022; Volet et al., 2019). From a CVT perspective, these differences are relevant because the same type of learning event may imply different levels of perceived control, task value, autonomy, and support depending on the broader instructional arrangement. In the present study, these distinctions are captured analytically as teaching implementation (lecture-based vs. flipped learning) and learning mode (individual, peer, and teacher-led learning events), and we therefore examined both their main effects and their interaction.

Interaction with teachers and peers is an important source of students' emotions in education (Pekrun & Linnenbrink-Garcia, 2014; Volet et al., 2019). For example, enthusiastic teachers can enhance the motivational quality of teaching which can help meet students' needs and promote students' perceived value (Pekrun, 2024; Pekrun & Perry, 2014). Cooperation with peers is another important source of emotions (Volet et al., 2019). Socioemotional challenges can rise in collaboration, for example, due to troubles in pertaining to shared goals, division of labor, and group dynamics (Näykki et al., 2014). However, sharing responsibility, accomplishing difficult assignments together, and sharing emotions with peers can enhance positive emotions (Hartikainen et al., 2022). Thus, working in groups can help in seeing achievement as a group performance and help students' cope with difficult tasks, which is crucial for one to not perceive a lack of control (Pekrun, 2024; Pekrun & Perry, 2014). In addition, emotional contagion can occur in classrooms. For example, Frenzel et al. (2024) found that the cross-recurrence of university students' and teachers' facial expressions of joy was clearly above chance.

Although EDA has been the most used method to study emotions with wearable devices in education (Ba & Hu, 2023), it has not been widely applied in authentic learning settings in HE, especially when emotions are regarded. For example, as active learning methods aim to foster students' active disposition to learn (Drew & Mackie, 2011) in addition to behavioral, cognitive, and social engagement (Watkins et al., 2007), they could be assumed to cause more activation than deactivation in students. However, there is only some research comparing students' EDA in active learning versus lecture-based focused teaching. Villanueva et al. (2018) found that HE students' EDA diminished in authentic lecture-based instruction in the progress of the semester, while it increased during individual and collaborative active learning activities (see also Horvers et al., 2021). In addition, students' EDA and negative topic emotions correlated moderately in the first of the five engineering design activities. In addition, Goussain et al. (2025) examined eight students in both active learning and lecture-based sessions and found that EDA was generally higher during active learning, though individual differences were observed.

Cooperation with other students has been also connected to increases in EDA (Moya-Albiol et al., 2013; Sariñana-González et al., 2017). Hoogerheide et al. (2019) found that university students' EDA was higher when they taught to an imaginary peer student than when they were studying the material or solving a problem alone. Elfering and Grebner (2011) examined graduate students' whose EDA rose when they were giving oral presentations. Other studies, while not examining emotions per se, have connected higher EDA with learning gains (Sung et al., 2023), satisfaction (Giannakos et al., 2020), and engagement in more active forms of learning (e.g., Dubovi, 2022).

A process-oriented perspective on EDA is particularly relevant for learning-event research. EDA has been used not only to examine associations between physiological arousal and learning outcomes, but also to study how arousal varies during the learning process (Horvers et al., 2021). This distinction is central to the present study. If EDA is averaged across an entire learning event, temporal differences in physiological activation may be obscured. Prior learning-context studies suggest that physiological arousal can fluctuate across task phases and learning activities, reflecting changes in cognitive effort, attention, motivational activation, or affective involvement during the activity (Horvers et al., 2021; Pijeira-Díaz et al., 2018; Puusepp et al., 2024). Accordingly, segment-wise EDA analysis can provide information about whether physiological arousal is maintained, decreases, or increases over the course of a learning event, rather than only indicating the overall arousal level of that event.

3. The present study

This research examines students' emotions and physiological arousal in learning events from two complementary temporal perspectives. First, we examine post-event emotional valence as students' subjective affective evaluation after a learning event. Second, we examine physiological arousal during the learning event by modeling segment-wise EDA dynamics. This design allows us to distinguish between how students felt after a learning event and how their physiological arousal unfolded while the event was taking place. We set two research questions (RQs) with related hypotheses: RQ1 focuses on post-event emotional valence (H1–H5), whereas RQ2 focuses on within-event physiological arousal indexed by EDA (H6–H10).

3.1. Research question 1

For RQ1, we wanted to know *how students' emotional valence and arousal before learning event, learning mode (individual, peer, teacher-led), implementation of teaching (flipped vs. lecture-based), and physiological arousal (EDA) are related to students' emotional valence after a learning event?*

We first investigated state-level stability of students' emotional valence during learning events, as earlier considerations argue for situational and temporal variability (Eaton & Funder, 2001; Stone et al., 1999), while recent research suggests stability (Keller et al., 2025). Based on this, we expect that:

H1. Students' outcome-related emotional valence before a learning event is positively related to their emotional valence afterwards (RQ1).

Similarly, we expected that the activation level of emotions before a learning event would also be related to students' emotional valence afterwards. We expect that:

H2. Students' emotional arousal before a learning event is related to their emotional valence afterwards (RQ1).

As student collaboration in the courses was more focused on working on tasks than for example on larger open-ended projects, we expect that collaboration with peers would enhance students' control over studying and, thus, support students' positive emotions (e.g., Hartikainen et al., 2022). Additionally, we expect that teachers' guidance in teacher-led situations could enhance students' perceived control through the motivational quality of studying (Pekrun & Perry, 2014). Additionally, we expect teacher-led situations to be associated with more positive emotions than studying alone which has triggered negative emotions due to students' high responsibility and lack of support (Chen et al., 2023; O'Flaherty et al., 2015; Shekhar et al., 2020) or than studying with peers which can involve socio-emotional challenges (Näykki et al., 2014). Based on this, we expect that:

H3. Learning mode is associated with post-event emotional valence, with peer and teacher-led learning expected to be associated with more positive valence than individual learning (RQ1).

Previous research has linked positive student emotions to flipped learning designs (Baytiyeh, 2017; Jeong et al., 2016), however, improper implementation and student resistance can produce negative affect (Karabulut-Ilgu et al., 2018; O'Flaherty et al., 2015). Accordingly, appropriate support is needed to enhance emotional valence in flipped designs (Chen et al., 2023; Shekhar et al., 2020). While both flipped learning and lecture-based implementations included student-centered teaching components (e.g., videos, small group discussions), flipped learning implementation was particularly focused on structured support of conceptual understanding. Accordingly, it aimed at providing students autonomy and cumulative successes, in line with CVT (Pekrun & Perry, 2014). Thus, we expect that:

H4. Students in flipped learning group report more positive emotional valence after learning events than the students in lecture-based group (RQ1).

Because physiological arousal can accompany both positively and negatively valenced experiences (Harley et al., 2019; Kreibig, 2010; Muldner et al., 2010), we treated the EDA–post-event valence association as an empirical question. Following this, we set a following non-directional hypothesis:

H5. Students' event-level EDA is associated with emotional valence after the learning event (RQ1).

3.2. Research question 2

For RQ2, because physiological arousal is a distinct CVT component, we examined EDA explicitly as an arousal outcome during learning events by modeling within-event segment dynamics and their interaction with teaching implementation, thus we investigated *how EDA varies within learning events, and to what extent is EDA associated with teaching implementation, learning mode, and pre-event self-reported emotional valence and arousal?* (RQ2).

Based on research indicating variability in arousal/EDA during learning situations and tasks (Horvers et al., 2021; Pijeira-Díaz et al., 2018; Puusepp et al., 2024; VaezMousavi et al., 2009; see), we expect that:

H6. Students' EDA differs across within-event segments (beginning: t_0, middle: t_1, end: t_2) (RQ2).

Because various active learning approaches aim at increasing students' activation (Drew & Mackie, 2011; Watkins et al., 2007; Wright et al., 2022) and there is some evidence suggesting higher EDA during active learning compared to lecture-based sessions (Goussain et al., 2025), we expect to see differences in students' mean EDA in flipped and lecture-based groups.

H7. Flipped learning implementation is associated with higher EDA across learning events compared to lecture-based implementation (RQ2).

Previous research has found that social interaction is associated to high EDA (Elfering & Grebner, 2011; Hoogerheide et al., 2019; Moya-Albiol et al., 2013; Sariñana-González et al., 2017). Therefore, we expect that:

H8. Learning mode is associated with mean EDA, such that social learning modes (peer learning and teacher-led learning) are expected to be associated with higher EDA than individual learning (RQ2).

Based on the research indicating that both positive (Ketonen et al., 2023) and negative (Huber & Bannert, 2023) activating emotions can induce higher arousal/EDA (in addition to negative deactivating emotions being associated with lower arousal; Stempfer et al., 2025), we expect that:

H9. Students' pre-event self-reported emotional state (aggregated into valence and arousal variables) predicts EDA during the following learning event, with emotional negative or positive arousal (rather than valence) expected to show the stronger positive association with EDA (RQ2).

Previous research has indicated that lecture-based instruction might lead to a decrease in EDA in time while active learning activities might increase EDA (Villanueva et al., 2018). Therefore, we hypothesize that:

H10. Teaching implementation moderates within-event EDA dynamics such that lecture-based events show a clearer decline in EDA from t0 to t2 than flipped events (RQ2).

4. Methods

4.1. Participants

The present study sample came from 69 engineering HE students of a large Finnish research university, selected from a broader cohort of 367 first-year students enrolled in engineering mathematics courses. These students were divided into two teaching-implementation groups based on their study programs, which determined whether they studied mathematics using a lecture-based model or a flipped model. The full cohort had been grouped by study programs due to practical and pedagogical constraints. The lecture-based group primarily included students from electrical engineering, bioengineering, and information technology, while the flipped model group consisted mainly of students from automation engineering, mechanical engineering, materials science and environmental and energy engineering. Although the specializations differed, all students belonged to a man-dominated, technology-oriented HE population, and both groups included students from programs with varying entrance thresholds, indicating comparable diversity in academic preparedness. A pedagogical advantage of this grouping was that the flipped model students could continue using their existing tutoring groups from orientation studies as mathematics learning groups, supporting group cohesion.

From this larger population, 69 students participated in the present study. Their average age was 21.17 years ($SD = 2.16$), and 26 of them were women (37.7%). Lecture-based learning group consisted of 34 participants (9 women, 26.5%, $M_{age} = 21.62$ years, $SD = 3.15$) and flipped learning group had 35 participants (17 women, 48.6%, $M_{age} = 20.81$ years, $SD = 0.68$).

4.2. Context

The context of this study was a linear algebra course, which is the second mathematics course out of four consecutive, compulsory first-year engineering mathematics courses at a research-intensive university. The topics on this course included vectors, vector spaces, basic matrix

algebra, eigenvalues and eigenvectors. The course lasted seven weeks, with a final examination in the eighth week, and was worth 5 credits (based on European Credit Transfer and Accumulation System, ECTS). The students were divided between two distinct teaching implementations. One group studied using a modern lecture-based method, while others followed a teaching method based on flipped learning.

In the lecture-based model, the introduction to new topics was initiated in a teacher-led format during lectures. Following this, students worked independently or together with peers on online problems and pen-and-paper problems. They then gathered in tutorials, where solutions to some of the homework problems were discussed, and they worked in small groups on more advanced problems.

