Learning and Instruction 103 (2026) 102310

Contents lists available at ScienceDirect

Learning and Instruction

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

Multimodal perspectives on affective dynamics in an intelligent tutoring system

Anja Henke [a,b,*], Jason M. Harley [c,d], Negar Matin [d], Johann Chevalère [a,b,e], Verena V. Hafner [b,e], Niels Pinkwart [f,g], Rebecca Lazarides [a,b]

[a] University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476, Potsdam, Germany
[b] Cluster of Excellence "Science of Intelligence" (SCIoI), Technische Universität Berlin, Marchstr. 23 – MAR 5-2, 10587, Berlin, Germany
[c] Department of Surgery, Faculty of Medicine and Health Sciences, McGill University, Montreal General Hospital, 1650 Cedar Avenue, T5-110, Montreal, QC, H3G 1A4, Canada
[d] Research Institute of the McGill University Health Centre (RI-MUHC), Montreal General Hospital, 1650 Cedar Ave, R1.112, Montreal, QC, H3G, 1A4, Canada
[e] Laboratoire de Psychologie Sociale et Cognitive (LAPSCO) UMR 6024, CNRS & Université Clermont-Auvergne, 34 Av. Carnot, TSA 60401 - 63001 Clermont-Ferrand Cedex 1, France
[f] Humboldt-University Berlin, Unter Den Linden 6, 10099, Berlin, Germany
[g] Educational Technology Lab, German Research Center for Artificial Intelligence (DFKI), Kesselhaus in der Lanolinfabrik, Salzufer 15/16, 10587, Berlin, Germany

ARTICLE INFO

Keywords:
Multimodal assessment
Emotion
Affective activation
Prompt compliance
Affective dynamics

ABSTRACT

Background: Research has highlighted the relevance of affective processes for learning but often relies on self-reports without accounting for inter- and intraindividual dynamics. Multimodal research allows for a more comprehensive assessment of these dynamics. Most multimodal studies, however, rely on lab contexts, reducing ecological validity of results. The present study aims to address this gap.
Aims: We examine situated inter- and intraindividual variability in concurrent activity emotions and affective activation and their interrelations with learning behaviour (prompt compliance) and domain knowledge after learning in an intelligent tutoring system.
Sample: The sample consisted of 83 students (M_{age} = 15.52, SD = 1.90; girls = 56.6 %) from four secondary schools.
Methods: We combined logfile data (prompt compliance), with electrodermal activity assessment (affective activation), standardized tests data (domain knowledge), and two types of self-report data: experience-sampling to capture concurrent activity emotions over time and topic-related emotions. Two-level dynamic structural equation models were applied.
Results: Results revealed concurrent emotions and prompt compliance to be self-predictive over time. Topic-related boredom and confusion were linked to concurrent boredom and confusion. Person-level affective activation negatively predicted person-level domain knowledge after learning, suggesting high activation may deplete cognitive resources. Topic-related boredom and confusion positively predicted increases in aggregated affective activation, suggesting these emotions drive activation.
Conclusions: The study shows that once an emotion is established, it can persist for the learning session and that especially initial boredom can be detrimental for following learning processes and that particularly affective activation can deplete learning performance (domain knowledge after learning).

1. Introduction

A large amount of research within the last decades showed that achievement emotions are important for learning behaviours and successful learning outcomes (for reviews, see also: Frenzel et al., 2024; Harley & Pekrun, 2024; Pekrun, 2024). Based on Pekrun's control-value

This article is part of a special issue entitled: Implementing Multimodal Learning Analytics published in Learning and Instruction.

* Corresponding author. University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476, Potsdam, Germany.
E-mail addresses: anja.henke@uni-potsdam.de (A. Henke), jason.harley@mcgill.ca (J.M. Harley), negar.matin@mail.mcgill.ca (N. Matin), johann.chevalere@uca.fr (J. Chevalère), hafner@informatik.hu-berlin.de (V.V. Hafner), pinkwart@hu-berlin.de (N. Pinkwart), rebecca.lazarides@uni-potsdam.de (R. Lazarides).

https://doi.org/10.1016/j.learninstruc.2025.102310
Received 29 November 2024; Received in revised form 2 December 2025; Accepted 23 December 2025
Available online 30 January 2026

theory (CVT; 2006), current research on achievement emotions has taken a ‘situative turn’, emphasising that learning-related emotions vary across situations and situated contexts (Pekrun & Marsh, 2022). However, research often over-relies on self-reports (e.g., Goetz et al., 2010). A trade-off of using only student self-reports is that interrelations to other self-report assessed constructs such as student engagement are overestimated or corrupted by common-method biases (Podsakoff et al., 2003). Multimodal assessments are therefore needed to enrich research designs and enable more differentiated perspectives on learning (Hung et al., 2013).

Multimodal assessments also enable research on both interindividual and intraindividual learning because assessment methods such as electrodermal activity (EDA), video observations, or eye-tracking cover fine-grained nuances in situation-specific processes inaccessible through self-reports (see also Dindar et al., 2020). Since the emergence of multimodal assessment in several fields including education more than 20 years ago (e.g., Jewitt et al., 2001; Norris, 2004), multimodal methods have allowed more comprehensive insights into learning. However, one drawback of most research using multimodal approaches is that most studies are conducted in laboratory settings instead of real classrooms reducing their ecological validity (Cukurova et al., 2020).

The present study aims to address this gap by applying situated theoretical approaches based on CVT (Pekrun, 2006) in investigating situation-specific variability (function of time) in achievement emotions. Concretely, we examine cross-lagged longitudinal interrelations of situated achievement emotions and compliance with scaffolding prompts in an intelligent tutoring system (ITS), an adaptive learning technology, in an ecological classroom setting. We thereby account for the nested nature of such processes by focusing on within- and between-person processes using a multimodal approach assessing achievement emotions through EDA measures and experience sampling methods (ESM; Moeller et al., 2023) and prompt compliance through logfile data. Domain knowledge after learning is measured using a standardised knowledge test. This innovative design enables uncovering theory-driven situation-specific affective and emotion-related dynamics in learning with new technologies in real classrooms.

2. Emotions in achievement situations

According to CVT (Pekrun, 2006, 2019, 2024; Pekrun et al., 2007), achievement emotions are central to learning processes in academic settings as they recursively influence behaviour and achievement. Achievement emotions are thereby multifaceted with physiological activation as one core aspect (Pekrun et al., 2011, 2023). Generally, positive activating emotions like enjoyment are associated with higher learning performance through higher cognitive activation while the opposite is true for negative deactivating emotions such as boredom (Camacho-Morles et al., 2021; Pekrun, 2024). However, effects can be ambiguous or situation-specific, especially for activating negative emotions like confusion which has been shown both as related and unrelated to performance in different studies (Vogl et al., 2020) and frustration which recent research has suggested may be virtually unrelated to performance (Camacho-Morles et al., 2021). However, according to the theory of cognitive disequilibrium (D’Mello & Graesser, 2010, 2012) learners activating negative emotions such as confusion (Pekrun, 2024) might also lead to a state of engagement or flow when learners solve challenges.

Although research on more fine-grained relations between emotional experiences, learning processes, and outcomes is lacking, studies have shown substantial situated variability of emotions within learners. For example, Goetz et al. (2016) revealed that about 26 % of the variability in students situated enjoyment and about 23 % of the variability of situated boredom were explained by the situation. Vilhunen et al. (2022) revealed that about 38 % variability in boredom was explained by situations, not persons (38 % for frustration, 42 % for confusion). Further, when analysing affective structures within and between individuals, Brose et al. (2015) found substantial differences suggesting that emotions should be examined on both a within-person and between-person level. Accordingly, studies should account for intraindividual situation-specific changes and differences in achievement-related activity emotions (Berweger et al., 2022). Although studies using ESM allow analysing fin-grained time-related changes of emotions, they mostly rely on self-reported data (Gabriel et al., 2019). Self-reports, however, may not capture multiple facets of emotion and multimodal research particularly research that combines ESM and non-self-report data would be beneficial (see also: Harley et al., 2015).

3. Multimodal assessment of achievement emotions

Multimodal assessment in research is concerned with integrating multimodal signals for analysis (D'Mello, 2020), allowing, for example, analysing learning processes in great detail through physiological data (Sharma & Giannakos, 2020). Especially EDA has been demonstrated as useful in examining emotions (Harley et al., 2015; Horvers et al., 2024; Moreno et al., 2024) and complementing self-reports (Dindar et al., 2020; Harley et al., 2015, 2019b). However, a systematic review by Horvers et al. (2021) revealed mixed findings concerning associations between information from EDA and self-reports on emotional experiences which may be due to different operationalisations of the constructs by the studies.

More precisely, at the moment EDA is an established physiological method to assess affective activation or arousal and although some studies have shown associations between affective states (mostly stress) and EDA, EDA cannot (yet) be considered a good indicator of discrete emotions or emotional valence (Fowles, 2008; Horvers et al., 2021). A number of studies (Ganapathy et al., 2021; Kołodziej et al., 2019; Mercado-Diaz et al., 2024; Veeranki et al., 2024) have tried to relate EDA signals with discrete emotions using elaborate machine learning and deep learning approaches, but results lack accuracy or validity for distinguishing between several emotions. However, Kołodziej et al. (2019) and Ganapathy et al. (2021) suggest that in the future, EDA may have the potential to be utilized to identify discrete emotions or valence without requiring additional situational information. Thus and in line with Pekrun et al. (2023), EDA-derived affective activation can only be understood as a measurement of one distinct component of emotions.

Thus, although multimodal studies using both self-reports and EDA data offered a clearer view of emotions while learning, EDA should be understood as measuring one of several expression components (physiological) of emotional experiences that also include experiential (e.g., self-report) and behavioural (e.g., facial expression coding; Harley et al., 2015). Whereas EDA relates to learning outcomes and behaviors, mapping it directly to specific emotions remains challenging (Horvers et al., 2021, 2024). In that sense, EDA has been established as an indicator of emotional, or affective activation (Sequeira et al., 2009).

Affective activation is one response component of emotion to a given stimulus and can be described as being psychophysiologically awake and alert (Niven & Miles, 2013). As suggested by CVT (Pekrun, 2006), activating emotions can have differing effects on learning outcomes depending on the situation and the associated discrete emotion. In line with this assumption, research has found both positive and negative relations between affective activation on academic performance (see also: Horvers et al., 2021). First advances in research on learning behaviour have shown relations between learning strategy application and physiological activation (Malmberg et al., 2021). However, more research is needed. The present study therefore includes distinctive yet complementary components of emotion to gain a more comprehensive picture of affective processes when learning with adaptive learning technologies.

So far, multimodal assessments in educational emotion research are promising for gaining a better and highly-granular understanding of affective processes. Educational research has only recently adopted this approach, mostly in lab settings, leaving knowledge about interrelations of emotional components and learning processes scarce and limited. Therefore, directions of moderating effects as well as possible moderating effects remain currently unclear. New (adaptive) technologies are promising for implementing such research because they can improve data assessment through logfiles (Paquette et al., 2021), which offer insights that can help illuminate relations between variables and their situated contexts.