In the flipped learning model, students began exploring new topics independently through written materials, videos, and concept-building tasks. The weekly problem sets were similar to those of the students in the lecture-based model. The flipped learning group gathered with their designated, permanent small group of about eight students in tutorials, where they worked on the problems, and later in the week in a so-called prime time session. During the prime-time session, students discussed the week's topics with the instructor and engaged in advanced group tasks with their small group. Additionally, the students in the flipped learning model performed self- and peer-assessments of their understanding and homework solutions at the end of each topic. Both groups had the possibility to visit Mathematics Support Center, where they could work on homework problems alone or with peers and get support from the teaching assistants.

There were significant differences between the activities of the two student groups during both the independent study phase and the teacher-led instruction phase. The lecture-based group learned the theory through teacher-led lectures, whereas the flipped learning group studied the theory independently, although support was available. In contrast, the lecture-based group's independent work primarily involved solving homework problems. In the teacher-led sessions, the flipped learning group worked in groups, with a focus on discussion and collaborative work. In both groups, the students worked also informally and in a self-organized manner with their peers while completing various types of tasks. The forms of this self-directed group work did not differ between the groups.

Self- and peer assessment were present in both learning models but implemented differently in ways that aligned with each pedagogical approach. In the flipped learning group, written self-assessments and peer evaluations were integrated into weekly tasks via a learning management system Moodle, reflecting the model's emphasis on student responsibility. In contrast, the lecture-based group engaged in oral peer feedback during exercise sessions, and self-reflection was left to students' discretion. In the remainder of this article, we use “lecture-based model” to describe the overall lecture-based course format, whereas “teacher-led learning events” refer to specific classroom activities in which the teacher or a teaching assistant actively guided students; such teacher-led learning events could occur in both instructional models.

4.3. Data collection

Before the data collection, a research permission, including ethics approval, was obtained from the University's Research Ethics Board (decision number 2019/08/26). Participants were recruited in the beginning of the courses. They were informed about the details of the study, and the students gave an informed consent to participate voluntarily in the study. They were instructed to wear the Moodmetric smart ring (SR) for a duration of two weeks, equivalent to 10 working days. During this period, they were asked to signal via *LearningTracker* (LT; Nokelainen & Hartikainen, 2020) smartphone application the start and end times of the learning events during the day. The participants were asked to report learning events only on the linear algebra course, not on the other courses students were attending concurrently. Before and after each event, participants used LT to self-report their emotional state. This process involved ecological momentary assessment (EMA, see Stone et al., 1999) with an event-contingent design, as outlined by Bolger and Laurenceau (2013).

By using LT, participants selected one emotional state from a list of emotions (Pekrun & Perry, 2014). Emotions related to learning were asked before (“When thinking about my learning in this sequence, I now feel …”) and after (“When thinking about my learning after this sequence, I feel …”) each learning event, following the CVT framework of prospective and retrospective outcome emotions (Pekrun & Perry, 2014). Available learning-related emotions prior to the learning event were: 1. Hope (positive, activating), 2. Joy (positive, activating), 3. Relief (positive, deactivating), 4. Anxiety (negative, activating), 5. Hopelessness (negative, deactivating). Focusing on learning as an outcome in a learning situation, these are prospective outcome emotions which, in terms of CVT, are dependent on perceived control (hope, anticipatory joy, anticipatory relief) or the lack of it (anxiety, hopelessness) (Pekrun & Perry, 2014).

Learning-related emotions available after the learning event were: 1. Joy (positive, activating), 2. Pride (positive, activating), 3. Contentment (positive, deactivating), 4. Relief (positive, deactivating), 5. Shame (negative, activating), 6. Anger (negative, activating), 7. Sadness (negative, deactivating), 8. Disappointment (negative, deactivating), and 9. Gratitude (positive, activating). These emotions are retrospective outcome emotions in the CVT framework including control-independent emotions (joy, sadness), emotions dependent on matching expectations and outcome (disappointment, relief), and emotions stemming directly from success and failure of the situation (pride, shame, gratitude, anger) (Pekrun & Perry, 2014).

Although the pre- and post-lists contained different counts of positive and negative labels (pre: three positive and two negative, post: five positive and four negative), each student selected exactly one discrete emotion before and one after learning event. Emotional valence was aggregated into pre and post binary variables reflecting positive (1) versus negative (0) emotions. Because the binary code depends only on the chosen label for that event – and not on how many labels were offered – the unequal sizes of pre and post emotion lists do not bias the resulting aggregated variables.

As shown in “Emotional valence” section in Table S1 (see Supplementary Material Section 1), before learning events, 65.6% of responses were classified as positive and 34.4% as negative ($n = 546$). After learning events, the proportion of positive valence increased to 72.2%, with 27.8% classified as negative. This distribution is reflected in the binary predictor variable (preEmVal) and the binary outcome variable (postEmVal) used in the regression analyses. These proportions indicate that both pre- and post-event valence variables were moderately balanced, ensuring that the subsequent logistic regression analyses were not affected by class imbalance.

In addition to valence, each selected discrete emotion was categorized by arousal (activating vs. deactivating) based on the CVT valence × arousal taxonomy (Pekrun et al., 2023). This yielded binary self-reported arousal variables for each event: pre-event activation (preEmAr: 1 = activating, 0 = deactivating) and post-event activation (postEmAr: 1 = activating, 0 = deactivating). For pre-event emotions, hope, joy, and anxiety were coded as activating, whereas relief and hopelessness were coded as deactivating. For post-event emotions, joy, pride, shame, anger, and gratitude were coded as activating, whereas contentment, relief, sadness, and disappointment were coded as deactivating. As shown in the “Emotional arousal” section in the Supplementary Material Table S1, students' anticipatory learning-related emotions were predominantly activating before learning events (71.6%; $n = 391$). After learning events, the selected emotions were predominantly deactivating (70.5%; $n = 385$), indicating lower self-reported emotional arousal in the post-event reports. This pattern is consistent with a shift from forward-looking, action-oriented states (e.g., hope and anxiety) prior to the learning event toward more settling, outcome-focused states (e.g., relief and contentment) after the event.

The same trend is visible in the valence × arousal quadrants: before learning events, positive–activating emotions were most common (48.2%;

$n = 263$), whereas after learning events positive–deactivating emotions became the modal category (49.3%; $n = 269$), alongside an increase in negative–deactivating emotions (21.2%; $n = 116$). To test whether activation contributed beyond valence, we also created two mutually exclusive "activating quadrant" indicators for pre-event emotions - negative–activating (anxiety) and positive–activating (hope/joy) - with deactivating emotions (relief/hopelessness) as the reference category. Table S1 shows that before learning events positive–activating emotions (48.2%) were about twice as common as negative–activating emotions (23.4%), while deactivating emotions comprised 28.4% of responses (17.4% positive–deactivating; 11.0% negative–deactivating), indicating that anticipatory activation was more often positively valenced.

Participants also reported the number of persons involved in each learning event: 0 = By myself (231 events, 42.3%), 1 = With someone else (40 events, 7.3%), 2 = In a group of 3–30 persons (220 events, 40.3%), 3 = In a group of over 30 persons (55 events, 10.1%). They also reported the place for each learning event: 0 = Classroom (285 events, 52.2%), 1 = Public space at the campus (79 events, 14.5%), 2 = Public space outside campus (2 events, 0.4%), 3 = Private apartment (180 events, 33.0%). Based on these two time-varying variables, a new "learning mode" variable was constructed with three nominal categories: 0 = Individual learning in any place (213 events, 39.0%), 1 = Peer learning in any place (52 events, 9.5%), 2 = Teacher-led learning in a group of three or more students in a classroom (281 events, 51.5%). To clarify the distribution of learning modes across teaching implementations, Table 1 presents the number of learning events in each learning-mode category separately for the lecture-based and flipped learning groups. All combinations of teaching implementation and learning mode were present, including individual learning events in the flipped learning group. However, the distribution of learning modes differed between the two teaching implementations, $\chi^2(2, 546) = 14.47$, $p < .001$, Cramér's $V = 0.16$, indicating a small association between teaching implementation and learning mode.

For statistical modeling, learning mode was represented using two mutually exclusive indicator predictors with individual learning as the reference category (peer vs. individual; teacher-led vs. individual). All single-choice fields including emotions were mandatory (open field question for additional information was not mandatory) to start or stop the EMA session.

Participants provided 557,228 1-min interval EDA measurements from the SR. Descriptive statistics for the minute-level EDA data are reported in Supplementary Material Section 2, Table S2, including overall values and subgroup summaries by teaching implementation (lecture-based vs. flipped). These descriptives provide context for the subsequent segment-level EDA analyses (RQ2). The double-normalized EDA index (DNE; Salonius et al., 2024) was used in this study because it is appropriate for longitudinal, minute-level analysis of EDA and describes typical features of EDA raw signals on a scale from 1 to 100. It integrates both tonic (skin conductance level, SCL) and phasic (skin conductance response, SCR) characteristics. Accordingly, the DNE should not be interpreted as a pure tonic or phasic EDA measure, but as a composite indicator of sympathetic arousal derived from both SCL- and SCR-related information.

As described by Pakarinen et al. 2019, (pp. 2191–2195), the Moodmetric/Nuanic ring records EDA at three measurements per second (3 Hz) and extracts both tonic and phasic EDA features. The ring first normalizes the raw signal by dividing it by the slow-changing SCL, thereby emphasizing the faster SCR-related changes. The DNE then applies a second normalization at the individual level – comparing each momentary value to that person's typical EDA pattern – yielding a unitless index (1–100) that reflects sympathetic nervous-system activation (Salonius et al., 2024). Across the full measurement period, the temporal sampling rate corresponded to approximately 100.3 million raw 3-Hz measurement time points (557,228 min × 60 s × 3 measurements per second), from which the 1-min DNE values used in the analyses were produced. This approach therefore combines SCL and SCR information through sequential within-signal and within-person normalization steps (Pakarinen et al., 2019, pp. 2191–2195; Salonius et al., 2024; Torniainen et al., 2015, pp. 6433–6436), although the preprocessing and feature extraction make the index particularly sensitive to phasic, SCR-related changes. Typically, the average EDA index level is expected to be around 50. The data were screened for valid values (skin conductance level, SCL: 0.02–100 micro-Siemens; skin conductance reactions per minute, SCRn: 0–25; see Boucsein, 2012) and merged with the longitudinal self-report data from LT with *mergeSRLT* software (Nokelainen & Hartikainen, 2020). Participant-level EDA group comparisons indicated no meaningful baseline differences between instructional groups (see Supplementary Material Section 2; Table S3).

For LT, participants (lecture-based $n = 34$, flipped learning $n = 35$) provided 546 learning events of self-assessed diary data, of which 383 events occurred during arousal measurements (i.e., 29.9% of the events contained only diary data).

As the sample was man-dominating (62.3% vs. 37.7%), it was no surprise that women reported smaller n of diary events (207, 37.9%, $M = 7.96$, $SD = 3.49$) than men did (339, 62.1%, $M = 7.88$, $SD = 4.05$). Lecture-based group self-reported 303 diary events (55.5%), while the flipped learning group's number of events was 243 (44.5%). Average duration of the learning events was 90.91 min ($SD_{min} = 46.42$, $Min_{min} = 10.05$, $Max_{min} = 381.24$, $SD_{hours} = 0.77$, $Min_{hours} = 0.17$, $Max_{hours} = 6.35$). Baseline equivalence between the lecture-based and flipped groups was evaluated at the participant level. We tested whether gender composition differed between groups using a Pearson χ^2 test and quantified the magnitude of any difference using Cramér's V. This comparison was not statistically significant, $\chi^2(1, n = 69) = 3.59$, $p = .058$, $V = 0.23$. We also tested whether age differed between groups using a Welch's two-sample t-test, which likewise indicated no statistically significant difference, $t(21.46) = 1.15$, $p = .264$.

4.4. Statistical analyses

Prior to analysis, the merged data were examined for missing values using several methods (see Supplementary Material Section 3.1). In the data, each row represents a single learning event (LT entry) for one student, containing both self-reported measures collected immediately before and after the event and the corresponding aggregated physiological features (EDA) recorded during that event. For the RQ2 segment-level EDA models, learning-event rows were expanded to three segment rows per event (t_0–t_2). Missing value analysis revealed that 70.1% of the rows were complete while 29.9% had missing values only in the EDA variables; these missing values, likely due to user behavior, were not imputed.

The assumptions for the multilevel binomial logistic regression (RQ1) were confirmed by finding a linear relationship between predictors and the log odds of the outcome, nearly zero residuals indicating independent errors, low multicollinearity, and satisfactory model fit with normally distributed random effects (Supplementary Material Section 3.2). For the RQ2, linear mixed-effects models predicting segment-level EDA, we examined conditional residual diagnostics (residuals vs. fitted values and Q–Q plots) for the final segment × teaching implementation model to assess approximate normality and detect major heteroscedasticity or influential outliers (Supplementary Material Section 3.2).

Table 1
Learning events by teaching implementation and learning mode.

Teaching implementation	Individual *N* (%)	Peer *N* (%)	Teacher-led *N* (%)	Total *N* (%)
Lecture-based	100 (33.0)	25 (8.3)	178 (58.7)	303 (100.0)
Flipped learning	113 (46.5)	27 (11.1)	103 (42.4)	243 (100.0)
Total	213 (39.0)	52 (9.5)	281 (51.5)	546 (100.0)

Note. Learning mode was coded at the event level. Individual learning served as the reference category in the regression models. The distribution of learning modes differed between teaching implementations, $\chi^2(2, 546) = 14.47, p < .001$, Cramér's $V = 0.16$.

An analysis of between-student variation using the intraclass correlation coefficient (*ICC*) showed that 8.9% of the variance in post-event emotional valence (RQ1; postEmVal) was attributable to clustering at the student level, supporting multilevel modeling of learning events nested within students (Supplementary Material Section 3.3). For the segment-level EDA outcome (RQ2; EDA_seg), variance decomposition from the unconditional random-intercepts model indicated substantial clustering across students and learning events (*ICC* = 0.64; Supplementary Material Section 3.3), further justifying mixed-effects modeling.

A simulation-based multilevel power analysis was conducted for all hypotheses (see Supplementary Material Section 3.4). For RQ1 (H1–H5), power was 76.0% (H1), 98.6% (H2), 98.6% (H3), 49.4% (H4), and 97.8% (H5). For RQ2 (H6–H10), power was 100.0% (H6), 86.4% (H7), 100.0% (H8), 100.0% (H9), and 100.0% (H10). All analyses were conducted in the *R* statistical computing environment (R Core Team, 2020).

To address RQ1 on how pre-event emotional valence (H1) and arousal (H2), learning mode (H3), teaching implementation (H4), and physiological arousal (H5) were related to emotional valence after learning events, two-level correlation and multilevel binomial logistic regression analyses were conducted. The two-level bi-variate Kendall's τ_b correlations (see Table 2) were calculated using *correlation* package (Makowski et al., 2022), and the multilevel binomial logistic regression models were fitted with *lme4::glmer* function (Bates et al., 2015).

To address RQ2, we modeled electrodermal activity (EDA; double-normalized EDA index, DNE, 1–100) as a continuous physiological arousal outcome across within-event segments (t_0: start, t_1: middle, t_2: end) to further examine (a) within-event arousal dynamics (segment main effect), (b) whether these dynamics differed by teaching implementation (segment × teaching implementation), and (c) how pre-event self-reported emotional valence and arousal related to physiologically measured arousal during the learning event. Segment averages (t_0–t_2) were reshaped from wide to a segment-long format (three rows per learning event), with learning events nested within students. Segment-long models were estimated as linear mixed-effects models with random intercepts for student and learning event (event-within-student). Within-event residual correlation across segments was modeled using a compound-symmetry structure, and to account for heteroscedasticity across segments, residual variances were allowed to differ by segment. Model comparisons for the segment main effect and the segment × teaching implementation interaction were evaluated using likelihood-ratio tests (LRT) under maximum likelihood estimation; fixed-effect estimates are reported with 95% confidence intervals.

In the multilevel binomial logistic regression models related to RQ1 (see Table 3, Models 0–5), the response variable was post-event emotional valence ("postEmVal"). The outcome was moderately balanced (72.2% positive vs. 27.8% negative), so we did not apply class weighting or resampling. In the segment-long Gaussian linear mixed-effects models related to RQ2 (see Table 4, Models 0–2i), the response variable was segment-level mean EDA (double-normalized EDA index, 1–100), computed as the average EDA within each within-event segment (t_0–t_2) across 1161 segment observations. The distribution was broad ($M = 60.57$, $SD = 16.08$). These segment-level averages are higher than the minute-level descriptives in Table S2 ($M = 50.69$) and the participant-level means in Table S3 (around 49), which is expected because the segment-level values are computed only from the 34,842 min recorded during learning events, whereas Tables S2–S3 summarize all valid wear-time minutes; physiological arousal during active learning is expected to exceed overall wear-time levels. Models were built incrementally (Raudenbush & Bryk, 2002), starting with the unconditional Model 0.

Predictors were added incrementally as follows: event index for each student (Model 1), pre-event emotional valence (Model 2), learning mode represented by two indicator predictors with individual learning as the reference category (peer vs. individual; teacher-led vs. individual; Model 3), teaching implementation (lecture-based vs. flipped learning; Model 4), and event-level EDA segment averages (Model 5).

As a robustness check for Model 2, pre-event self-reported emotional arousal was tested as an additional predictor using (a) a binary activation indicator (preEmAr) and (b) activating-quadrant indicators (positive–activating; negative–activating; reference = deactivating). Because learning mode comprised three nominal categories (individual, peer, teacher-led), it was entered using two indicator predictors (peer vs. individual [reference]; teacher-led vs. individual). Following Bolger and Laurenceau (2013), each time-varying predictor was decomposed into centered-within-cluster (cwc, within-student) and centered-between-cluster (cbc, between-student) components to separate event-specific deviations from stable between-student differences. The cwc component represents the event-specific deviation from the student's own mean (within-student association) and cbc component represents the student's person-mean (between-student association, grand-mean centered).

For each learning event ($n = 546$, with varying durations), EDA was recorded at 1-min intervals and as a per-event minute-level series (variable "M_{seq}"). This minute-level M_{seq} series for each event was then split into three equal-sized segments, and the average EDA value for each segment was calculated and stored in separate variables: "Mt_0" (beginning), "Mt_1" (middle), and "Mt_2" (end). Age and gender effects were tested but not found significant.

Table 2
Two-level bivariate Kendall's τ_b correlations among time-varying variables by teaching implementation.

	1. preEmVal	2. postEmVal	3. preEmAr	4. Mt_0	5. Mt_1	6. Mt_2
Lecture–based sample ($n = 34$)						
1. preEmVal	–	0.12*	0.05	0.04	0.02	0.00
2. postEmVal	0.40***	–	0.02	0.03	0.13**	0.08
3. preEmAr	−0.13**	−0.09	–	−0.03	−0.05	−0.03
4. Mt_0	0.14**	−0.12*	0.01	–	0.35***	0.22***
5. Mt_1	0.09	−0.04	0.11*	0.72***	–	0.38***
6. Mt_2	0.10*	−0.05	0.07	0.69***	0.83***	–
Flipped learning sample ($n = 35$)						
1. preEmVal	–	0.04	0.09	−0.02	0.00	0.03
2. postEmVal	0.12*	–	0.07	−0.02	0.01	−0.05
3. preEmAr	−0.06	−0.02	–	−0.04	−0.00	0.02
4. Mt_0	−0.04	−0.07	0.02	–	0.44***	0.15**
5. Mt_1	−0.10	−0.08	−0.13*	0.73***	–	0.28***
6. Mt_2	−0.11*	−0.02	0.06	0.58***	0.59***	–

Note: Upper diagonal entries show centered-within-cluster (cwc; within-student) correlations, and lower diagonal entries show centered-between-cluster (cbc; between-student) correlations. $\tau_b > |0.06|$ = weak association, $\tau_b > |0.19|$ = moderate, $\tau_b > |0.33|$ = strong, $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$. preEmVal = Emotional valence before learning event; postEmVal = Emotional valence after learning event; preEmAr = Emotional arousal before learning event; Mt_0, Mt_1, Mt_2 = Average electrodermal activity (EDA) during the first, middle, and last 33.3% of each learning event, respectively.

Table 3
Multilevel logistic regression models predicting emotional valence after learning events.

Predictors	Model 0	Model 1	Model 2	Model 3	Model 4	Model 4i	Model 5
	B [95% CI]	*B* [95% CI]	*B* [95% CI]	*B* [95% CI]	*B* [95% CI]	*B* [95% CI]	*B* [95% CI]
Intercept	**1.05** [0.80, 1.31]	**1.10** [0.72, 1.48]	**1.12** [0.74, 1.49]	**1.18** [0.79, 1.57]	**1.17** [0.70, 1.64]	**1.22** [0.73, 1.70]	**0.97** [0.40, 1.54]
Learning event index		−0.01 [−0.07, 0.05]	−0.01 [−0.07, 0.05]	−0.02 [−0.08, 0.04]	−0.02 [−0.08, 0.04]	−0.02 [−0.08, 0.04]	0.01 [−0.07, 0.08]
Pre-event valence (cwc)			**0.74** [0.25, 1.22]	**0.75** [0.26, 1.24]	**0.75** [0.26, 1.24]	**0.82** [0.32, 1.32]	**0.70** [0.11, 1.28]
Pre-event valence (cbc)			**1.01** [0.17, 1.85]	**0.98** [0.12, 1.83]	**0.97** [0.12, 1.83]	**0.99** [0.11, 1.86]	**1.01** [0.02, 1.99]
Peer vs. individual learning (cwc)				0.56 [−0.37, 1.50]	0.56 [−0.37, 1.50]	0.93 [−0.19, 2.05]	0.87 [−0.40, 2.13]
Peer vs. individual learning (cbc)				−0.30 [−2.08, 1.48]	−0.29 [−2.09, 1.51]	−1.26 [−4.04, 1.52]	−2.02 [−5.13, 1.09]
Teacher-led vs. individual learning (cwc)				**0.49** [0.03, 0.95]	**0.49** [0.03, 0.95]	**1.13** [0.48, 1.79]	**1.08** [0.30, 1.86]
Teacher-led vs. individual learning (cbc)				0.15 [−0.93, 1.23]	0.16 [−0.94, 1.25]	0.14 [−1.22, 1.51]	0.14 [−1.43, 1.71]
Teaching implementation					0.02 [−0.47, 0.50]	−0.03 [−0.53, 0.46]	0.24 [−0.42, 0.91]
Peer vs. ind. learning (cwc) × Teaching impl.						−0.77 [−2.87, 1.34]	0.15 [−2.40, 2.70]
Peer vs. ind. learning (cbc) × Teaching impl.						1.64 [−2.11, 5.38]	3.15 [−1.05, 7.36]
Teacher-led vs. ind. learning (cwc) × Teaching impl.						**−1.31** [−2.23, −0.38]	−1.03 [−2.14, 0.08]
Teacher-led vs. ind. learning (cbc) × Teaching impl.						−0.01 [−2.33, 2.31]	1.67 [−1.34, 4.68]
Mt_0 EDA beginning segment (cwc)							−0.01 [−0.04, 0.01]
Mt_0 EDA beginning segment (cbc)							−0.01 [−0.07, 0.05]
Mt_1 EDA middle segment (cwc)							0.02 [−0.00, 0.05]
Mt_1 EDA middle segment (cbc)							0.03 [−0.04, 0.09]
Mt_2 EDA end segment (cwc)							−0.00 [−0.02, 0.02]
Mt_2 EDA end segment (cbc)							−0.01 [−0.06, 0.04]
Random Effects							
σ^2	3.29	3.29	3.29	3.29	3.29	3.29	3.29
τ_{00}	0.32	0.31	0.24	0.24	0.24	0.26	0.17
ICC	0.09	0.09	0.07	0.07	0.07	0.07	0.05
N students	69	69	69	69	69	69	62
Observations	546	546	546	546	546	546	395
Deviance	639.03	638.93	624.32	619.25	619.25	610.67	439.24
AIC	643.03	644.93	634.32	637.25	639.25	638.67	479.24