4. Emotions and affective processing when learning with new technologies

Emotions are crucial for learning and have become increasingly relevant in research on new learning technologies (see also: Lajoie et al., 2020). Advances towards emotion-aware tools have been made to, for example, promote learning-enhancing emotions (Harley et al., 2017). New learning technologies like Intelligent tutoring systems (ITS) often employ prompts as a means to scaffold student learning (Roll et al., 2011). Scaffolding prompts have been shown to potentially foster performance (Kardan & Conati, 2015), which in turn could foster positive emotional experiences according to CVT (Pekrun, 2006). However, previous work (Clarebout & Elen, 2006; Harley et al., 2017) proposed that students' experiences and use of support tools (e.g., scaffolding prompts) in adaptive learning environments may depend on characteristics of the learning situation and the learner, but the authors outline that related findings remain inconclusive.

Further, research has shown that adolescents' negative affect on the person and positive affect on the situation level can affect how learners use learning technologies, more specifically, by negatively predicting prompt compliance (Murray et al., 2023; Sokolovsky et al., 2013). Compliance in this context refers to learner behaviour and whether a learner followed or did not follow an ITS' prompts and feedback (Harley et al., 2018; Kardan & Conati, 2015). Additionally, Murray et al. (2023) have shown that stressful situations, which according to Martinez et al. (2017) are accompanied by increased activation, can reduce prompt compliance. However, so far, the mechanisms between emotion, prompt compliance, and learning performance (e.g., domain knowledge after learning) are under-researched.

As new learning technologies ease analysing behaviour through logfile data (Paquette et al., 2021) and enable ESM studies without needing additional devices, they offer ideal conditions for research on behaviour and affective learning processes. However, ITSs as one example of such technology often require high self-regulation by learners (Segedy et al., 2015) and offer ample opportunity to experience a range of emotions.

5. The present study

The present study builds on theoretical work (D'Mello & Graesser, 2010; Harley, Pekrun, et al., 2019; Pekrun, 2006) and empirical work (outlined above) on emotions and affective processing in learning. We aim to better understand situation-specific variability in students' concurrent activity emotions in achievement situations and situated affective dynamics by focusing on interrelations with learning behaviour. The study aims to provide insights into ecological learning in adaptive settings and authentic environments, using a multimodal approach to improve future learning analytics technology. We use data from a research project which tested a traditional performance-adaptive ITS (Condition A) against a novel version of an ITS adaptive towards emotion and performance (Condition B) to answer the following research questions:

RQ1. How are concurrent activity emotions in achievement situations (boredom, enjoyment, confusion, frustration) and affective activation reciprocally related with prompt compliance at the within-person level?

H1. At the within-person level, enjoyment as a positive, activating emotion will be positively related to subsequent prompt compliance. Negative emotions (boredom, confusion, and frustration), in turn, are hypothesized to be negatively related to subsequent prompt compliance. We anticipate higher levels of negatively-valenced than positively-valenced affective activation. Therefore, we hypothesize affective activation to be negatively related to subsequent prompt compliance.

RQ2. Do concurrent activity emotions in achievement situations (boredom, enjoyment, confusion, frustration) and affective activation affect domain knowledge after learning at the between-person level?

H2. At the between-person level, we hypothesize that aggregated situated positive emotions (enjoyment) and aggregated situated prompt compliance will positively predict domain knowledge after learning and the reverse will be true for negative emotions (boredom, confusion, and frustration).

RQ3. Are retrospective topic-related emotions in achievement situations related to concurrent activity emotions in achievement situations, affective activation, and prompt compliance?

H3a. We hypothesize that the retrospective topic-related experience of a given emotion (enjoyment, boredom, confusion, frustration) is positively related to its respective concurrent experience. Positive emotions (enjoyment) are positively related to prompt compliance while the inverse is true for negative emotions (boredom, confusion, frustration).

H3b. We hypothesize that the retrospective topic-related experience of activating emotions (enjoyment, confusion, frustration) is positively related to and the topic-related experience of deactivating emotion (boredom) is negatively related to affective activation.

6. Methods

6.1. Sample

The present study uses data from $N = 83$ students from four public and private secondary schools in Germany ($M_{age} = 15.52$, $SD = 1.90$; girls = 56.6 %; male = 43.4 %). Originally, 69 additional students participated in the study, but had to be excluded from analyses due to compromised EDA or logfile data. This participant attrition itself reflects challenges of multimodal (especially physiological data) research in schools. The study was part of an intervention project in which students were randomly assigned to learn with one of two differently adaptive ITS: A traditional ITS using information on learners' progress and performance (Condition A) and an experimental ITS using information on learners' progressive performance and emotional experience (Condition B) to offer scaffolding prompts. Although not the focus of the present study, we tested how group membership affected outcome variables prior to analysis. The a priori analyses showed no significant differences in between these two conditions ($p_{\text{two-tailed}} \geq 0.108$; full results provided in Supplemental Material A).

Participants, and if required, their legal guardians, consented to participation prior to the study. This study received approval from the University of Potsdam Ethics Committee (number 82/2020) and from the Brandenburg Ministry of Education, Youth and Sports (number 82-E1-2020). The study complies with APA ethical standards in the treatment of the human participants.

6.2. Procedure

Data were obtained within a science week at each participating school within a real classroom setting. The procedure is depicted in Fig. 1. Within the project week, students worked with the ITS *Betty's Brain* (Biswas et al., 2016) which aims to help students learn about climate change by teaching a virtual peer agent with the help of a virtual tutor agent and has been used in a number of previous studies (e.g., Biswas et al., 2016; Munshi et al., 2018; Munshi et al., 2023; Segedy, 2014; Segedy et al., 2015). More precisely, learners were tasked to build

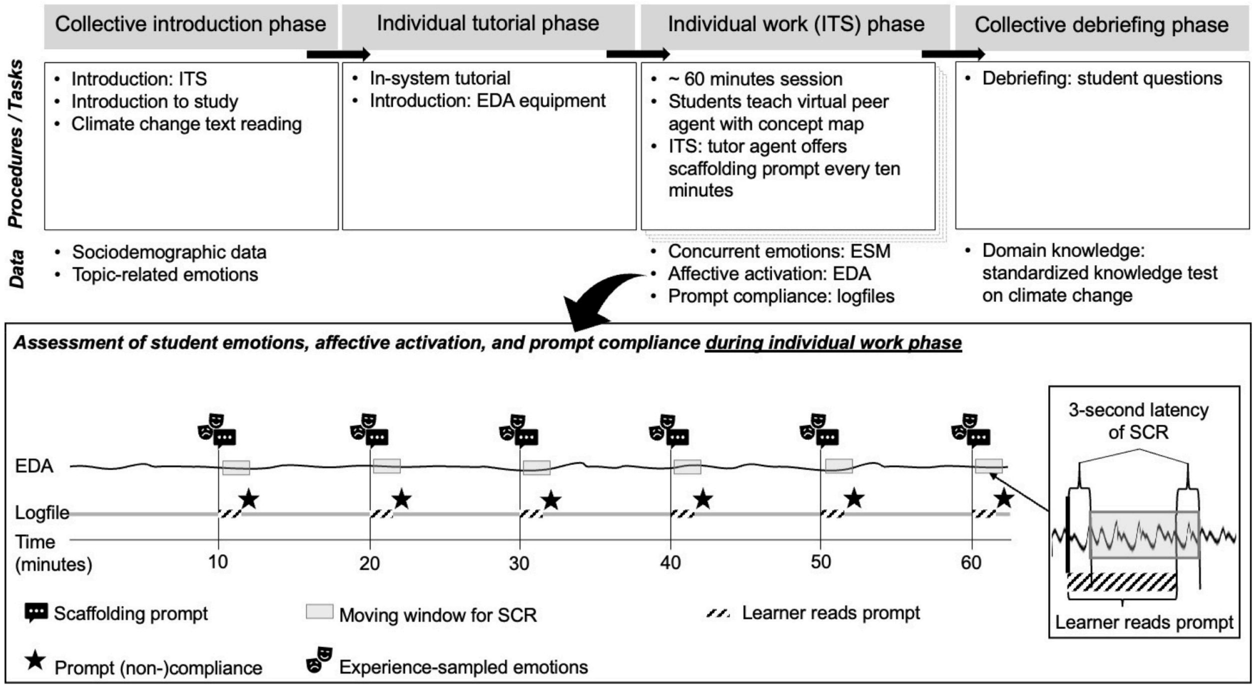

Fig. 1. Design of Study and Data Assessment within a Science Week of a Participating School; ESM and Moving Window are Detailed.

a concept map in the ITS using information from a virtual textbook and aid provided by a virtual tutor. Students could assess their concept maps' accuracy by asking the virtual peer agent to answer quiz questions posed by the virtual tutor agent. Students themselves did not answer quizzes. This teachable agent framework encourages metacognitive reflection, as students iteratively revise their maps based on Betty's quiz performance and mentor feedback. The system provides cognitive strategy prompts, aiming to foster self-regulated learning rather than deliver direct corrections (for a more comprehensive description of the ITS see also: Biswas et al. (2016)).

At the beginning of the science week, students were introduced to the ITS and the concept map method by the research team. They also read a short text about climate change and answered a questionnaire including questions about their retrospective topic-related emotions (enjoyment, boredom, confusion, frustration) as well as sociodemographics. After, students were randomly assigned to condition A or B. Students then followed an individual in-system tutorial for 1 hr to help them understand how to use the ITS. After, students worked individually for up to four 1-hr sessions. To ensure that learners could complete their current tasks without being abruptly interrupted, a few additional minutes were budgeted for each session to help students finish the learning and study-related tasks, enabling a more complete assessment of the learning process and also accounting for individual differences in pace.

During the 1-hr-long learning sessions, students were asked about their concurrent activity emotions (enjoyment, boredom, confusion, frustration) every 10 min (ESM; Moeller et al., 2023). Further, after reporting their current emotional experiences, the virtual peer agent answered a quiz evaluated by the virtual tutor agent. Following this evaluation, the virtual tutor agent offered a scaffolding prompt from a selection of prompts of four different levels of detail, ranging from very broad suggestions like reading more about a topic (level 1) to a detailed worked example presenting a clear correction to the concept map (source and target concept as well as their causal relation) and directing the learning to the relevant passage in the virtual text book (level 4). For example, scaffolding prompts, see Supplement B. The scaffolding prompt that was delivered was determined by a backend matrix (rule--based AI) which was unique between the traditional condition A and experimental condition B (see Supplement B).

Whereas prompts in condition A only targeted each learner's performance at each 10-min interval and followed the progressive hint delivery strategy described and used by, for example, Segedy (2014) with Betty's Brain. Condition B, on the other hand, offered scaffolding prompts based on an emotion-performance matrix at each 10-min interval. However, the prompts themselves did not differ between conditions, only the rationale behind the delivery differed.