Note. *B* = Unstandardized fixed-effect estimate on the log-odds scale; 95% CI = 95% confidence interval for *B*; Intercept = Estimated log-odds of positive post-event emotional valence when predictors are set to zero or reference values; cwc = centered-within-cluster component, representing the event-specific deviation from the student's own mean; cbc = centered-between-cluster component, representing the student's person mean, grand-mean centered; Individual learning = The reference category for the learning-mode contrasts (peer and teacher-led learning); Teaching implementation = Coded as 0 = lecture-based and 1 = flipped learning; EDA = Electrodermal activity, measured using the double-normalized EDA index ranging from 1 to 100; Mt_0, Mt_1, Mt_2 = Average EDA during the first, middle, and last 33.3% of each learning event, respectively; σ^2 = Residual variance of the binomial outcome on the latent scale, fixed to $\pi^2/3$; τ_{00} = Variance of the cluster means (between-students); *ICC* = Intraclass correlation, representing the proportion of latent outcome variance attributable to between-student clustering; Observations = Number of learning events from students; Deviance = A likelihood-based index of model fit, with smaller values indicating better relative fit among comparable models; AIC = Akaike Information Criterion, used to compare relative model fit while accounting for model complexity. Bolded values indicate estimates which 95% CIs do not include zero.

5. Results

5.1. Correlation results

For the lecture-based group, within-student correlations showed that higher pre-event emotional valence was weakly associated with higher post-event valence ($\tau_b = 0.12$, $p = .014$), whereas the corresponding between-student association was moderate ($\tau_b = 0.40$, $p < .001$; see Table 2). In the flipped learning group, the within-student association between pre- and post-event valence was not statistically significant, whereas the between-student association was weak and positive ($\tau_b = 0.12$, $p = .028$). Pre-event emotional arousal showed only small associations with post-event valence in both groups.

EDA measures showed varying relationships with emotional valence. Within students, post-event emotional valence was weakly associated with average EDA during the middle part of learning events in the lecture-based group (Mt1.cwc; $\tau_b = 0.13$, $p = .005$), whereas this association was not evident in the flipped learning group. In contrast, the EDA segment measures were strongly interrelated at the between-student level in both instructional groups, indicating stable individual differences in overall EDA level across event phases. At the within-student level, the EDA segments also covaried, consistent with

Table 4
Multilevel linear mixed models predicting EDA during learning events.

Predictors	Model 0	Model 1	Model 2	Model 2i
	B [95% CI]	*B* [95% CI]	*B* [95% CI]	*B* [95% CI]
Intercept	**59.37** [56.88, 61.86]	**60.33** [57.72, 62.93]	**58.86** [55.29, 62.43]	**59.90** [56.27, 63.53]
Middle (t_1) vs. beginning (t_0) segment		−0.72 [−2.08, 0.64]	−0.72 [−2.08, 0.64]	**−2.07** [−3.87, −0.28]
End (t_2) vs. beginning (t_0) segment		**−2.05** [−3.41, −0.70]	**−2.06** [−3.41, −0.70]	**−3.80** [−5.59, −2.01]
Teaching implementation			2.93 [−1.99, 7.84]	0.49 [−4.68, 5.66]
Middle vs. beginning segment × Teaching implementation				**3.14** [0.41, 5.87]
End vs. beginning segment × Teaching implementation				**4.03** [1.31, 6.75]
Random Effects				
σ^2	93.23	92.11	92.11	91.04
τ_{00} learning event	92.89	93.39	93.39	93.63
τ_{00} student	72.36	72.18	70.19	70.22
ICC	0.64	0.64	0.64	0.64
N students	62	62	62	62
N learning events	395	395	395	395
Observations	1161	1161	1161	1161
Deviance	9186.63	9177.61	9176.25	9167.02
AIC	9194.63	9189.61	9190.25	9185.02

Note. *B* = Unstandardized fixed-effect estimate; 95% CI = 95% confidence interval for *B*; EDA = electrodermal activity, measured using the double-normalized EDA index ranging from 1 to 100; Intercept = Estimated mean EDA in the reference condition when predictors are set to zero or to their reference values; Segments = Within-event time segments, with each learning event divided into three equal parts; the beginning segment (t_0) served as the reference category, and the middle (t_1) and end (t_2) segments refer to the second and last 33.3% of each learning event, respectively; Teaching implementation = Coded as 0 = lecture-based and 1 = flipped learning; σ^2 = residual within-observation variance; τ_{00} = random-intercept variance at the learning-event and student levels; *ICC* = Intraclass correlation coefficient, representing the proportion of EDA variance attributable to clustering at the learning-event and student levels.; Observations = number of segment-level EDA observations; Deviance = A likelihood-based index of model fit, with smaller values indicating better relative fit among comparable models; AIC = Akaike Information Criterion, used to compare relative model fit while accounting for model complexity. Bolded values indicate estimates which 95% CIs do not include zero.

coherent within-event arousal dynamics. Overall, the group-specific results suggest that while EDA measures are internally consistent across phases, their association with post-event emotional valence is weak and appears mainly in the lecture-based group. For more details about the correlation results, see Supplementary Material Section 4.

5.2. RQ1: Predictors of emotional valence after learning events

A series of multilevel binomial logistic regression models were then fitted to predict post-event emotional valence. The response variable ("postEmVal") was the binary emotional valence after learning events. A total of 546 learning events (LT events) were analyzed in Models 0–4i; Model 5 was estimated on a complete-case subsample with 395 events from 62 students due to missing EDA segment data. Of the full set of 546 events, 207 (37.9%) were attended by women and 339 (62.1%) by men. By teaching implementation, 303 (55.5%) events were part of the lecture-based group, and 243 (44.5%) events were part of the flipped learning group. The models were built incrementally to evaluate the influence of different predictors, as shown in Table 3. In these models, the estimates (*B*) represent log-odds, which indicate the logarithm of the odds of experiencing positive emotional valence after a learning event. For easier interpretation, odds ratios (*OR*s) and their 95% confidence intervals (CIs) were also reported, as they provide a more intuitive understanding of the effects.

The baseline model (Model 0) included only random intercepts for students. This model established a significant intercept ($B = 1.05$, 95% CI [0.80, 1.31], $p < .001$; $OR = 2.86$, 95% CI [2.23, 3.71]), indicating a general tendency towards positive emotional valence after learning events. An *OR* of 2.86 means that the odds of students experiencing positive emotional valence after a learning event are almost three times higher than experiencing negative emotional valence. The random intercepts had a variance of $\tau_{00} = 0.32$, suggesting significant between-student variability in post-event emotional valence.

Model 1 introduced the event index (time-order/detrending variable) as a predictor. The inclusion of this variable did not significantly improve the model fit, suggesting that the time of learning events did not impact post-event emotional valence. The variance of the random intercepts remained similar to Model 0.

Adding emotional valence before learning event (Model 2) showed that both within-student centered ($B = 0.73$, 95% CI [0.25, 1.22], $p = .003$) and between-student centered ($B = 1.01$, 95% CI [0.17, 1.85], $p = .018$) measures were significant predictors, supporting H1. The *OR* of 2.08 (95% CI [1.28, 3.40]) for within-student centered valence indicates that a positive emotional valence before learning event is associated with a doubling of the odds for positive valence afterwards. Similarly, an *OR* of 2.75 (95% CI [1.18, 6.35]) for between-student centered valence suggests that students who generally feel more positive before learning events are much more likely to feel positive afterwards. The variance of the random intercepts decreased to $\tau_{00} = 0.24$, indicating that some of the between-student variability was explained by pre-event emotional valence.

To examine whether pre-event emotional arousal added explanatory value beyond valence, we extended Model 2 by adding (a) a binary activation indicator (deactivating, activating) and (b) two mutually exclusive "activating quadrant" indicators (negative–activating; positive–activating; reference = deactivating states). Neither extension improved model fit relative to Model 2 (activation indicator: $\chi^2(2) = 0.69$, $p = .71$; quadrant indicators: $\chi^2(4) = 5.53$, $p = .24$; AIC increased in both cases). Thus, H2 was not supported: pre-event valence remained the only robust pre-event emotional predictor of post-event valence in this model-building sequence. Accordingly, arousal predictors were not retained in subsequent models.

Model 3 incorporated learning mode as a predictor. As shown in Table 1, all three learning modes occurred in both teaching implementations, although teacher-led learning events were proportionally more common in the lecture-based group and individual learning events were proportionally more common in the flipped learning group.

Learning mode was entered using two mutually exclusive indicator variables with individual learning as the reference category (peer vs. individual; teacher-led vs. individual), each decomposed into within-student (cwc) and between-student (cbc) components. The within-student teacher-led (vs. individual) contrast was a positive predictor of post-event emotional valence ($B = 0.49$, 95% CI [0.03, 0.95], $p = .038$; $OR = 1.63$, 95% CI [1.03, 2.60]), indicating that teacher-led events were associated with higher odds of positive post-event emotional valence than individual events within the same student. This finding provided partial support for H3, although the uncertainty interval was relatively wide. The within-student peer contrast and both between-student learning mode components were not supported, suggesting no reliable evidence that peer learning or students' typical learning-mode profiles were associated with post-event emotional valence. The variance of the random intercepts remained relatively unchanged, suggesting that the added learning-mode predictors explained little additional between-student variability in post-event valence.

To investigate H4, Model 4 added teaching implementation (0 = lecture-based, 1 = flipped learning). It was not a significant predictor,

indicating no evidence for a main effect of teaching implementation on post-event valence. Thus, H4 was not supported. Model 4i further tested whether teaching implementation moderated the event-level (within-student; cwc) and person-level (between-student; cbc) learning-mode contrasts (peer vs. individual; teacher-led vs. individual) providing marginal overall improvement over Model 4 (LRT $p = .073$). The only supported moderation concerned the within-student teacher-led contrast: The teaching implementation × teacher-led (cwc) interaction was negative ($B = -1.31$, 95% CI [−2.23, −0.38], $p = .006$; $OR = 0.27$, 95% CI [0.11, 0.68]), indicating that the teacher-led advantage observed in the lecture-based group was substantially attenuated in the flipped group.