The backend matrix for condition A was designed to increase the level of detail at any 10-min interval where learners had not completed a chapter of the ITS task for at least two consecutive prompt deliveries. On the other hand, the backend matrix of condition B was designed in such way that it aimed to prevent task-disengagement in learners. For example, students closer to task-disengagement (operationalised based on their performance and emotional state; D'Mello & Graesser, 2012) received more detailed prompts (e.g., "You need to add the link *precipitation reduces droughts to the map*. This portion of the science book called "Droughts and Water Cycle" explains the relationship. See if you can figure out which part of this passage explains the relationship between precipitation and droughts. [display of the relevant section of the virtual text book]"; level 4). On the other hand, students who were high-performing and closer to high task engagement (operationalised based on their performance and emotional state) received less detailed scaffolding prompts (e.g., "I still think you need to take a look at the concept precipitation. Something's not quite right with it."; level 1). The backend matrices used in the two conditions and example scaffolding prompts for the four levels of detail can be found in Supplement B. Both conditions were identical, except for their distinct backend matrices, thus learners where unaware of their condition assignment while working with the ITS (single blind study).

While working with the ITS, students could freely choose how to approach the task. After the individual sessions, students returned to the classroom to answer a standardized knowledge test as well as for a collective debrief during which the study was explained once more and students could ask questions. The intervention took place in the learners' regular classrooms and the desks were in a configuration typical of a traditional direct instruction setting, which in most cases corresponded to the standard seating plan used in these classrooms. During the intervention, at least one regular teacher was present. The intervention was led by one researcher and no more than two researchers were present in one classroom at a time while the learners worked on the independent learning tasks. Although teachers were present, researchers (mostly trained teachers or teacher students) took on the role of teachers and intervened when learners had technical issues or comprehension issues.

6.3. Measures

6.3.1. Prompt compliance

Consistent with Kardan and Conati (2015) and Harley et al. (2018), students' prompt compliance was assessed by processing logfile data. The prompts provided by the ITS every 10 min as well as learner behaviour following each scaffolding prompt were retrieved from the logfile data. Students' behaviour following each scaffolding prompt instance was evaluated to either follow or disregard a given prompt. For example, if a prompt suggested removing a specific causal link from the concept map and the learner removed said causal link, the behaviour was coded as compliant. Conversely, if a learner removed another causal link or opened the virtual textbook, for example, their behaviour was coded as non-compliant.

6.3.2. Retrospective topic-related emotions

Prior to working with the ITS and after reading a short text about climate change, students reported their retrospective topic-related emotions using the epistemically-related emotion scale (Pekrun et al., 2017) with an instruction adapted to suit the questionnaire design (Supplemental Material C). A five-point Likert-scale asking students to which degree they experienced discrete emotions ranging from “not at all” (0) to “very strongly” (4) was used. The assessed emotions were internally consistent: enjoyment ($\alpha = .834$), boredom ($\alpha = .649$), confusion ($\alpha = .816$), frustration ($\alpha = .874$). We will refer to this construct as topic-related emotion henceforth.

6.3.3. Concurrent activity emotions

Students’ concurrent activity emotions (enjoyment, boredom, frustration, and confusion) were assessed every 10 min using ESM (Moeller et al., 2023). We used the short version of the epistemic emotions questionnaire (Pekrun et al., 2017) directly asking students to which degree they experienced these four discrete activity emotions on a five-point Likert-scale using a similar wording used for topic-related emotions but only as single-item-assessment and adapted to suit the in-system assessment (Supplemental Material C). Henceforth, this construct will be referred to as concurrent emotion.

6.3.4. Affective activation

During the learning sessions, students wore wrist bracelets with electrodes connected to the palm of their non-dominant hand to measure EDA (Movisens EdaMove 4). The raw EDA data had been sampled at 32 Hz but was downsampled to 8 Hz to remove artifacts. We then applied a lowpass filter (cut-off frequency of 1 Hz at order 1), manually removed any remaining artifacts, conducted continuous decomposition analysis, an analysis rather less prone to artifacts (Benedek & Kaernbach, 2010) using Ledalab 3.4.9 (Benedek & Kaernbach, 2006-2016), and retrieved skin conductance responses (SCR). Using the moving-window method proposed by Matin et al. (2024) as well as (Moreno et al., 2024), we calculated the average frequency density of SCRs associated with scaffolding prompt delivery, applying a 3-s latency period (see Fig. 1). Specifically, after a scaffolding prompt was delivered, we waited 3 s to allow for a physiological response delay.

We then measured the frequency of SCRs from that point until the learner's next action in the ITS. This SCR frequency was averaged over that time window and is referred to as the average frequency density of SCRs. Affective activation as a physiological response is a subconscious reaction as a component of emotions (see also: Scherer, 2009). Such implicit emotional processes do not measure the same component of emotion that self-reports do. Thereby, measuring physiological emotional responses in addition to self-reported emotional experience offers a broader picture of learners' emotional processes while learning. Another benefit of assessing such physiological measures is that they do not interfere with a learning task as they do not require students to answer questions while also working on a learning task. Further, such measures also allow finer measurement granularity as their measurement can be continuous. However, since we aligned and linked students' physiological data to the ESM data and behavioural data in the present study, we do not make use of said finer granularity.

6.3.5. Domain knowledge after learning

Following working with the ITS, students answered a set of multiple-choice questions as well as an open-ended question about climate change to assess their domain knowledge after learning. The contents of the self-developed standardized test encompassed only constructs and relations between constructs that were contents in the ITS. Thereby, we aimed to assess learner's domain knowledge directly related to the contents of the ITS independent of the quality of the concept map they built within the ITS. In the self-developed standardized knowledge test, students could score up to 20 points from both multiple-choice questions and the open-ended question combined. The test was developed in alignment with recommendations and regulations for design of school examinations in Germany in collaboration with geography teachers. The internal consistency of the test for the sample used in the present study is $\alpha = .774$. The open-ended question was graded by two independent raters (interrater agreement at $\alpha = .949$) based on a coding scheme by Novak and Gowin (1984). The coding scheme by Novak and Gowin (1984) was developed for concept maps. The open question used in the test, was designed in such way, that learners could answer it by noting down concepts and relations as had been constructed in the concept map within the ITS. Therefore, when grading the open-ended question, the answers were understood as written versions of concept maps and thus their grading scheme was used.

6.4. Statistical analyses

We focused on data from the first learning session. An a priori power analysis was performed using G*Power (Faul et al., 2009) to determine the minimum sample size necessary for detecting the incremental contribution of a single predictor within a series of multiple regression models comprising three predictors. Based on Cohen's (1988) conventions, assuming an alpha level of 0.05, statistical power of 0.80, and a medium effect size ($f^2 = 0.15$, approximately corresponding to a change in R^2 of 0.13), the required minimum sample size was estimated at 55 participants. A subsequent post hoc power analysis, based on the actual sample size of 83 participants and an observed effect size of $f^2 = 0.10$ ($\Delta R^2 \approx 0.09$), indicated achieved power of 0.81. Accordingly, the study was adequately powered to detect effects approaching medium magnitude and greater.

To answer our research questions, we specified four two-level dynamic structural equation models (DSEM; Asparouhov et al., 2017; McNeish & Hamaker, 2020) separately for each emotion (both retrospective topic-related emotion and the according recurrent activity emotion) to allow parsimony and clarity. Further, this approach aligns with correlational results suggesting no unified association between topic-related and concurrent emotions in our dataset. Two-level DSEM allow to model autoregressive, cross-lagged, and cross-level relations of nested intensive longitudinal data. We used a two-level approach because situations ($n = 438$) were nested in students ($n = 83$). Prior analyses of ICC1 (Table 1) indicated that a substantial amount of students’ emotions in the learning situation was attributable to their individual characteristics. Thus, multilevel modelling was applied to account for the hierarchical structure of the data. Models were estimated using Bayesian estimation with 3.000 iterations and thinning set at 5 and the TINTERVAL-function in Mplus 8.10 (Muthén & Muthén,

Table 1
Correlations, ICC1, and descriptive statistics for study variables.

Variables	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Prompt compliance	–	.01	.04	−.04	.03	−.01	–	–	–	–	–
2. Affective activation	.33	–	.05	.02	−.02	.03	–	–	–	–	–
Concurrent											
3. Boredom	−.22	−.38**	–	.02	.32**	.34**	–	–	–	–	–
4. Enjoyment	−.10	−.06	−.37*	–	−.09	−.25**	–	–	–	–	–
5. Confusion	−.03	−.28	.48**	−.26	–	.53**	–	–	–	–	–
6. Frustration	.17	−.15	.61**	−.29	.85**	–	–	–	–	–	–
Topic-related											
7. Boredom	−.05	−.13	.17	.14	.10	.09	–	–	–	–	–
8. Enjoyment	.25	−.15	.07	.18	.09	.13	.22	–	–	–	–
9. Confusion	−.07	−.08	.23	.16	.32**	.28*	.53**	.04	–	–	–
10. Frustration	−.24	−.10	−.01	.03	.14	.15	.11	−.35**	.52**	–	–
11. Domain knowledge	−.17	−.26	−.14	−.06	−.09	−.06	−.03	.02	−.16	.14	–
ICC1	.077	.230	.595	.592	.606	.605	–	–	–	–	–
M	0.25	1.03	1.61	1.64	2.04	1.88	1.21	1.53	1.15	1.62	13.05
SD	1.30	0.80	1.30	1.24	1.15	1.70	0.69	0.84	0.84	1.10	5.72
Range	0–1	0–5	0–4	0–4	0–4	0–4	0–4	0–4	0–4	0–4	0–20

Note. Estimates below diagonal show correlations on the between level; estimates above the diagonal show correlations on the within level.
Means and standard deviations reported for between level.
$^{*}p < .05$. $^{**}p < .01$.

1998-2023).

In our models, we modelled within-person autoregressions for concurrent emotion, prompt compliance, and affective activation with a lag interval of 10 min (ESM interval). On the within level, we modelled cross-lagged relations between concurrent emotion, prompt compliance, and affective activation (RQ1). On the between-level, we included time-aggregated concurrent situation-specific achievement emotions, prompt compliance, domain knowledge after learning operationalised through the standardized knowledge test (as the outcome variable) and topic-related emotions. Particularly, we modelled relations between slopes and time-aggregated levels of concurrent activity related emotion and domain knowledge after learning (RQ2). Additionally, we modelled between- and cross-level effects testing whether topic-related emotion (between level) related to the aggregations (between level) and changes (autoregressive slopes; within level) of learners' concurrent emotions (enjoyment, boredom, confusion, frustration), prompt compliance, and affective activation over the learning session (RQ3). Further, we tested whether learners' aggregated (between level) or change in (autoregressive slopes, within level) concurrent emotion (enjoyment, boredom, confusion, frustration), prompt compliance, and affective activation over the learning session related to domain knowledge after learning (between level). Fig. 2 depicts the structure of the DSEMs.