In the lecture-based group, teacher-led (vs. individual) events were associated with higher odds of positive post-event valence ($B = 1.13$, 95% CI [0.47, 1.79], $p < .001$; $OR = 3.11$, 95% CI [1.61, 6.01]), whereas little evidence of a corresponding teacher-led advantage was observed in the flipped learning group. Peer-related moderation was not supported. Fig. 1 illustrates this interaction: predicted probabilities increased from individual to teacher-led events in the lecture-based group, whereas they were comparatively similar and relatively high across learning modes in the flipped learning group.

Model 5 added EDA segment predictors (Mt_0, Mt_1, Mt_2; within- and between-student components) to the Model 4i structure and was estimated on a sample of 395 events (from 62 students) due to missing EDA segment data. Overall, adding EDA segment predictors did not meaningfully improve model fit relative to Model 4i, and the EDA segment effects were not robustly supported (H5). The within-student middle-segment effect (Mt_1.cwc) was slightly positive, but the estimated effect was small and the 95% confidence interval narrowly included zero, indicating no robust evidence for a mid-event EDA–valence association in this specification. Key effects from earlier models were largely replicated: pre-event emotional valence remained a positive predictor at both the within-student and between-student levels, and the within-student teacher-led contrast remained positive. However, after adding the EDA segments, the teaching implementation × teacher-led (cwc) interaction was attenuated and no longer clearly supported ($B = -1.03$, 95% CI [−2.14, 0.08]). The random-intercept variance in Model 5 was $\tau_{00} = 0.17$ ($ICC = 0.05$), indicating modest between-student heterogeneity in post-event emotional valence in this specification.

To visualize the descriptive association between physiological arousal and post-event emotional valence, we created an event-level mean EDA indicator (the average of the minute-level Mseq series within each learning event) and generated model-implied predicted probabilities (Fig. 2). The left-hand panel plots predicted probabilities at three representative arousal levels (M − 1 SD, M, M + 1 SD) and suggests a steeper positive arousal–valence gradient in the lecture-based group. In the flipped group, predicted probabilities were comparatively high already at lower EDA levels and changed less across the EDA range, suggesting a flatter descriptive arousal–valence association. The right-hand panel extends this visualization by learning mode. The most stable-looking pattern appeared for teacher-led events, where predicted probabilities increased with EDA in both groups. For individual learning events, predicted probabilities were generally higher in the flipped group across the EDA range, whereas the peer-learning panel showed the most divergent but also the least stable pattern. Because peer-learning events were comparatively rare and the EDA-related terms were not robustly supported in Model 5, Fig. 2 should be interpreted as an exploratory visualization rather than confirmatory evidence of moderated EDA effects.

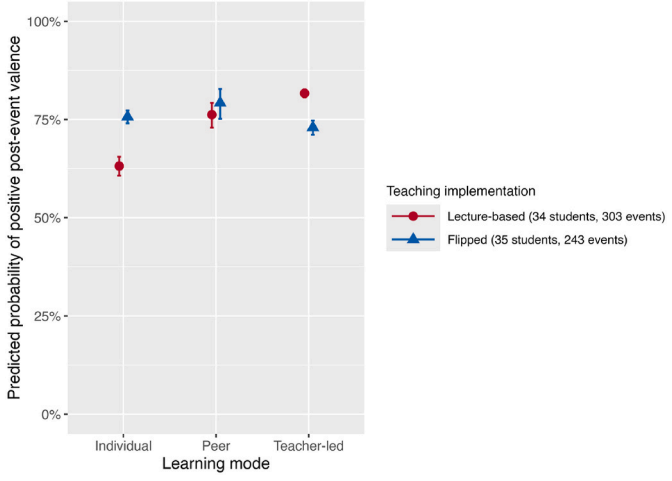


Fig. 1. Predicted probability of positive post-event emotional valence by learning mode and teaching implementation

Note: Points show model-implied marginal predictions from the multilevel logistic regression (Model 4i in Table 3) including index and pre-event valence, holding covariates at their reference/mean values; error bars denote 95% confidence intervals.

In summary, the results supported H1 and partially supported H3. H4 was not supported as a main effect, although Model 4i indicated a context-dependent moderation pattern: the teacher-led (vs. individual) within-student contrast was stronger in the lecture-based group than in the flipped learning group. Regarding H5, Model 5 did not provide clear, model-level support for EDA segment predictors overall, but the within-student mid-event EDA effect (Mt_1.cwc) was consistently positive and very close to the conventional significance threshold, suggesting a potential (albeit small) association between higher-than-usual mid-event physiological arousal and higher odds of positive post-event valence.

5.3. RQ2: Predictors of physiological arousal during learning events

A series of linear mixed-effects models were fitted to predict within-event physiological arousal indexed by EDA. The response variable ("EDA_seg") captured segment-level mean EDA within each learning event by dividing each event into three equal temporal thirds (each 33.3% of event duration) and calculating the mean EDA for each third: t_0 (beginning), t_1 (middle), and t_2 (end). The analytic dataset comprised 1161 segment observations nested within 395 learning events contributed by 62 students. All models included random intercepts for student and learning event, accounting for clustering of segment observations within events and for repeated events within students. Model building proceeded incrementally, and the statistically supported models are summarized in Table 4 (Models 0, 1, 2, and 2i). Fixed-effect estimates (B) represent mean differences in EDA units (DNE index points) associated with predictors (and their interactions), and uncertainty is reported with 95% CIs.

The baseline model (Model 0) included only random intercepts for students and learning events, representing the overall mean segment-level EDA across all students, learning events, within-event segments, and teaching implementations. The intercept was statistically significant ($B = 59.37$, 95% CI [56.88, 61.86], $p < .001$). Variance decomposition from this unconditional model indicated substantial between-student and between-event heterogeneity ($ICC = 0.64$), showing that segment-level EDA was strongly clustered within students and between learning events, beyond within-event segment-to-segment variation.

To test H6 (EDA differs across within-event segments), Model 1 added segment contrasts (t_1 vs. t_0; t_2 vs. t_0). In Model 1, the end segment (t_2) showed reliably lower EDA than the beginning segment (t_0), $B = -2.05$, 95% CI [−3.41, −0.70], $p = .003$, whereas the middle segment (t_1) did not differ from the beginning. Thus, H6 was partially supported: EDA was lower by the end of the event than at the beginning, but there was no clear evidence of a monotonic decline already by the middle third in the segment-only specification.

To test H7 (implementation is associated with mean EDA across learning events), Model 2 added teaching implementation (flipped vs. lecture-based) as a main effect on top of segment. Implementation was not supported as a uniform mean shift in EDA ($B = 2.93$, 95% CI [−1.99, 7.84], $p = .247$), indicating no reliable evidence that flipped events differed from lecture-based events in average EDA when pooling across

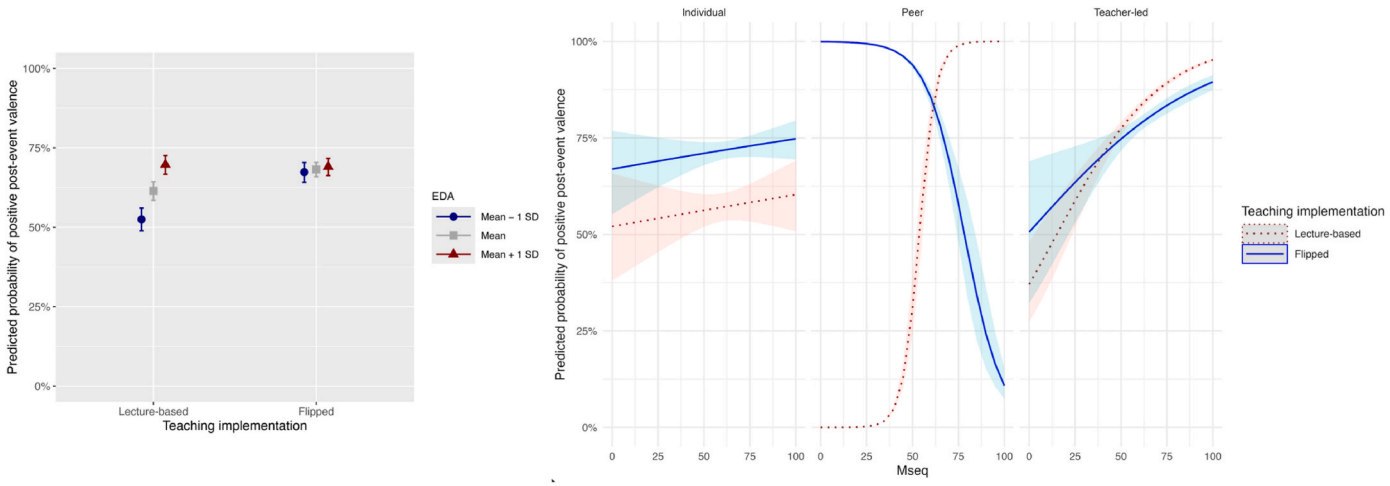


Fig. 2. Exploratory model-implied predicted probability of positive post-event emotional valence as a function of electrodermal activity (EDA) and teaching implementation

Note: The left panel shows model-implied predicted probabilities of positive post-event emotional valence at the mean and ±1 *SD* of average EDA during the learning event (M_{seq}) by teaching implementation (lecture-based, flipped learning). The right panel shows the corresponding predictions by learning mode. Points, error bars, and shaded areas indicate predicted probabilities and 95% confidence intervals.

segments. Accordingly, H7 was not supported in the main-effect sense.

The key implementation-related hypothesis was H10 (implementation moderates within-event EDA dynamics), evaluated by adding the segment × implementation interaction in Model 2i. The interaction model (Model 2i) provided a better fit than the corresponding model without the interaction (Model 2), $\chi^2(2) = 9.23$, $p = .010$, indicating that implementation-related differences were expressed primarily as differences in the within-event EDA pattern rather than as a constant offset across the event. In Model 2i, the implementation main effect remained non-significant, but both interaction terms were statistically significant: t1 (middle) × implementation ($B = 3.14$, 95% CI [0.41, 5.87], $p = .025$) and t2 (end) × implementation ($B = 4.03$, 95% CI [1.31, 6.75], $p = .004$). These positive interaction coefficients indicate that the segment-related decline observed under lecture-based instruction was attenuated in the flipped condition, yielding comparatively flatter within-event EDA dynamics under flipped implementation. Consistent with this interpretation, the lecture-based reference slope terms in Model 2i showed lower EDA in the middle (t_1) and end (t_2) segments relative to the beginning (t_0). Taken together, these results support H10: teaching implementation moderated within-event EDA change, such that lecture-based events showed a clearer decline from the beginning to later segments than flipped events.

Fig. 3 illustrates the model-implied EDA trajectories from Model 2i. The lecture-based trajectory declined from the beginning toward the end of the learning event, whereas the flipped trajectory was comparatively flatter, consistent with the segment × implementation interaction.

Beyond the segment and implementation effects reported in Table 4, additional models were estimated to address H8 (learning mode effects) and H9 (pre-event self-reported valence and activating emotions as predictors of within-event EDA), including tests of whether learning mode or pre-event self-reported emotional states moderated within-event EDA dynamics (see Supplementary Material Section 3). Learning mode was evaluated by adding peer-learning and teacher-led learning indicators, decomposed into within-student and between-student components, to Model 2i. These predictors did not improve model fit, nor did extensions including segment × learning-mode or implementation × learning-mode interactions. Thus, H8 was not supported. Pre-event emotional state predictors, including pre-event emotional valence and positive- and negative-activating emotion indicators, were tested in the same incremental manner. These extensions also did not improve fit relative to Model 2i, providing no model-level evidence that pre-event self-reported valence or activating emotions predicted within-event EDA or moderated the segment-wise EDA trajectory. Accordingly, H9 was not supported.