7. Results

7.1. Descriptive statistics and correlations

Descriptive and correlative statistics for study variables are presented in Table 1. On the within level, prompt compliance was uncorrelated with any study variables. On the between level, prompt compliance and domain knowledge after learning were uncorrelated with any study variables. Overall, we found few significant correlations. On the between level, affective activation was significantly moderately negatively correlated with concurrent boredom. Further, topic-related confusion significantly positively correlated with concurrent frustration and concurrent confusion. We found no statistically significant correlations between the other topic-related emotions and their respective concurrent counterpart. Topic-related confusion was further statistically significantly positively associated with concurrent frustration and topic-related boredom. Again, there were no statistically significant associations between other topic-related and current emotions. Further, we tested whether there were group differences in time spent on the learning task or the average number of ESM measurements between conditions A and B. T-test results showed no significant differences in time spent on the learning task ($M_{\text{condition A}}$ = 1:07:43, SD = 0:09:46; $M_{\text{condition B}}$ = 1:04:58, SD = 0:12:20; p = .290) and significant group differences for number of ESM measures per participant ($M_{\text{condition A}} = 5.31$, $SD = 1.06$; $M_{\text{condition B}} = 5.26$, $SD = 1.21$; $p = .826$).

7.2. Two-level dynamic structural equation models

Table 2 presents results for the four two-level DSEMs. The four models offered low to high variance explanation (see Table 2) on the within level and generally low variance explanation on the between level but could explain variances for domain knowledge after learning on the between level well.

The **DSEM for enjoyment** showed significant positive autoregressive effects for concurrent enjoyment and prompt compliance but not for affective activation. Cross-lagged effects were not significant. On the between level, aggregated concurrent enjoyment and prompt compliance did not while aggregated affective activation significantly negatively did predict domain knowledge after learning. Cross level effects between topic-related enjoyment and changes in concurrent enjoyment, affective activation, and prompt compliance were not significant. Further, changes in concurrent enjoyment, affective activation, and prompt compliance did not predict domain knowledge after learning.

The **DSEM for boredom** showed no significant cross-lagged effects but significant positive autoregressive effects for concurrent boredom and concurrent prompt compliance on the within level. On the between level, topic-related boredom positively significantly related to aggregated concurrent boredom but not to aggregated affective activation and aggregated prompt compliance. Further, aggregated concurrent boredom and aggregated affective activation negatively significantly predicted domain knowledge after learning. Aggregated prompt compliance was not significantly related to domain knowledge after learning. Cross level effects were significant for topic-related boredom and the autoregressive boredom slope (negative relation) and the autoregressive affective activation slope (positive relation).

The **DSEM for confusion** showed significant positive autoregressive effects for concurrent confusion and prompt compliance but no significant cross-lagged effects. On the between level, topic-related confusion only significantly positively related to learners' aggregated concurrent confusion. Domain knowledge after learning was negatively predicted by aggregated affective activation but not concurrent confusion or prompt compliance. Cross level effects on domain knowledge after

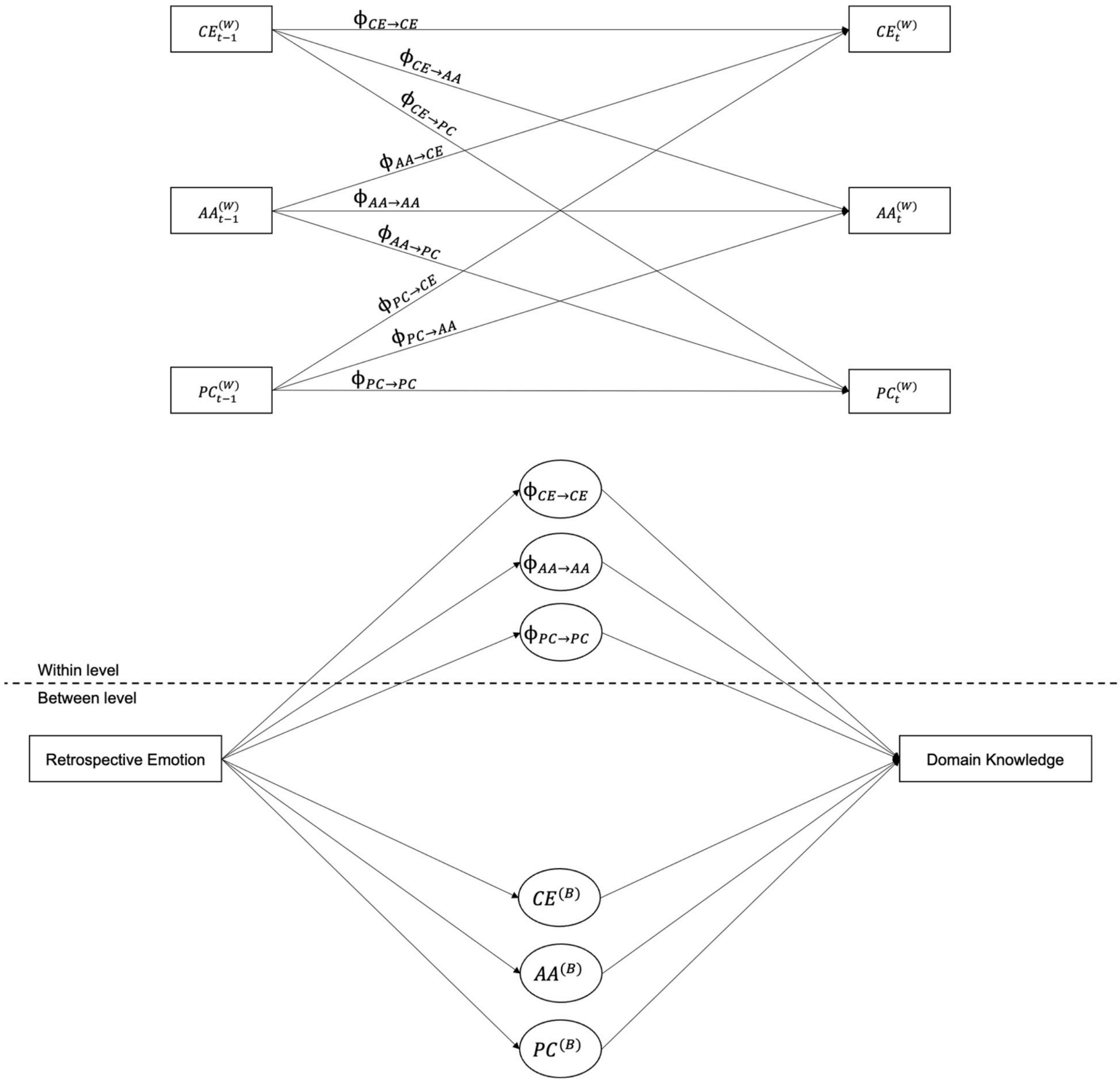

Fig. 2. *Schematic Representation of the Examined Relations in the DSEMs*
Note. CE = Concurrent emotion, AA = Affective activation (EDA SCRs), PC = prompt compliance, ϕ = Slope, $^{(W)}$ = within level, $^{(B)}$ = between level, $_{t\text{-}1}$ = timepoint prior to a given timepoint, $_{t}$ = given timepoint.

learning were also not significant but topic-related confusion positively related to the autoregressive slope of affective activation but not the autoregressive slopes of confusion or prompt compliance.

The **DSEM for frustration** showed significant positive autoregressive effects for concurrent frustration and prompt compliance but was not significant for affective activation and the cross-lagged effect on the within level. Similar to the between level of DSEM for enjoyment, topic-related confusion was not significantly related to the aggregated concurrent frustration, affective activation, and prompt compliance. Domain knowledge after learning was only significantly predicted by affective activation (negative relation). Similar to the DSEM for enjoyment, no significant cross level effects were found. It should be noted that the model for frustration was less stable than the other models and results may be interpreted with caution.

8. Discussion

Previous work has highlighted the relevance of emotions and affective processes for learning (e.g., Harley, Pekrun, et al., 2019; Pekrun, 2024), especially as part of learning analytical tools like ITS (Harley et al., 2017). However, research on fine-grained dynamics of these multi-component processes and their roles for effective learning remain under-researched, especially in regards to multimodal inter- and intra-personal differences in real-classroom settings. The present study showed that situated discrete concurrent emotions, can be self--predictive — but only the discrete experience (ESM data) not the emotion's physical component (EDA data), within learners. Further, prompt compliance on the within level was also self-predictive. Although we found no significant relations between situated affective activation and concurrent emotions, affective activation negatively predicted domain knowledge after learning, measured as a post-test knowledge test, between learners. Further, students' topic-related boredom and confusion before the learning task positively related to concurrent boredom and confusion respectively. Lastly, both topic-related boredom and confusion positively predicted the autoregressive slopes of affective activation and topic-related boredom negatively predicted the autoregressive slope

Table 2
DSEM results.