In summary, the results provided partial support for H6, no support for H7, support for H10, and no model-level support for H8 or H9. The clearest RQ2 finding was that teaching implementation moderated within-event EDA dynamics: lecture-based events showed a clearer decline across segments than flipped events, as reflected in the segment × implementation interaction in Table 4 and the corresponding trajectory pattern in Fig. 3. Learning mode and pre-event self-reported affect did not explain additional variance beyond the supported segment × implementation model.

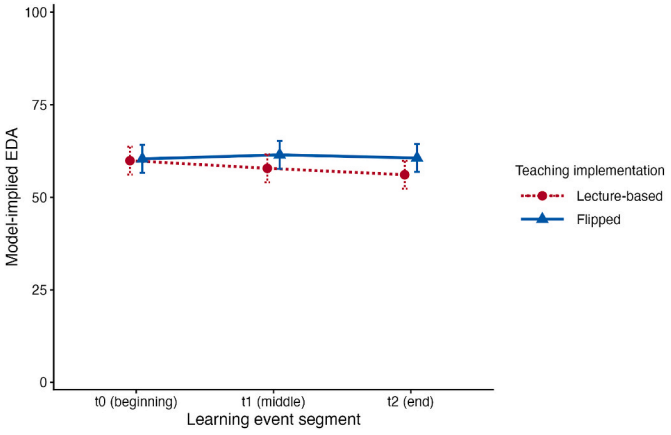


Fig. 3. Model-implied estimated marginal means of EDA across learning-event segments by teaching implementation

Note. EDA = electrodermal activity (double-normalized index range 1–100); Learning event segment = within-event time segment (t_0 = first 33.3% of the learning event duration, t_1 = middle 33.3% of the learning event, t_2 = last 33.3% of the learning event). Teaching implementation is distinguished jointly by color, line type, and marker shape: the lecture-based implementation is shown as a red dotted line with circle markers, and the flipped implementation as a blue solid line with triangle markers. Values are estimated marginal means from the final linear mixed-effects model (Table 4, Model 2i) with random intercepts for student and learning event. Error bars indicate 95% confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

6. Discussion

This section considers the two complementary outcome perspectives – post-event emotional valence (RQ1) and within-event physiological arousal dynamics indexed by EDA (RQ2) – before addressing limitations and directions for future research.

6.1. RQ1: emotional valence after learning events

Regarding RQ1, the findings supported the hypothesis that students' emotional valence before a learning event is positively related to their emotional valence afterwards (H1). Thus, although state-level emotions have situational and temporal variability (Eaton & Funder, 2001), in this setting, the emotional valence of students remained relatively stable during the learning events, as found in some other contexts, as well (Keller et al., 2025).

Second, the findings did not support the hypothesis that pre-event emotional arousal would predict post-event emotional valence beyond pre-event valence (H2). Descriptively, the associations between pre-event arousal and pre-event valence were small in magnitude: the within-student correlations were $\tau_b = 0.05$ in the lecture-based group and $\tau_b = 0.09$ in the flipped group, while the corresponding between-student correlations were $\tau_b = -0.13$ and $\tau_b = -0.06$, respectively (Table 2). Thus, the absence of an incremental pre-event arousal effect does not appear to reflect strong overlap between pre-event arousal and valence. This is consistent with previous work showing that the relation between self-reported valence and arousal is generally modest and varies considerably across individuals (Kuppens et al., 2013). In the present data, pre-event valence remained the more robust pre-event emotional predictor of post-event valence.

Third, the findings provided partial support for the hypothesis that students reported more positive emotional valence after learning events in social settings (peer or teacher-led events) than after individual learning events (H3). However, the support for H3 was not uniform across all social learning modes: in the model-level results, the clearest evidence concerned teacher-led (vs. individual) learning at the within-student level, whereas peer-learning effects were less consistent. Factors contributing to this finding might lie in the course practices. Because this was the first time the flipped format was used alongside the lecture-based model, teachers' support for students was paid extra attention to in both teaching implementations. Thus, teacher guidance could enhance students' perceived control (Pekrun & Perry, 2014) compared to studying alone or with peers, which include risks for negative emotions (Chen et al., 2023; Näykki et al., 2014; O'Flaherty et al., 2015; Shekhar et al., 2020).

Fourth, the findings showed that teaching implementation (lecture-based vs. flipped learning) was not a significant predictor of students' emotional valence, which was against our expectations (H4). This was surprising considering that positive emotions have been previously connected to flipped designs (Jeong et al., 2016). However, this was the first time the students encountered a new methodology of flipped learning which often affects the emotional experience (e.g., Kilgour et al., 2016). First-year students have also been found to have lower readiness for flipped learning methods (Hao, 2016). Additionally, the teaching staff was engaged with the methodology for the first time and simultaneously made considerable efforts for the lecture-based instruction to be student-centered and supportive of learning, as well. These factors might have influenced why the students' emotional valence did not differ between the groups.

Nevertheless, because of this result, we wanted to investigate the two instructional groups in more detail. Introducing the interaction between learning mode (individual, peer, and teacher-led learning) and teaching implementation indicated that the impact of learning mode on emotional valence differed between the lecture-based and flipped learning groups (see Fig. 1). In the lecture-based group, teacher-led learning events were associated with more positive post-event emotional valence than individual learning events. In the flipped learning group, however, students reported comparatively positive post-event emotional valence across individual, peer, and teacher-led learning events. Thus, the moderation effect should not be interpreted only as a teacher-led advantage in the lecture-based implementation, but also as suggesting that post-event emotional valence was less dependent on learning mode in the flipped implementation. Accordingly, learning-mode effects should be interpreted in relation to the broader teaching implementation in which they occurred, rather than as fully context-independent effects. One possible interpretation is that the flipped learning implementation provided a more coherent structure for individual preparation, peer collaboration, and teacher-supported consolidation, thereby making the learning process emotionally more stable across different learning modes. For example, the flipped learning design provided students autonomy but also paid attention to the cognitive quality of tasks by providing clear instructions and included a structure that offered cumulative successes (see Pekrun & Perry, 2014). These efforts can promote students' perceived competence and control also in individual learning.

However, the present data do not allow us to identify the specific interactional mechanisms underlying these learning-mode differences. For example, teacher-led events may have supported post-event emotional valence through teacher guidance, task clarification, perceived control, social support, or affective processes such as emotional contagion (Frenzel et al., 2024). Future studies should examine these mechanisms more directly using finer-grained observations of interaction quality and teacher–student or peer interaction during learning events.

Taken together, this pattern indicates that it is not only the broader instructional philosophy or methodology that matters for emotions or other learning outcomes, but also more specific features of instruction, such as the support provided during individual study phases. These more detailed differences might explain why some flipped learning experiments manage to support learning, and some do not (Kapur et al., 2022). Consequently, more precise features of instruction, not only between flipped learning and other methodologies, but also between different kinds of flipped learning practices, should be more thoroughly studied in the future. For example, the role of fixed small groups, different support strategies, types of learning materials, and assessment have not been researched in-depth to date. Thus, while H4 was not supported as a main effect of teaching implementation, the interaction analyses suggest a more nuanced, context-dependent role of teaching implementation in shaping students' emotional valence.

Finally, the findings did not provide clear support for the non-directional hypothesis that event-level EDA is associated with post-event emotional valence (H5) at the model level. Although some visualizations and exploratory patterns suggested potentially different EDA–valence relations across teaching implementations and learning modes (see Fig. 2), the multilevel models including EDA indicators did not provide robust overall evidence for a reliable EDA effect on post-event valence. This suggests that physiological arousal, while theoretically relevant to emotional experience (Harley et al., 2019; Kreibig, 2010; Muldner et al., 2010), may not map straightforwardly onto students' post-event valence in this context once pre-event valence, learning mode, and implementation-related factors are taken into account.

While the interaction effect between average EDA during learning events and teaching implementation was not significant, visualizing the results indicated certain differences between groups. In the lecture-based group, higher EDA was associated more strongly with positive valence (see left-hand side of Fig. 2). In the flipped learning group, the levels of EDA did not significantly differentiate the post-event emotional valence. This descriptive pattern may suggest that, in the flipped learning group, positive post-event valence was less contingent on physiological arousal level than in the lecture-based group. In comparison, lower EDA in the lecture-based group was associated with more negative emotions. The students in the lecture-based group might require more physiological activation to experience positive emotions. This can be challenging in case of lectures that are not often particularly arousing (see Villanueva et al., 2018).

Further analysis showed that EDA was on a higher level in the flipped learning group during individual learning events (see right-hand side of Fig. 2). This might reflect that flipped learning students were more engaged with their individual studying because it required them to be active and autonomous in studying the new content and making tasks (see e.g., Dubovi, 2022). During peer learning, the results are mixed as higher EDA relates to negative valence in the flipped learning group and positive valence in the lecture-based group. This finding should be interpreted with caution, as the number of peer learning events was smaller than the other two learning events. Finally, during the teacher-led learning events, higher EDA was associated with positive valence in both groups. Thus, when students were actively engaged in these instructional events in classrooms, positive emotions were more likely. These EDA-related patterns should be interpreted as descriptive rather than confirmatory evidence for H5.

Taken together, the RQ1 findings support H1, do not support H2, provide partial support for H3, do not support H4 as a simple main effect (while suggesting moderation by learning mode), and do not provide clear model-level support for H5. Overall, the results suggest that post-event emotional valence was most consistently related to students' pre-event emotional valence and to contextual features of the learning event, whereas physiological arousal showed a more complex and less robust relation to post-event valence in these models. These findings indicate the value to design learning environments that students perceive positively by paying attention to course structure, teacher guidance and students' perceptions of control.

6.2. RQ2: Physiological arousal dynamics during learning events

Second, this study examined within-event physiological arousal dynamics (EDA) as a distinct outcome (RQ2), focusing on segment-wise change during learning events (t_0–t_2) (H6), teaching implementation (lecture-based vs. flipped) (H7), learning mode (H8), and pre-event self-reported affective state (H9). The findings provide a more process-sensitive picture of students' arousal than the post-event valence models in RQ1. In particular, the results suggest that the most robust differences emerged in the temporal patterning of arousal within learning events (H10), rather than in a uniform mean difference between teaching implementations (H7) or in direct associations with learning mode (H8) or pre-event self-reported affective state (H9).

Regarding H6, the results provided partial support for the expectation that EDA differs across within-event segments. In the segment-only model, EDA was lower at the end of the learning event (t_2) than at the beginning (t_0), whereas the middle segment (t_1) did not differ reliably from the beginning. This suggests that within-event physiological arousal is dynamic rather than static, but that the change pattern is not simply linear across the event. This kind of pattern is consistent with prior work showing that EDA in learning settings can vary across phases of activity and task progression, rather than changing monotonically over time (e.g., Pijeira-Díaz et al., 2018; Villanueva et al., 2018).