	Enjoyment			Boredom			Confusion			Frustration		
	B	95 % CI	p	B	95 % CI	p	B	95 % CI	p	B	95 % CI	p
Within level (time)												
Autoregressive effects												
$\phi_{CE \to CE}$	0.60	[0.37,0.87]	<.001	0.36	[0.18,0.56]	<.001	0.47	[0.30,0.71]	<.001	0.69	[0.37,0.96]	<.001
$\phi_{AA \to AA}$	−0.09	[-0.24,0.04]	.091	−0.07	[-0.23,0.07]	.200	−0.09	[-0.23,0.04]	.105	−0.08	[-0.22,0.04]	.090
$\phi_{PC \to PC}$	0.09	[0.00,0.22]	.023	0.10	[0.01,0.24]	.011	0.09	[0.01,0.22]	.015	0.06	[0.00,0.19]	.028
Cross-lagged effects												
$\phi_{CE \to PC}$	0.01	[-0.07,0.10]	.365	0.00	[-0.07,0.08]	.490	0.01	[-0.08,0.10]	.438	0.00	[-0.11,0.12]	.493
$\phi_{CE \to AA}$	−0.06	[-0.21,0.09]	.200	0.02	[-0.10,0.13]	.400	−0.09	[-0.21,0.05]	.108	−0.02	[-0.21,0.15]	.426
$\phi_{AA \to CE}$	−0.01	[-0.05,0.04]	.382	−0.02	[-0.10,0.04]	.240	0.02	[-0.03,0.09]	.204	0.01	[-0.04,0.07]	.361
$\phi_{AA \to PC}$	0.00	[-0.06,0.07]	.498	0.00	[-0.07,0.07]	.490	0.00	[-0.07,0.06]	.475	0.00	[-0.06,0.06]	.486
$\phi_{PC \to CE}$	0.01	[-0.05,0.07]	.338	0.00	[-0.06,0.06]	.495	0.01	[-0.06,0.07]	.391	0.00	[-0.07,0.06]	.443
$\phi_{PC \to PC}$	0.04	[-0.06,0.13]	.216	0.05	[-0.05,0.14]	.179	0.04	[-0.06,0.14]	.205	0.04	[-0.07,0.14]	.228
Between level (student)												
Effects of topic-related activity emotion												
Topic-related emotion → $M_{CE^{(B)}}$	0.26	[-0.10,0.61]	.071	0.30	[0.00,0.58]	.023	0.39	[0.08,0.66]	.007	0.26	[-0.39,0.99]	.162
Topic-related emotion → $M_{AA^{(B)}}$	−0.12	[-0.38,0.16]	.205	−0.14	[-0.42,0.15]	.164	−0.06	[-0.33,0.21]	.329	−0.12	[-0.38,0.15]	.203
Topic-related emotion → $M_{PC^{(B)}}$	0.27	[-0.38,0.77]	.215	−0.20	[-0.72,0.47]	.287	−0.15	[-0.68,0.49]	.336	−0.14	[-0.66,0.48]	.328
Effects on domain knowledge after learning												
$CE^{(B)}$ → domain knowledge after learning	−0.19	[-0.55,0.21]	.158	−0.26	[-0.54,0.04]	.046	0.00	[-0.33,0.35]	.493	−0.07	[-0.62,0.76]	.382
$AA^{(B)}$ → domain knowledge after learning	−0.26	[-0.50,0.01]	.030	−0.33	[-0.57,-0.04]	.013	−0.29	[-0.52,-0.02]	.018	−0.28	[-0.53,0.01]	.026
$PC^{(B)}$ → domain knowledge after learning	0.11	[-0.39,0.57]	.328	0.06	[-0.44,0.55]	.393	0.09	[-0.44,0.57]	.358	0.11	[-0.38,0.56]	.330
Effects of topic emotion on autoregressions												
Topic-related emotion → $\phi_{CE \to CE}$	−0.13	[-0.66,0.50]	.314	−0.50	[-0.89,0.05]	.041	−0.28	[-0.77,0.33]	.191	−0.33	[-0.94,0.47]	.196
Topic-related emotion → $\phi_{AA \to AA}$	0.12	[-0.80,0.92]	.406	0.80	[-0.02,1.00]	.027	0.83	[0.26,1.00]	.004	0.65	[-0.30,0.99]	.085
Topic-related emotion → $\phi_{PC \to PC}$	0.60	[-0.43,0.98]	.120	0.45	[-0.57,0.96]	.205	0.60	[-0.33,0.97]	.098	0.35	[-0.63,0.94]	.244
Effects of autoregressions on domain knowledge after learning												
$\phi_{CE \to CE}$ → domain knowledge after learning	−0.27	[-0.72,0.40]	.166	0.04	[-0.60,0.71]	.447	−0.32	[-0.78,0.21]	.106	−0.29	[-0.85,0.55]	.227
$\phi_{AA \to AA}$ → domain knowledge after learning	−0.11	[-0.83,0.65]	.385	−0.06	[-1.07,0.72]	.438	−0.30	[-1.07,0.60]	.218	−0.19	[-1.11,0.68]	.330
$\phi_{PC \to PC}$ → domain knowledge after learning	−0.04	[-0.66,0.68]	.456	0.03	[-0.71,0.99]	.471	−0.02	[-0.86,0.78]	.484	0.04	[-0.69,0.81]	.463
Explained variances	R^2	**95 % CI**	p	R^2	**95 % CI**	p	R^2	**95 % CI**	p	R^2	**95 % CI**	p
Within level												
Concurrent emotion	0.44	[0.25,0.77]	<.001	0.29	[0.17,0.42]	<.001	0.36	[0.22,0.56]	<.001	0.53	[0.26,0.92]	<.001
Affective activation	0.15	[0.06,0.25]	<.001	0.18	[0.09,0.28]	<.001	0.15	[0.07,0.26]	<.001	0.16	[0.08,0.27]	<.001
Prompt compliance	0.09	[0.06,0.15]	<.001	0.08	[0.05,0.14]	<.001	0.08	[0.05,0.15]	<.001	0.10	[0.05,0.24]	<.001
Between level												
Concurrent emotion	0.07	[0.00,0.38]	<.001	0.09	[0.00,0.34]	<.001	0.15	[0.01,0.44]	<.001	0.08	[0.00,0.97]	<.001
Affective activation	0.02	[0.00,0.14]	<.001	0.02	[0.00,0.17]	<.001	0.01	[0.00,0.12]	<.001	0.02	[0.00,0.14]	<.001
Prompt compliance	0.09	[0.00,0.60]	<.001	0.08	[0.00,0.53]	<.001	0.06	[0.00,0.47]	<.001	0.05	[0.00,0.44]	<.001
Domain knowledge after learning	0.51	[0.15,0.90]	<.001	0.47	[0.15,0.91]	<.001	0.53	[0.17,0.91]	<.001	0.55	[0.17,0.93]	<.001
$\phi_{CE \to CE}$	0.05	[0.00,0.51]	<.001	0.25	[0.00,0.79]	<.001	0.09	[0.00,0.59]	<.001	0.14	[0.00,0.88]	<.001
$\phi_{AA \to AA}$	0.14	[0.00,0.86]	<.001	0.63	[0.02,0.99]	<.001	0.69	[0.07,0.99]	<.001	0.43	[0.00,0.98]	<.001
$\phi_{PC \to PC}$	0.37	[0.00,0.96]	<.001	0.24	[0.00,0.93]	<.001	0.37	[0.00,0.95]	<.001	0.18	[0.00,0.88]	<.001
Autoregressive Slopes	M			M			M			M		
$\phi_{CE \to CE}$	1.63			1.19			19.15			16.72		
$\phi_{AA \to AA}$	−0.77			−0.73			−0.66			−0.70		
$\phi_{PC \to PC}$	1.01			1.03			0.78			0.69		

Note. CE = Concurrent emotion, AA = Affective activation (EDA SCRs), PC = Prompt compliance.

of boredom.

8.1. Concurrent activity emotions, affective activation, and prompt compliance

Contrasting **H1**, we did not find any relations between components of emotions and prompt compliance. One possible explanation could be other factors such as prompt comprehension). For example, previous studies (Munshi et al., 2018, 2023) found that (non-) compliance was often a result of low comprehension—especially among lower-performing learners. However, we do not have student reported data or other measurements indicative of prompt comprehension and thus we cannot estimate in how far (non-) comprehension of the scaffolding prompts may have played a role in our study. Another possible explanation may be low regulation abilities of learners. Prior research (e.g., Baker et al., 2010; Camacho-Morles et al., 2021; Harley, Pekrun, et al., 2019, see also: Schweder & Raufelder, 2019) showed that achievement emotions can impact student behaviour, especially in those with lower regulation abilities and younger students (Camacho-Morles et al., 2021; Harley, Pekrun, et al., 2019). However, our study does not assess self-regulation which may act as an explaining factor.

8.2. Concurrent activity emotions, affective activation, and domain knowledge after learning

Contrasting **H2**, emotional experiences during learning were unrelated to domain knowledge after learning. However, affective activation, as a physiological component of emotion, during learning negatively predicted domain knowledge after learning. One possible explanation could be rooted in high activation using cognitive resources which decreases available resources for cognitive learning processes (Meinhardt & Pekrun, 2003). Another explanation could be based in differing effects of components of emotional experiences; in this sense, it may be possible that the physiological component of emotions, not the experiential (self-report) drives effects on domain knowledge after learning—especially when self-reported emotions are only moderate rather than high in level. Since affective activation was significantly negatively related to domain knowledge after learning, too high affective activation may be best avoided while interacting with ITSs like Betty's Brain. One solution could be interventions by teachers or the ITS aiming at reducing affective activation (or physiological arousal) when detected. For example, in instances of high affective activation, learners may be encouraged to take a short break, participate in a short breathing exercise or similar activities as such meditating exercise at the beginning of a class have been shown to help learners feel more calm (Gardner & Kerridge, 2019), or receive guidance (either from the ITS or a present teacher) on regulating them as teacher support has been shown to be positively related to emotion regulation (Somerville et al., 2024).

8.3. Retrospective topic-related emotions, concurrent activity emotions, affective activation, and prompt compliance

H3a and **H3b** were partially confirmed. While topic-related enjoyment and frustration were not related to their respective concurrent emotions, boredom and confusion were. Therefore, students who went into the learning task after reporting experiencing higher levels of boredom or confusion about climate change (topic) also experienced higher levels of concurrent boredom or confusion respectively while learning with Betty's Brain about the same topic. One possible explanation for the relation between topic-related and concurrent emotions for boredom and confusion is that learners experienced them more in response to the learning content. This potential explanation may also be indicated by the results of the correlational analysis. The lack of significant associations between retrospective topic-related and concurrent activity emotions, except for confusion, may also reflect differences in object foci of these constructs. Whereas retrospective emotions were more directed at the topic (climate change), concurrent activity emotions were more likely elicited by the activity (working with the ITS). This difference in object focus may explain said lack of significant associations and relations.

The positive correlation between topic-related confusion, enjoyment and frustration on the other hand may have been related more to other aspects of their interaction, including interacting with the pedagogical agents; whereas the self-report was specific about the target (object foci) emotion: the topic. The concurrent emotion self-report, on the other hand, was simply how they were feeling right then—not what triggered their emotional response. Research has shown that learners can and do report experiencing significantly different emotions toward different targets, even when the self-reports are asked one after the other (Harley et al., 2020). Further, self-report emotions directed toward different targets can have different statistically significant relations with antecedents and outcomes (Harley, Liu, et al., 2019). Our findings highlight the importance of factoring in both time frame and target of an emotion in measurement and interpretation (Harley, Pekrun, et al., 2019; Harley & Pekrun, 2024).

8.4. Additional findings

Additional results showed that topic-related boredom negatively predicted the autoregressive slope of concurrent boredom. More precisely, higher levels of topic-related boredom related to less increases of concurrent boredom over time; or in other words related to a more stable experience of boredom while learning. Such effect of learners' topic-related emotion on concurrent emotion was only found for boredom but not enjoyment, confusion, or frustration. Further, topic-related boredom and confusion positively predicted the autoregressive slopes of affective activation in both models. Although this finding is not surprising for confusion as an *activating* achievement emotion (Pekrun, 2024), the finding is surprising for boredom as a *deactivating* achievement emotion. More precisely, both topic-related boredom and confusion prior to the learning task related to higher increases in affective activation while learning.

8.5. Affective dynamics during learning with an ITS

Overall, our study shows that (a) once prompt compliance in a learner was 'established', it persisted, (b) that prompt compliance was not clearly driven by emotional states or affective activation, (c) and was not a clear driver of learning performance (i.e., domain knowledge after learning). Although our study cannot offer conclusive insights on why learners comply with scaffolding prompts (or not), it is possible that their prompt compliance is related to how the prompts are delivered or on learner characteristics outside of the scope of the present study.

Given that affective activation while learning was negatively related to domain knowledge after learning, these findings suggest that effects of initial topic-related emotions may spill over into the entire learning session. This spilling may be especially true for boredom, a highly prevalent emotion associated with learning (Goetz et al., 2024). Such persistence of boredom has also been found by Baker et al. (2010) and has further been found to be more likely than for other cognitive affective states to persist across three different computer-based learning environments (Baker et al., 2010). Moreover, Bench and Lench (2013) describe that when other emotional states or goals fade, boredom becomes increasingly prominent until something novel (e.g., a new task) is presented to the learner. D'Mello et al. (2009) propose a vicious cycle of negative cognitive-affective states persisting or transition to other negative cognitive-affective states rather than transitioning to more positive cognitive-affective states. Further, Tze et al. (2016) provide meta-analytical evidence that boredom negatively relates to academic performance. In line with the above outlined previous research, our findings therefore highlight the high relevance of boredom for learning. Further, our findings also point towards the distinctively difference relations between different components of emotion and learning. Especially against the background of findings by Goetz and colleagues (2014) suggesting the existence of differently activating types of boredom, our findings provide novel potential support with physiological data for these findings and thus go beyond the works outlined above by including several modalities of emotional experiences.

Although D'Mello and Graesser (2012) describe a process of task-disengagement between confusion and boredom, they also describe deeper learning in high-activation situations caused by challenge (leading to confusion). Therefore, it is possible that high affective activation here also indicates (over-)challenge which results in task-disengagement and thus lower domain knowledge after learning—explaining the influence of confusion. Overall, prompt compliance in the present study was rather low and even when learners complied with a prompt, the provided scaffolding may have been insufficient to help learners adequately. If the scaffolding prompts failed to adequately meet learners' instructional needs or foster deeper understanding, learners may have been over-challenged. Further, although the prompts were designed to offer cognitive strategies, they may not have been able to sufficiently foster self-regulated learning and thus learners may have over-relied on the prompts to finish the learning task or relied on behaviour unconducive to learning (e.g., behavioural patterns associated with gaming the system). Additionally, it may be possible that prompt compliance may have been related to domain knowledge after learning but such effect may have been too small to be detected or may have been over-shadowed by other relations in the model (e.g., negative emotions during learning may have over-powered small effects of the scaffolding prompts) in the present sample.

In theory, the adaptive environment of the ITS should prevent both over-challenge and task-disengagement (in both conditions). Given the overall low prompt compliance of learners in the present study, these findings highlight the importance of offering adequate scaffolding and support for future learning analytics, by for example adopting methods to promote unsolicited hint usage (Maniktala et al., 2020). Unfortunately, the present study cannot explain effects on prompt compliance except for its autoregressive relation. However, future research could try to better understand under which circumstances learners use scaffolding prompts effectively and whether this type of scaffolding used here is even accessible to students.

8.6. Limitations

The present study has a number of limitations. Although the study started with a rather large sample for studies using modern learning technologies in ecological classroom settings, a large part of the participants had to be excluded from analysis, reducing the sample size. Further, although we had not found differences between conditions A and B and thus analysed both conditions as one sample, this needs to be considered as a limitation. A previous investigation using a different subset of the study testing these two conditions against each other (Chevalére et al., 2023), only found differences in experiencing boredom in the second half of the first learning session among higher and lower performing learners (higher performing students experiences significantly lower boredom). While worth noting we see this as a minor limitation because it was only the second half of the first learning session (limited and specific period) and only among higher performing learners, not all learners. Further, these findings are not directly relevant to our research questions.

Though we find that students mainly do not follow the provided scaffolding prompts, the present study cannot explain the rather low compliance. If students misunderstood prompts, for example, this could have significantly impacted prompt compliance. Therefore, more research on prompt compliance as an outcome is needed. Additionally, although the interval of 10 min for experience sampling is more granular than pre-post-test study designs, it may not be sufficient to capture adequate images of transient or fleeting emotions. Future studies could consider video-based emotion detection as an additional fine-grained emotional information, although this modality would mostly provide information on affective expression as a component of emotion.

Moreover, though we examined several emotions, our statistical modelling does not account for mixed emotions, all components of emotion, nor the target of emotions for our concurrent emotion measure. Future research could focus on interrelations of emotions and more modalities simultaneously. Likewise, although we assessed a range of emotions using ESM, all assessed emotions had valence. However, previous research (e.g., Ahn & Harley, 2020; Baker et al., 2010; Harley et al., 2015) found that neutral states dominated prolonged interactions with ITS. Our study design enabled learners to report they were not experiencing the assessed emotions (boredom, enjoyment, confusion, frustration) to any degree ("not at all") but could not specifically indicate a neutral emotional experience. Due to the frequent repeated nature of the ESM the focus was on a subset of emotional states which did not include neutral, despite this state being domiant (Baker et al., 2010). Future studies should thus explore the role of neutral affective states on prompt compliance, and domain knowledge after learning as well as relations to affective activation. Further, longer-lasting interventions may also be able to help improve understanding affective dynamics over longer periods of time. For example, future studies could model three-level DSEMs to account for situational data nested within students in lessons.

It should also be noted that although ESM approaches can generally be considered reliable their main fallacy is that they rely on participants' response adherence (see also: Csikszentmihalyi & Larson, 2014), i.e., ESM relies on self-reports. Self-reports can be biased through, for example, the social desirability bias (Nederhof, 1985) or the systematic response bias (Baumgartner & Steenkamp, 2001). Previous research has further shown that self-reports may be especially biased among learners with higher emotion regulation (see also: Ciuk et al., 2015) which may impede the validity of ESM data of emotions in those learners with higher emotion regulation. Further, adolescents' cognition is not fully developed yet and its development aims to achieve a conscious mind including conscious emotion processing and regulation, among other things (see also: Steinberg, 2005). Thus, students may not always be able to adequately and consciously assess (and report) their emotional experiences.

8.7. Future directions

Although therefore ESM data may be biased, assessing emotions by means of self-reports is a well-established method (Pekrun & Bühner, 2014) with benefits such as practicality and the opportunity to assess different components of emotions (Pekrun, 2016). Further, ESM sampling intervals may be too coarse to capture transient emotional episodes. Previous well-cited research (Harley et al., 2015) has employed longer intervals of 14 min suggesting that a 10-min interval may not be considered as too coarse in general but may still be too coarse to capture transient emotional states. One solution to capturing transient emotions could be shortened sampling intervals. However, as any additional (secondary) task increases cognitive demand (see also: Brunken and Leutner, 2003), more frequent ESM measurements could impose too high demands on learners' cognition and thus impede how well they can process information and thus impede their learning (see also: de Jong, 2010; Kirschner, 2002; Paas & Van Merriënboer, 1994; Sweller et al., 1998). Thus, too frequent ESM sampling may contaminate data associated with learning by distracting learners or even lead to measuring emotions associated with answering the ESM prompts (e.g., frustration) rather than task-related emotions.

To increase validity of such measurements without disrupting learning, research could add additional (external) measurement modalities: For example, future studies could employ external validation methods like the Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP; Ocumpaugh et al., 2015) or Quick Red Fox (Hutt et al., 2022); alternatively, automatic facial expression software or video recordings with in-situ or post-hoc (automatic) emotion detection could be utilized. Both observer protocols and video codings could aid in verifying the ESM self-report data as well as reducing data granularity. However, two things should be noted regarding observer protocols and video recordings: Video recordings open ethical questions especially in terms of data protection as well as school policies and further both approaches may miss or mislabel emotional states as learners may not always express emotions through mimicry (e.g., Hawthorne effect).

On another note, using physiological data such as EDA can also aid in capturing affective information and also help verify data capturing discrete emotions, particularly, differentiating emotions along the activation/deactivation axes emotions are widely conceptualized on (in addition to valence). Unlike ESM, increasing sampling rate of such data may not induce additional distraction or strain on the learners while learning as EDA is measured using electrodes attached to the learners' skin. Sampling is automatized and does not require any specific action by the learner. EDA provides a continuous account of emotional arousal, whereas ESM provides rapid snapshots of emotional states, and retrospective emotions self-reports offer summaries; therefore, EDA may offer a more complete picture of one component of emotional experiences (affective activation) during learning that retrospective summaries or rapid snapshot captured by retrospective self-report or ESM. One downside, however, is that that EDA data tends to be noisy and requires appropriate processing to retrieve information relevant to the research being done.

However, EDA, for now, by itself cannot capture discrete emotional states (e.g., enjoyment vs. anxiety) or transitions between them and remains a measure of affective activation as a distinct expression component of emotional experiences. Future studies could triangulate modalities like self-report data, ESM, EDA, and observer protocols or video recordings. Such triangulation could aid support more comprehensive assessments of affective states and their different components as well as increase measurement validity like other studies have done (Harley et al., 2015). Other approaches such as assessing affective engagement by means of posture (D'Mello et al., 2009) or emotional valence by means of electromyography (Sato et al., 2020) may also be promising options for gaining a more comprehensive perspective. Overall, the assessment of emotions, especially as multicomponent phenomena, remains a complex task.

It may be possible that affective activation does not always impede learning but under certain circumstances (e.g., after a learning task; through enhanced interest). For example, Nielson and Powless (2007) found that the induction of arousal (positively or negatively valanced) within 30 min after a learning task, improved memory retrieval of the learning content one week later. Similarly, Ozcelik and Arslan-Ari (2024) found that arousal did improve transferable knowledge directly as well as indirectly through increased interest. Other findings indicate that instruction which fostered performance was also related to higher arousal although no direct links between arousal and performance were shown (Hoogerheide et al., 2019). Accordingly, more research is needed, including with other ITSs, to better understand the role of affective activation in fostering or impeding learning success. No single data assessment and no type of measurement is infallible, especially outside of lab settings. Regardless, multimodal approaches are important to increase ecological validity and gain a fuller picture of learning and learning processes – particularly in classroom settings. One issue researchers employing multimodal assessment methods are often confronted with is the different granularity of the data collected. Data with different granularity requires alignment and therefore often aggregation, decreasing the detail of measurements. Further, when aligning continuous and interval data, blind spots where the interval data does not cover the continuous data are to be expected. For example, a brief spike in physiological arousal or a sudden change in behavior may be signs of, for example, a moment of stress that occurs between scheduled ESM prompts and thus remains uncaptured in self-reports. Additionally, multimodal approaches often rely on technology which may be sensitive to external or learner-induced disturbances and therefore technical faults outside of highly controlled and supervised lab settings. For example, learners may subconsciously fidget with the EDA electrodes de- and re-attaching them without research staff noticing, potentially compromising data quality.

Thus, data (or information) loss when employing multimodal assessment at the current time in ecological classroom settings may be an inherent, or at least very challenging to completely mitigate, aspect of multimodal data collection. Therefore, such drawbacks as well es adverse effects on data quality need to be considered in multimodal studies—particularly in ecological classrooms. However, the question of whether meticulously clean data of ecological classroom settings is always necessary arises as classrooms and classroom data will always be noisy (see discussion by D'Mello et al., 2018). Similar to teachers having to read, interpret, and react to noisy information in the classroom, investigating noisy and multimodal data may also be understood as a necessary approach to helping to achieve stronger ecologically valid data, results, and interpretations.

Our study cannot answer why or when learners comply with scaffolding prompts and in how far this may benefit learning, given that overall prompt compliance was rather low in the present study, learners may have perceived the offered prompts as unhelpful or did not comprehend them and thus dismissed the offered scaffolding prompts. Future research should focus on optimizing scaffolding by exploring the specific conditions under which prompt delivery and design facilitate learning. This approach may enhance the perceived usefulness of prompts among learners, especially in light of our findings that emotional adaptivity alone may be insufficient to improve learning outcomes for everyone. For example, future studies could include student-rated utility of scaffolding prompts into their adaptive model to optimize prompt design (e.g., different degree of language complexity) or timing of delivery. Further, future ITS may offer scaffolding prompts with lowered linguistic demands or vary means of delivery (e.g., orally verbalized scaffolding prompts rather than written text-based scaffolding prompts.