H7 was not supported in the main-effect sense: teaching implementation (flipped vs. lecture-based) did not show a reliable uniform mean difference in EDA when averaged across segments. This "null finding" is theoretically meaningful because EDA reflects physiological arousal that may arise from multiple sources (e.g., cognitive effort, uncertainty, challenge, and social activation), and therefore it does not map directly onto a single instructional contrast in a simple way (Horvers et al., 2021; Pekrun, 2024). In other words, differences between teaching implementations may be more visible in how arousal unfolds over time than in average arousal levels.

Our results showed that H10 was supported: teaching implementation moderated within-event EDA dynamics. The lecture-based condition showed a clearer decline in EDA from the beginning toward later segments, whereas the flipped condition showed a comparatively flatter trajectory. This pattern is compatible with the idea that instructional design can shape the temporal regulation or maintenance of arousal during learning events, even when average EDA levels do not differ significantly between groups. This interpretation is broadly in line with prior findings suggesting that active-learning environments may evoke different arousal profiles than more traditional instruction, while also highlighting that phase structure and task timing are critical for understanding such differences (Goussain et al., 2025; Villanueva et al., 2018).

H8 (learning mode effects) was not supported after accounting for segment and the segment × implementation interaction. Although prior research suggests that social interaction and cooperation can be associated with higher physiological arousal (e.g., Hoogerheide et al., 2019; Sariñana-González et al., 2017), such effects are likely contingent on the nature and intensity of the interaction, task demands, and the social-evaluative context. In the present study, the broad learning-mode categories (individual, peer, teacher-led) may not have captured these finer-grained arousal-relevant features once temporal event structure and implementation-specific dynamics were modeled.

H9 was also not supported: There was no model-level evidence that pre-event self-reported emotional valence or arousal predicted within-event EDA, nor that these variables moderated segment-wise EDA change or implementation-related differences. This does not imply that self-reported emotions and physiological arousal are unrelated. Rather, it is consistent with evidence that relations between self-reports and EDA in learning contexts are often modest, heterogeneous, and dependent on emotion type, timing, and contextual features (Horvers et al., 2021; Ketonen et al., 2023). It is also conceptually consistent with work showing that arousal-related physiological signals may accompany both positively and negatively valenced emotional experiences, including activating negative emotions and lower-arousal deactivating states such as boredom (Huber & Bannert, 2023; Stempfer et al., 2025).

Taken together, the RQ2 findings extend the RQ1 results by showing that teaching implementation appears to matter more for the shape of physiological arousal across learning-event segments than for a simple mean-level difference in arousal. Importantly, the absence of robust EDA effects in the post-event valence model does not imply that EDA is uninformative. Rather, it suggests that EDA may be more informative as a process-level indicator of within-event arousal dynamics than as a direct proxy for post-event emotional valence. More broadly, the results support a process-oriented perspective on learning-related emotions and arousal: when EDA is examined at a finer temporal resolution, instructional differences may emerge as differences in trajectory rather than static intensity (Pekrun, 2024; Villanueva et al., 2018).

6.3. Limitations

Certain limitations should be considered when interpreting the findings of this study. *First*, the study was conducted at a single Finnish university, and the specific context of engineering HE might not be representative of other disciplines or educational settings. However, it is worth noting that natural science education - and particularly engineering mathematics - is relatively standardized across countries. The language of mathematics is inherently international, and the structure and content of engineering mathematics instruction tend to follow globally consistent norms. Therefore, the generalizability of the findings may be stronger than in more context-dependent educational domains.

Second, the teaching implementation was linked to degree programme membership for practical scheduling and pedagogical reasons. Consequently, differences between degree programmes, student cohorts, prior educational experiences, or programme-specific learning cultures may be confounded with teaching implementation. Therefore, findings concerning the flipped versus lecture-based comparison should not be interpreted as causal treatment effects. Rather, they describe associations observed in two naturally occurring implementations of the same engineering mathematics course.

Third, while the sample size of 69 students may appear limited for quantitative research, the resulting intensive longitudinal in-situ

physiological (34,842 min of EDA during learning events) and EMA data (546 learning events) provide a robust dataset for analysis. Simulation-based power analyses indicated sufficient power for most focal effects, including H2, H3, and H5 in RQ1, and all RQ2 hypotheses (H6–H10) at the selected benchmark effect sizes. Power was marginal for H1 (76.0%) and relatively low for the teaching implementation main effect in H4 (49.4%), suggesting that null or weak findings related to the overall teaching implementation main effect should be interpreted with caution.

Regarding the low power of H4, this likely reflects the fact that H4 targets a time-invariant between-student teaching implementation effect in a multilevel model, whereas much of the study's information comes from repeated within-student observations. In practice, repeated measurements increase precision mainly for within-student effects, but they do not increase the effective sample size for a between-student contrast to the same extent. Thus, for H4, power is driven more by the number of students (clusters) and treatment-group comparability than by the total number of observations. Accordingly, H4 should be treated as a more uncertain test than the other hypotheses, and future studies would benefit from increasing the number of students (especially at the treatment-group comparison level), improving balance between groups, and/or using designs that strengthen identification of treatment effects.

Fourth, the reliance on self-reported data for emotional states introduces potential biases such as social desirability and recall bias. Although EMA techniques were used to mitigate these issues, self-reported measures are inherently subjective and may not fully capture the complexity of students' emotional experiences. However, physiological measurement of emotional arousal was used to complement the self-report data.

Fifth, this study investigated students' pre and post outcome-related emotions, while students' activity-related emotions during the learning event were not explored. Thus, future studies should focus on studying these perspectives also in relation to students' concurrent or post-event emotions during learning.

7. Conclusions

This study investigated HE students' emotions during first-year engineering mathematics studies within the framework of CVT in courses implemented using either flipped learning or lecture-based pedagogy. The results showed that positive emotional valence before a learning event predicted more positive post-event emotional valence, and that learning events involving teacher-led instruction (relative to individual learning) were associated with more positive post-event emotional valence, particularly in the lecture-based group. In contrast, teaching implementation (lecture-based vs. flipped) did not show a clear main effect on post-event emotional valence, but the results suggested that implementation shaped how learning-mode differences were reflected in students' emotional experiences.

Regarding physiological arousal, the results did not provide clear model-level evidence for a robust direct association between event-level EDA and post-event emotional valence once other predictors were taken into account. However, when EDA was examined as a within-event outcome, the key difference between teaching implementations was not a uniform mean-level difference but a difference in segment-wise arousal trajectory: lecture-based events showed a clearer decline in EDA (i.e., physiological arousal) across event segments than flipped learning events. This highlights the value of examining arousal as a temporal process rather than only as an event-average indicator.

Overall, the study extends research on learning-related emotions in HE by combining EMA-based self-reports and wearable EDA data in authentic instructional settings. The findings suggest that students' emotional experiences are shaped by both pre-event emotional states and contextual features of learning events, and that teaching implementation may be especially visible in the temporal dynamics of physiological arousal during learning.

CRediT authorship contribution statement

Susanna Hartikainen: Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Petri Nokelainen:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Ilmari Puhakka:** Writing – review & editing, Writing – original draft, Investigation. **Riikka Kangaslampi:** Writing – review & editing, Writing – original draft, Conceptualization.

Declarations of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The supporting data for this study are not publicly available due to their sensitive nature and because participants did not provide written consent for their data to be shared publicly.

Acknowledgements

We thank Terhi Kaarakka, Elina Viro, Jani Hirvonen, and Simo Ali-Löytty for developing the teaching method based on flipped learning that was used in the intervention group.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2026.102474.

References

Ba, S., & Hu, X. (2023). Measuring emotions in education using wearable devices: A systematic review. *Computers & Education, 200*, Article 104797. https://doi.org/10.1016/j.compedu.2023.104797

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baytiyeh, H. (2017). The flipped classroom model: When technology enhances professional skills. *The International Journal of Information and Learning Technology, 34*(1), 51–62. https://doi.org/10.1108/IJILT-07-2016-0025

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* Guilford Press.

Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). Springer.

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review, 33*(3), 1051–1095. https://doi.org/10.1007/s10648-020-09585-3

Cao, H., Wu, Y., Guo, M., Tang, C., Yang, K., Jiang, Q., & Yu, M. (2025). The application of the flipped classroom teaching method in periodontal medical education. *Learning and Instruction, 98*, Article 102148. https://doi.org/10.1016/j.learninstruc.2025.102148

Chen, X., Usher, E. L., Roeder, M., Johnson, A. R., Kennedy, M. S., & Mamaril, N. A. (2023). Mastery, models, messengers, and mixed emotions: Examining the development of engineering self-efficacy by gender. *Journal of Engineering Education, 112*(1), 64–89. https://doi.org/10.1002/jee.20494

Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (4th ed., pp. 217–243). Cambridge University Press.

Drew, V., & Mackie, L. (2011). Extending the constructs of active learning: Implications for teachers' pedagogy and practice. *Curriculum Journal, 22*(4), 451–467. https://doi.org/10.1080/09585176.2011.627204

Driessen, E. P., Knight, J. K., Smith, M. K., & Ballen, C. J. (2020). Demystifying the meaning of active learning in postsecondary biology education. *CBE-Life Sciences Education, 19*(4), Article ar52. https://doi.org/10.1187/cbe.20-04-0068

Dubovi, I. (2022). Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology. *Computers & Education, 183*, Article 104495. https://doi.org/10.1016/j.compedu.2022.104495

Eaton, L. G., & Funder, D. C. (2001). Emotional experience in daily life: Valence, variability, and rate of change. *Emotion, 1*(4), 413–421. https://doi.org/10.1037/1528-3542.1.4.413

Elfering, A., & Grebner, S. (2011). Ambulatory assessment of skin conductivity during first thesis presentation: Lower self-confidence predicts prolonged stress response. *Applied Psychophysiology and Biofeedback, 36*(2), 93–99. https://doi.org/10.1007/s10484-011-9152-3

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111* (23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Frenzel, A. C., Dindar, M., Pekrun, R., Reck, C., & Marx, A. K. (2024). Joy is reciprocally transmitted between teachers and students: Evidence on facial mimicry in the classroom. *Learning and Instruction, 91*, Article 101896. https://doi.org/10.1016/j.learninstruc.2024.101896

Giannakos, M. N., Sharma, K., Papavlasopoulou, S., Pappas, I. O., & Kostakos, V. (2020). Fitbit for learning: Towards capturing the learning experience using wearable sensing. *International Journal of Human-Computer Studies, 136*, Article 102384. https://doi.org/10.1016/j.ijhcs.2019.102384

Goussain, B. G. C. D. S., de Moura, R. A., Luche, J. R. D., Andrade, H. D. S., Gomes, F. M., & Silva, M. B. (2025). Electrodermal activity as an indicator of student engagement: A comparative study of traditional and active learning environments. *Proceedings of the 23rd LACCEI international multi-conference for engineering, education and technology (LACCEI 2025)*. https://laccei.org/LACCEI2025-Mexico/meta/FP2423.html.