8.8. Conclusion

The present study underpins the relevance of emotions for learning processes in new technologies applied in real classroom settings using a multimodal dynamic approach. Especially experiencing topic-related boredom and confusion before starting a learning task were detrimental for learning performance (i.e., domain knowledge after learning) and increased affective activation. Although prior work (D'Mello & Graesser, 2012) suggests affective activation as a driver for deep learning, we found that high affective activation significantly reduced domain knowledge after learning and appears to have been driven by negative experiences in possible over-challenge scenarios. Especially initial topic-related boredom seems to be highly persistent in so far that it translates to higher levels of concurrent boredom while learning. Therefore, teachers should prevent or help students overcome topic-related boredom before starting a task which could lead to unproductive, boredom-related emotional activation. Further, affective activation as a physiological component of emotion was distinctively negatively related to domain knowledge after learning, highlighting the importance of considering distinct components of emotion in learning situations. By integrating ESM and particularly EDA data in an ecologically valid setting, the present study contributes to a more differentiated understanding of affective dynamics and their role in learning. Our findings emphasize the importance of affective activation and initial discrete emotions as those may shape affective dynamics and benefit or hinder learning performance (e.g, domain knowledge after learning).

Although prompt compliance was self-predictive on the situation level, it was unrelated to any study variables. Given ambiguous findings on both prompt compliance and its effectiveness for learning outcomes, future research should focus on understanding what constitutes effective prompts and effective prompt delivery to reduce the risk of ineffective learning caused by negative experiences of overchallenge related to confusion and boredom.

CRediT authorship contribution statement

Anja Henke: Writing – original draft, Visualization, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jason M. Harley:** Writing – review & editing, Methodology, Conceptualization. **Negar Matin:** Writing – review & editing, Resources, Data curation. **Johann Chevalère:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Verena V. Hafner:** Methodology, Funding acquisition. **Niels Pinkwart:** Methodology, Funding acquisition. **Rebecca Lazarides:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization.

Ethical statement

This study received approval from the University of Potsdam Ethics Committee (number 82/2020) and from the Brandenburg Ministry of Education, Youth and Sports (number 82-E1-2020). The study complies with APA ethical standards in the treatment of the human participants.

Funding statement

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 “Science of Intelligence” - project number 390523135. The funding source was not involved in the study design; in the collection, analysis, or interpretation of the data; in the writing of the report; or in the decision to submit the article for publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2025.102310.

Data availability

Data will be made available on request.

References

Ahn, B. T., & Harley, J. M. (2020). Facial expressions when learning with a queer history app: Application of the control value theory of achievement emotions. *British Journal of Educational Technology, 51*(5), 1563–1576. https://doi.org/10.1111/bjet.12989

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Baker, R. S. J.d., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223–241. https://doi.org/10.1016/j.ijhcs.2009.12.003

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national onvestigation. *Journal of Marketing Research, 38*(2), 143–156. https://doi.org/10.1509/jmkr.38.2.143.18840

Bench, S. W., & Lench, H. C. (2013). On the function of boredom. *Behavioral Sciences, 3*(3), 459–472. https://www.mdpi.com/2076-328X/3/3/459.

Benedek, M., & Kaernbach, C. (2006). *Ledalab*.

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Berweger, B., Born, S., & Dietrich, J. (2022). Expectancy-value appraisals and achievement emotions in an online learning environment: Within- and between-person relationships. *Learning and Instruction, 77*, Article 101546. https://doi.org/10.1016/j.learninstruc.2021.101546

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From design to implementation to practice a learning by teaching system: Betty's brain. *International Journal of Artificial Intelligence in Education, 26*(1), 350–364. https://doi.org/10.1007/s40593-015-0057-9

Brose, A., Voelkle, M. C., Lövdén, M., Lindenberger, U., & Schmiedek, F. (2015). Differences in the between–person and within–person structures of affect are a matter of degree. *European Journal of Personality, 29*(1), 55–71. https://doi.org/10.1002/per.1961

Brunken, R. L. P. J., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53–61. https://doi.org/10.1207/S15326985EP3801_7

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

Chevalére, J., Lazarides, R., Yun, H. S., Henke, A., Lazarides, C., Pinkwart, N., & Hafner, V. V. (2023). Do instructional strategies considering activity emotions reduce students' boredom in a comupterized open-ended learning environment? *Computer & Education, 196*, 1–15. https://doi.org/10.1016/j.compedu.2023.104741

Ciuk, D., Troy, A., & Jones, M. (2015). Measuring emotion: Self-reports vs. physiological indicators. *SSRN*.

Clarebout, G., & Elen, J. (2006). Tool use in computer-based learning environments: Towards a research framework. *Computers in Human Behavior, 22*(3), 389–411. https://doi.org/10.1016/j.chb.2004.09.007

Cohen, B. (1988). *Statistical power analysis for the behavioral sciences* (3 ed.). John Wiley & Sons.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology: The collected works of mihaly csikszentmihalyi* (pp. 35–54). Netherlands: Springer. https://doi.org/10.1007/978-94-017-9088-8_3.

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology, 51*(5), 1441–1449. https://doi.org/10.1111/bjet.13015

D'Mello, S. K. (2020). Multimodal analytics for automated assessment. In D. Yan, A. A. Rupp, & P. W. Foltz (Eds.), *Handbook of automated scoring: Theory into practice* (pp. 94–114). Chapman and Hall/CRC.

D'Mello, S. K., Bosch, N., & Chen, H. (2018). Multimodal-multisensor affect detection. *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition - Volume 2, 21*, 167–202. https://doi.org/10.1145/3107990.3107998. Association for Computing Machinery and Morgan & Claypool.

D'Mello, S. K., Chipman, P., & Graesser, A. (2009). Posture as a predictor of learner's affective engagement. *Proceedings of the Annual Meeting of the Cognitive Science Society, 29*. https://escholarship.org/uc/item/7hs9v2hr.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0

Dindar, M., Malmberg, J., Järvelä, S., Haataja, E., & Kirschner, P. A. (2020). Matching self-reports with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies, 25*(3), 1785–1802. https://doi.org/10.1007/s10639-019-10059-5

D'Mello, S., & Graesser, A. (2010). Modeling cognitive-affective dynamics with hidden markov models. *Proceedings of the 32nd annual cognitive science society*.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fowles, D. C. (2008). The measurement of electrodermal activity in children. In *Developmental psychophysiology: Theory, systems, and methods* (pp. 286–316). Cambridge University Press.

Frenzel, A. C., Goetz, T., & Stockinger, K. (2024). Emotions and emotion regulation. In *Handbook of educational psychology* (4th ed., pp. 219–244). Routledge.

Gabriel, A. S., Podsakoff, N. P., Beal, D. J., Scott, B. A., Sonnentag, S., Trougakos, J. P., & Butts, M. M. (2019). Experience sampling methods: A discussion of critical trends and considerations for scholarly advancement. *Organizational Research Methods, 22*(4), 969–1006. https://doi.org/10.1177/1094428118802626

Ganapathy, N., Veeranki, Y. R., Kumar, H., & Swaminathan, R. (2021). Emotion recognition using electrodermal activity signals and multiscale deep convolutional neural network. *Journal of Medical Systems, 45*(4), 49. https://doi.org/10.1007/s10916-020-01676-6

Gardner, P., & Kerridge, K. (2019). Everybody present: Exploring the use of an in-class meditation intervention to promote positive mental health among university students. *Canadian Journal of Community Mental Health, 38*(1), 9–21. https://doi.org/10.7870/cjcmh-2018-022

Goetz, T., Frenzel, A. C., Hall, N. C., Nett, U. E., Pekrun, R., & Lipnevich, A. A. (2014). Types of boredom: An experience sampling approach. *Motivation and Emotion, 38*(3), 401–419. https://doi.org/10.1007/s11031-013-9385-y

Goetz, T., Frenzel, A. C., Stoeger, H., & Hall, N. C. (2010). Antecedents of everyday positive emotions: An experience sampling analysis. *Motivation and Emotion, 34*(1), 49–62. https://doi.org/10.1007/s11031-009-9152-2

Goetz, T., Stempfer, L., Pekrun, R., van Tilburg, W. A. P., & Lipnevich, A. A. (2024). Academic boredom. In M. Bieleke, W. Wolff, & C. S. Martarelli (Eds.), *The routledge international handbook of boredom* (pp. 225–249). Routledge. https://doi.org/10.4324/9781003271536-18.

Goetz, T., Sticca, F., Pekrun, R., Murayama, K., & Elliot, A. J. (2016). Intraindividual relations between achievement goals and discrete achievement emotions: An experience sampling approach. *Learning and Instruction, 41*, 115–125. https://doi.org/10.1016/j.learninstruc.2015.10.007

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

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

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features. *International Journal of Artificial Intelligence in Education, 27*(2), 268–297. https://doi.org/10.1007/s40593-016-0126-8

Harley, J. M., Lajoie, S. P., Tressel, T., & Jarrell, A. (2020). Fostering positive emotions and history knowledge with location-based augmented reality and tour-guide prompts. *Learning and Instruction, 70*, Article 101163. j.learninstruc.2018.09.001.

Harley, J. M., Liu, Y., Ahn, T. B., Lajoie, S. P., Grace, A. P., Haldane, C., Whittaker, A., & McLaughlin, B. (2019). I've got this: Fostering topic and technology-related emotional engagement and queer history knowledge with a mobile app. *Contemporary Educational Psychology, 59*, Article 101790. https://doi.org/10.1016/j.cedpsych.2019.101790

Harley, J. M., & Pekrun, R. (2024). Managing emotions in education: The emotion regulation in achievement situations model. In *Handbook of emotion regulation* (3rd ed., pp. 536–544). The Guilford Press.

Harley, J. M., Pekrun, R., Taxer, J. L., & Gross, J. J. (2019). Emotion regulation in achievement situations: An integrated model. *Educational Psychologist, 54*(2), 106–126. https://doi.org/10.1080/00461520.2019.1587297

Harley, J. M., Taub, M., Azevedo, R., & Bouchet, F. (2018). Let's set up some subgoals: Understanding human-pedagogical agent collaborations and their implications for learning and prompt and feedback compliance. *IEEE Transactions on Learning Technologies, 11*(1), 54–66. https://doi.org/10.1109/TLT.2017.2756629

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., Molenaar, I., Van Der West, H., Bosse, T., & Lazonder, A. W. (2024). Multimodal measurements enhance insights into emotional responses to immediate feedback [Original Research]. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1294386

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), 7869. https://www.mdpi.com/1424-8220/21/23/7869.

Hung, H.-T., Chiu, Y.-C. J., & Yeh, H.-C. (2013). Multimodal assessment of and for learning: A theory-driven design rubric. *British Journal of Educational Technology, 44*(3), 400–409. https://doi.org/10.1111/j.1467-8535.2012.01337.x

Hutt, S., Baker, R. S., Ocumpaugh, J., Munshi, A., Andres, J. M. A. L., Karumbaiah, S., Slater, S., Biswas, G., Paquette, L., Bosch, N., & van Velsen, M. (2022). Quick red fox: An app supporting a new paradigm in qualitative research on AIED for STEM. In F. Ouyang, P. Jiao, B. M. McLaren, & A. H. Alavi (Eds.), *Artificial intelligence in STEM education: The paradigmatic shifts in research, education, and technology* (pp. 319–332). CRC Press.

Jewitt, C., Gunther, K., Jon, O., & Tsatsarelis, C. (2001). Exploring learning through visual, actional and linguistic communication: The multimodal environment of a science classroom. *Educational Review, 53*(1), 5–18. https://doi.org/10.1080/00131910123753

Kardan, S., & Conati, C. (2015). Providing adaptive support in an interactive simulation for learning. *An experimental evaluation proceedings of the 33rd annual ACM conference on human factors in computing systems*. https://doi.org/10.1145/2702123.2702424. Seoul, Republic of Korea.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*(1), 1–10. https://doi.org/10.1016/S0959-4752(01)00014-7

Kołodziej, M., Tarnowski, P., Majkowski, A., & Rak, R. J. (2019). Electrodermal activity measurements for detection of emotional arousal. *Bulletin of the Polish Academy of Sciences, Technical Sciences, 67*(4), 813–826. https://doi.org/10.24425/bpasts.2019.130190

Lajoie, S. P., Pekrun, R., Azevedo, R., & Leighton, J. P. (2020). Understanding and measuring emotions in technology-rich learning environments. *Learning and Instruction, 70*, Article 101272. https://doi.org/10.1016/j.learninstruc.2019.101272

Malmberg, J., Fincham, O., Pijeira-Díaz, H. J., Järvelä, S., & Gašević, D. (2021). Revealing the hidden structure of physiological states during metacognitive monitoring in collaborative learning. *Journal of Computer Assisted Learning, 37*(3), 861–874. https://doi.org/10.1111/jcal.12529

Maniktala, M., Cody, C., Barnes, T., & Chi, M. (2020). Avoiding help avoidance: Using interface design changes to promote unsolicited hint usage in an intelligent tutor. *International Journal of Artificial Intelligence in Education, 30*(4), 637–667. https://doi.org/10.1007/s40593-020-00213-3

Martinez, R., Irigoyen, E., Arruti, A., Martin, J. I., & Muguerza, J. (2017). A real-time stress classification system based on arousal analysis of the nervous system by an F-state machine. *Computer Methods and Programs in Biomedicine, 148*, 81–90. https://doi.org/10.1016/j.cmpb.2017.06.010

Matin, N., Törmänen, T., Ahn, B., Johnson, M., Sun, N.-Z., & Harley, J. M. (2024). *Physiological arousal in anti-harassment simulations: Addressing harassment in medical education international conference on motivation*. Switzerland: Bern.

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in mplus. *Psychological Methods, 25*(5), 610–635. https://doi.org/10.1037/met0000250

Meinhardt, J., & Pekrun, R. (2003). Attentional resource allocation to emotional events: An ERP study. *Cognition & Emotion, 17*(3), 477–500. https://doi.org/10.1080/02699930244000039

Mercado-Diaz, L. R., Veeranki, Y. R., Marmolejo-Ramos, F., & Posada-Quintero, H. F. (2024). EDA-Graph: Graph signal processing of electrodermal activity for emotional states detection. *IEEE Journal of Biomedical and Health Informatics, 28*(8), 4599–4612. https://doi.org/10.1109/JBHI.2024.3405491

Moeller, J., Dietrich, J., & Baars, J. (2023). The experience sampling method in the research on achievement-aelated emotions and motivation: Theoretical and methodological perspectives and empirical insights. In G. Hagenauer, R. Lazarides, & H. Järvenoja (Eds.), *Motivation and emotion in learning and teaching across educational contexts* (pp. 178–196). Routledge.

Moreno, M., Melo, L. P., Grewal, K., Matin, N., Azher, S., & Harley, J. M. (2024). Analyzing multimodal data to understand medical trainees' regulation strategies and physiological responses in high-fidelity medical simulation scenarios. *Metacognition and Learning, 19*(3), 1161–1213. https://doi.org/10.1007/s11409-024-09403-z

Munshi, A., Biswas, G., Baker, R., Ocumpaugh, J., Hutt, S., & Paquette, L. (2023). Analysing adaptive scaffolds that help students develop self-regulated learning behaviours. *Journal of Computer Assisted Learning, 39*(2), 351–368. https://doi.org/10.1111/jcal.12761

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018). Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments. *Proceedings of the 26th conference on user modeling, adaptation and personalization*. https://doi.org/10.1145/3209219.3209241. Singapore, Singapore.

Murray, A. L., Brown, R., Zhu, X., Speyer, L. G., Yang, Y., Xiao, Z., Ribeaud, D., & Eisner, M. (2023). Prompt-level predictors of compliance in an ecological momentary assessment study of young adults' mental health. *Journal of Affective Disorders, 322*, 125–131. https://doi.org/10.1016/j.jad.2022.11.014

Muthén, B., & Muthén, L. (1998). *Mplus version 8.10*.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3), 263–280. https://doi.org/10.1002/ejsp.2420150303

Nielson, K. A., & Powless, M. (2007). Positive and negative sources of emotional arousal enhance long-term word-list retention when induced as long as 30min after learning. *Neurobiology of Learning and Memory, 88*(1), 40–47. https://doi.org/10.1016/j.nlm.2007.03.005

Niven, K., & Miles, E. (2013). Affect arousal. In M. D. Gellman, & J. R. Turner (Eds.), *Encyclopedia of behavioral medicine* (pp. 50–52). New York: Springer. https://doi.org/10.1007/978-1-4419-1005-9_1089.

Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. Routledge. https://doi.org/10.4324/9780203379493

Novak, J., & Gowin, B. (1984). *Learning how to learn*. Cambridge University Press.

Ocumpaugh, J., Baker, R. S., & Rodrigo, M. M. T. (2015). *Baker rodrigo ocumpaugh monitoring protocol (BROMP) 2.0 technical and training manual*.

Ozcelik, E., & Arslan-Ari, I. (2024). Enhancing multimedia learning by emotional arousal. *The Journal of Experimental Education, 92*(2), 247–261. https://doi.org/10.1080/00220973.2023.2182263

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351–371. https://doi.org/10.1007/BF02213420

Paquette, L., Grant, T., Zhang, Y., Biswas, G., & Baker, R. (2021). Using epistemic networks to analyze self-regulated learning in an open-ended problem-solving environmentA. R. Ruis, & S. B. Lee (Eds.). *Advances in quantitative ethnography Cham*.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2016). Using self-report to assess emotions in education. In M. Zembylas, & P. A. Schutz (Eds.), *Methodological advances in research on emotion and education* (pp. 43–54). Springer International Publishing. https://doi.org/10.1007/978-3-319-29049-2_4.

Pekrun, R. (2019). Achievement emotions: A control-value theory perspective. In R. Patulny, A. Bellocchi, R. Olson, S. Khorana, J. McKenzie, & M. Peterie (Eds.), *Emotions in late modernity* (pp. 142–157). Routledge.

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

Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotions. In *International handbook of emotions in education* (pp. 561–579). Routledge/Taylor & Francis Group.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz, & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). Academic Press. https://doi.org/10.1016/B978-012372545-5/50003-4.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., & Marsh, H. W. (2022). Research on situated motivation and emotion: Progress and open problems. *Learning and Instruction, 81*, Article 101664. https://doi.org/10.1016/j.learninstruc.2022.101664

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., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The epistemically-related emotion scales. *Cognition & Emotion, 31*(6), 1268–1276. https://doi.org/10.1080/02699931.2016.1204989

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*(2), 267–280. https://doi.org/10.1016/j.learninstruc.2010.07.004

Sato, W., Kochiyama, T., & Yoshikawa, S. (2020). Physiological correlates of subjective emotional valence and arousal dynamics while viewing films. *Biological Psychology, 157*, Article 107974. https://doi.org/10.1016/j.biopsycho.2020.107974

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion, 23*(7), 1307–1351. https://doi.org/10.1080/02699930902928969

Schweder, S., & Raufelder, D. (2019). Positive emotions, learning behavior and teacher support in self-directed learning during adolescence: Do age and gender matter? *Journal of Adolescence, 73*, 73–84. https://doi.org/10.1016/j.adolescence.2019.04.004

Segedy, J. R. (2014). *Adaptive scaffolds in open-ended computer-based learning environments [doctoral dissertation*. Nashville, Tennessee: Vanderbilt University].

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal of Learning Analytics, 2*(1), 13–48. https://doi.org/10.18608/jla.2015.21.3

Sequeira, H., Hot, P., Silvert, L., & Delplanque, S. (2009). Electrical autonomic correlates of emotion. *International Journal of Psychophysiology, 71*(1), 50–56. https://doi.org/10.1016/j.ijpsycho.2008.07.009

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell Us about learning? *British Journal of Educational Technology, 51*(5), 1450–1484. https://doi.org/10.1111/bjet.12993

Sokolovsky, A. W., Mermelstein, R. J., & Hedeker, D. (2013). Factors predicting compliance to ecological momentary assessment among adolescent smokers. *Nicotine & Tobacco Research, 16*(3), 351–358. https://doi.org/10.1093/ntr/ntt154

Somerville, M. P., Midouhas, E., Delprato, M., & Whitebread, D. (2024). Help me if you can i'm feeling down: Supporting children's emotion regulation and well-being in the primary classroom. *School Mental Health, 16*(2), 577–591. https://doi.org/10.1007/s12310-024-09668-4

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences, 9*(2), 69–74. https://doi.org/10.1016/j.tics.2004.12.005

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Tze, V. M. C., Daniels, L. M., & Klassen, R. M. (2016). Evaluating the relationship between boredom and academic outcomes: A meta-analysis. *Educational Psychology Review, 28*(1), 119–144. https://doi.org/10.1007/s10648-015-9301-y

Veeranki, Y. R., Diaz, L. R. M., Swaminathan, R., & Posada-Quintero, H. F. (2024). Nonlinear aignal processing methods for automatic emotion recognition using electrodermal activity. *IEEE Sensors Journal, 24*(6), 8079–8093. https://doi.org/10.1109/JSEN.2024.3354553

Vilhunen, E., Turkkila, M., Lavonen, J., Salmela-Aro, K., & Juuti, K. (2022). Clarifying therelation between epistemic emotions and learning by using experience sampling method and pre-posttest design [Original Research]. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.826852

Vogl, E., Pekrun, R., Murayama, K., & Loderer, K. (2020). Surprised–curious–confused: Epistemic emotions and knowledge exploration. *Emotion, 20*(4), 625–641. https://doi.org/10.1037/emo0000578

Anja Henke

Anja Henke investigates adaptive technologies and robots in real-life social learning contexts.

Prof. Dr. Jason M. Harley

Jason Harley researches emotions in medical learning situations using AI, VR, and simulation.

Negar Matin

Negar Matin investigates emotions using physiological signals.

Jun.-Prof. Dr. Johann Chevalère

Johann Chevalère researches learning, memory, executive functions, emotions and motivation in technology-enhanced education.

Prof. Dr. Verena V. Hafner

Verena Hafner investigates sensorimotor interaction and learning, joint attention, predictive models, and exploration strategies.

Prof. Dr. Niels Pinkwart

Niels Pinkwart researches educational technology and AI-supported learning.

Prof. Dr. Rebecca Lazarides

Rebecca Lazarides investigates motivational-affective learning processes and technologies for learning and instruction.