Hao, Y. (2016). Exploring undergraduates' perspectives and flipped learning readiness in their flipped classrooms. *Computers in Human Behavior, 59*, 82–92. https://doi.org/10.1016/j.chb.2016.01.032

Harley, J. M., Jarrell, A., & Lajoie, S. P. (2019). Emotion regulation tendencies, achievement emotions, and physiological arousal in a medical diagnostic reasoning simulation. *Instructional Science, 47*(2), 151–180. https://doi.org/10.1007/s11251-018-09480-z

Hartikainen, S., Pylväs, L., & Nokelainen, P. (2022). Engineering students' perceptions of teaching: Teacher-created atmosphere and teaching procedures as triggers of student emotions. *European Journal of Engineering Education, 47*(5), 814–832. https://doi.org/10.1080/03043797.2022.2034750

Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & van Gog, T. (2019). Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology, 111*(1), 45–56. https://doi.org/10.1037/edu0000272

Horvers, A., Tombeng, N., Bosse, T., Lazonder, A. W., & Molenaar, I. (2021). Detecting emotions through electrodermal activity in learning contexts: A systematic review. *Sensors, 21*(23), Article 7869. https://doi.org/10.3390/s21237869

Huang, C. (2011). Achievement goals and achievement emotions: A meta-analysis. *Educational Psychology Review, 23*, 359–388. https://doi.org/10.1007/s10648-011-9155-x

Huber, K., & Bannert, M. (2023). What happens to your body during learning with computer-based environments? Exploring negative academic emotions using psychophysiological measurements. *Journal of Computers in Education, 10*(1), 189–215. https://doi.org/10.1007/s40692-022-00228-w

Jeong, J. S., González-Gómez, D., & Cañada-Cañada, F. (2016). Students' perceptions and emotions toward learning in a flipped general science classroom. *Journal of Science Education and Technology, 25*(5), 747–758. https://doi.org/10.1007/s10956-016-9630-8

Kapur, M., Hattie, J., Grossman, I., & Sinha, T. (2022). Fail, flip, fix, and feed – Rethinking flipped learning: A review of meta-analyses and a subsequent meta-analysis. *Frontiers in Education, 7*, Article 956416. https://doi.org/10.3389/feduc.2022.956416

Karabulut-Ilgu, A., Jaramillo Cherrez, N., & Jahren, C. T. (2018). A systematic review of research on the flipped learning method in engineering education. *British Journal of Educational Technology, 49*(3), 398–411. https://doi.org/10.1111/bjet.12548

Keller, M. M., Yanagida, T., Lüdtke, O., & Goetz, T. (2025). How similar are students' aggregated state emotions to their self-reported trait emotions? Results from a measurement burst design across three school years. *Educational Psychology Review, 37*(1), Article 26. https://doi.org/10.1007/s10648-025-09995-1

Ketonen, E. E., Salonen, V., Lonka, K., & Salmela-Aro, K. (2023). Can you feel the excitement? Physiological correlates of students' self-reported emotions. *British Journal of Educational Psychology, 93*, 113–129. https://doi.org/10.1111/bjep.12534

Kilgour, J. M., Grundy, L., & Monrouxe, L. V. (2016). A rapid review of the factors affecting healthcare students' satisfaction with small-group, active learning methods. *Teaching and Learning in Medicine, 28*(1), 15–25. https://doi.org/10.1080/10401334.2015.1107484

Kozanitis, A., & Nenciovici, L. (2023). Effect of active learning versus traditional lecturing on the learning achievement of college students in humanities and social sciences: A meta-analysis. *Higher Education, 86*(6), 1377–1394. https://doi.org/10.1007/s10734-022-00977-8

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin, 139*(4), 917–940. https://doi.org/10.1037/a0030811

Li, R., Lund, A., & Nordsteien, A. (2023). The link between flipped and active learning: A scoping review. *Teaching in Higher Education, 28*(8), 1993–2027. https://doi.org/10.1080/13562517.2021.1943655

Lo, C. K., & Hew, K. F. (2019). The impact of flipped classrooms on student achievement in engineering education: A meta-analysis of 10 years of research. *Journal of Engineering Education, 108*(4), 523–546. https://doi.org/10.1002/jee.20293

Makowski, D., Wiernik, B. M., Patil, I., Lüdecke, D., & Ben-Shachar, M. S. (2022). *Correlation: Methods for correlation analysis* (0.8.3) [R package] https://CRAN.R-project.org/package=correlation.

Moya-Albiol, L., De Andrés-García, S., Sanchis-Calatayud, M. V., Sariñana-González, P., Ruiz-Robledillo, N., Romero-Martínez, Á., & González-Bono, E. (2013). Psychophysiological responses to cooperation: The role of outcome and gender. *International Journal of Psychology, 48*(4), 542–550. https://doi.org/10.1080/00207594.2012.666552

Muldner, K., Burleson, W., & VanLehn, K. (2010). "Yes!": Using tutor and sensor data to predict moments of delight during instructional activities. In P. De Bra, A. Kobsa, & D. Chin (Eds.), *Lecture notes in computer science, 6075* pp. 159–170). Springer. https://doi.org/10.1007/978-3-642-13470-8_16. *User modeling, adaptation, and personalization*.

Näykki, P., Järvelä, S., Kirschner, P. A., & Järvenoja, H. (2014). Socio-emotional conflict in collaborative learning—A process-oriented case study in a higher education context. *International Journal of Educational Research, 68*, 1–14. https://doi.org/10.1016/j.ijer.2014.07.001

Nokelainen, P., & Hartikainen, S. (2020). LearningTracker [Mobile app]. App Store Version 2.0.1. https://apps.apple.com/us/app/learning-tracker/id1221361135.

O'Flaherty, J., Phillips, C., Karanicolas, S., Snelling, C., & Winning, T. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education, 25*, 85–95. https://doi.org/10.1016/j.iheduc.2015.02.002

Pakarinen, T., Pietilä, J., & Nieminen, H. (2019). Prediction of self-perceived stress and arousal based on electrodermal activity. *Proceedings of the 41st annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. https://doi.org/10.1109/EMBC.2019.8857621

Pekrun, R. (2019). Inquiry on emotions in higher education: Progress and open problems. *Studies in Higher Education, 44*(10), 1806–1811. https://doi.org/10.1080/03075079.2019.1665335

Pekrun, R. (2024). Control-Value theory: From achievement emotion to a general theory of human emotions. *Educational Psychology Review, 36*(3), Article 83. https://doi.org/10.1007/s10648-024-09909-7

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 1–10). Taylor and Francis. https://doi.org/10.4324/9780203148211.

Pekrun, R., Marsh, H. W., Elliot, A. J., Stockinger, K., Perry, R. P., Vogl, E., Goetz, T., van Tilburg, W. A. P., Lüdtke, O., & Vispoel, W. P. (2023). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology, 124*(1), 145–178. https://doi.org/10.1037/pspp0000448

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Taylor and Francis.

Pijeira-Díaz, H. J., Drachsler, H., Kirschner, P. A., & Järvelä, S. (2018). Profiling sympathetic arousal in a physics course: How active are students? *Journal of Computer Assisted Learning, 34*(4), 397–408. https://doi.org/10.1111/jcal.12271

Puusepp, I., Tammi, T., Linnavalli, T., Huotilainen, M., Laine, S., Kuusisto, E., & Tirri, K. (2024). Changes in physiological arousal during an arithmetic task: Profiles of elementary school students and their associations with mindset, task performance and math grade. *Scientific Reports, 14*(1), Article 1606. https://doi.org/10.1038/s41598-024-51683-7

Quinlan, K. M. (2016). How emotion matters in four key relationships in teaching and learning in higher education. *College Teaching, 64*(3), 101–111. https://doi.org/10.1080/87567555.2015.1088818

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models. Applications and data analysis methods* (2nd ed.). Sage.

Roos, A.-L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2021). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review, 33*(2), 579–618. https://doi.org/10.1007/s10648-020-09543-z

Salonius, H., Kytölä, P., Pakarinen, S., Mäkikangas, A., & Saija, M. (2024). Associations of continuous electrodermal activity with job burnout and work engagement. *SSRN*. https://doi.org/10.2139/ssrn.4698943

Sariñana-González, P., Romero-Martínez, Á., & Moya-Albiol, L. (2017). Cooperation induces an increase in emotional response, as measured by electrodermal activity and mood. *Current Psychology, 36*(2), 366–375. https://doi.org/10.1007/s12144-016-9426-2

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Schutz, P. A., Hong, J. Y., Cross, D. I., & Osbon, J. N. (2006). Reflections on investigating emotion in educational activity settings. *Educational Psychology Review, 18*(4), 343–360. https://doi.org/10.1007/s10648-006-9030-3

Shekhar, P., Borrego, M., DeMonbrun, M., Finelli, C., Crockett, C., & Nguyen, K. (2020). Negative student response to active learning in STEM classrooms: A systematic review of underlying reasons. *Journal of College Science Teaching, 49*(6), 45–54. https://doi.org/10.1080/0047231X.2020.12290664

Stempfer, L., Stoll, S. E., Fries, J., Pekrun, R., & Goetz, T. (2025). A systematic review and multilevel meta-analysis of the relationship between boredom and arousal. *Communications Psychology, 3*, Article 182. https://doi.org/10.1038/s44271-025-00358-8

Stone, A. A., Shiffman, S. S., & DeVries, M. W. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of Hedonic psychology* (pp. 26–39). Russell Sage Foundation.

Sung, G., Bhinder, H., Feng, T., & Schneider, B. (2023). Stressed or engaged? Addressing the mixed significance of physiological activity during constructivist learning. *Computers & Education, 199*, Article 104784. https://doi.org/10.1016/j.compedu.2023.104784

Talbert, R. (2017). *Flipped learning: A guide for higher education faculty*. Routledge. https://doi.org/10.4324/9781003444848

Torniainen, J., Cowley, B., Henelius, A., Lukander, K., & Pakarinen, S. (2015). Feasibility of an electrodermal activity ring prototype as a research tool. *Proceedings of the 37th*

annual international conference of the IEEE Engineering in Medicine and Biology Society. EMBC. https://doi.org/10.1109/EMBC.2015.7319865

VaezMousavi, S. M., Barry, R. J., & Clarke, A. R. (2009). Individual differences in task-related activation and performance. *Physiology & Behavior, 98*(3), 326–330. https://doi.org/10.1016/j.physbeh.2009.06.007

Villanueva, I., Campbell, B. D., Raikes, A. C., Jones, S. H., & Putney, L. G. (2018). A multimodal exploration of engineering students' emotions and electrodermal activity in design activities. *Journal of Engineering Education, 107*(3), 414–441. https://doi.org/10.1002/jee.20225

Volet, S., Seghezzi, C., & Ritchie, S. (2019). Positive emotions in student-led collaborative science activities: Relating types and sources of emotions to engagement in learning. *Studies in Higher Education, 44*(10), 1734–1746. https://doi.org/10.1080/03075079.2019.1665314

Wagner, M., & Urhahne, D. (2021). Disentangling the effects of flipped classroom instruction in EFL secondary education: When is it effective and for whom? *Learning and Instruction, 75*, Article 101490. https://doi.org/10.1016/j.learninstruc.2021.101490

Watkins, C., Lodge, C., & Carnell, E. (2007). *Effective learning in classrooms*. Sage.

Wright, S., Dean, K. L., & Forray, J. M. (2022). Negative student emotions and educator skill in experiential education: A taxonomy of classroom activities. *Higher Education, 83*(5), 987–1002. https://doi.org/10.1007/s10734-021-00720-9

Dr. Susanna Hartikainen is researcher at Tampere University. Her research interests include active learning methods and learning-related emotions in HE.

Dr. Petri Nokelainen is professor at Tampere University. His research interests include the modeling of intensive longitudinal data.

Ilmari Puhakka is doctoral researcher at Tampere University. His research interests include learning and well-being at work and in HE.

Dr. Riikka Kangaslampi is university lecturer in mathematics at Tampere University. She conducts longitudinal mathematics education research.