Liu, X., Ye, J., & Zhang. Y.(2023). Multimodal neurophysiological representations of high school students' situational interest: A machine learning approach. *Journal of Educational Technology Development and Exchange*, *16*(2), 108-125. https://doi.org/10.18785/jetde.1602.7

Multimodal Neurophysiological Representations of High School Students' Situational Interest: A Machine Learning Approach

Xiaobo Liu
Tsinghua University, Beijing

Jingmeiqi Ye
Beijing No.4 High School, Beijing

Yu Zhang *
Tsinghua University, Beijing
zhangyu2011@tsinghua.edu.cn

Abstract: *Interest plays a vital role in students' learning performance. Accurately measuring situational interest in the classroom environment is important for understanding the learning mechanism and improving teaching. However, self-report measurements frequently encounter issues of subjectivity and ambiguity, and it is hard to collect dynamic self-report scales without disturbance in the naturalistic environment. Thanks to the development of neuroscience and portable biosensors, it has become possible to represent psychological states with neurophysiological features in the classroom environment. In this study, multimodal neurophysiological signals, including electroencephalograph (EEG), electrodermal activity (EDA), and photoplethysmography (PPG), were applied to represent situational interest under both laboratory (Study 1) and naturalistic (Study 2) paradigms. A total of 33 features were extracted, and 7 different statistical indicators were calculated for each of them across all the epochs. Among these features, 47 in Study 1 and 49 in Study 2 demonstrated significant correlation with self-report situational interest. Employing a machine learning model, the analysis yielded a mean absolute error (MAE) of 0.772 and mean squared error (MSE) of 0.883 for the dataset in Study 1. However, the model was not robust on data from Study 2. These findings offer empirical support for the conceptual framework of situational interest, demonstrate the potential of neurophysiological data in educational assessments, and also highlight the challenges in naturalistic paradigm.*

Keywords: situational interest, neurophysiological representation, naturalistic paradigm, machine learning

1.Introduction

Interest is a critical factor in student learning (Harp & Mayer, 1997; Renninger, 1992). It not only enhances performance by directly fostering higher engagement, but influences the development of intrinsic motivation as well, thereby facilitating the long-term learning (Renninger, 2000; Renninger & Hidi, 2015). The concept of interest is broadly categorized into two types: situational interest and individual interest, representing state and trait components, respectively (Hidi, 2000; Schraw et al., 2001). Interest develops and becomes internalized through a four-stage model (Hidi & Renninger, 2006). It begins with triggered situational interest, progresses to maintained situational interest, then to emerging individual interest, and finally, to well-developed individual interest. The classroom serves as a primary setting for both learning and developing students' academic interest. With insights into students' situational interest, teachers can better design their lessons to improve the learning experience (Rotgans & Schmidt, 2011; Tsai et al., 2008). Similarly, if students have a more accurate and objective understanding of their academic interest, they can plan their career with less confusion.

The effectiveness of these applications hinges on the accurate measurement of interest. Similar to many psychological concepts, situational interest is usually measured through self-report scales. For instance, Wang and Adesope (2016) developed a scale to measure the four stages of interest as outlined in Hidi's model. However, Azevedo (2018) characterized interest as a short-term spike, which may not align with the longer duration of typical lessons. Therefore, post-stimuli scales are challenged to capture dynamic changes without interrupting the learning process. To effectively measure situational interest, a real-time, minimally invasive tool is needed.

Thanks to the development of portable biosensors, neurophysiological data have demonstrated its potential in assessing students' situational interest. Laboratory experiments have shown that neurophysiological features can effectively represent basic affective and cognitive states (Ayres et al., 2021; Wang et al., 2022). In real-world scenarios, these associations have also been validated (Shui et al., 2021), particularly in the real-classroom learning (Chen et al., 2023; Dikker et al., 2017; Zhang et al., 2018). In relevant theories, situational interest is seen as inducing and accompanied by a range of fundamental affective and cognitive aspects, such as attention, positive affect, and cognitive processes (Ainley & Hidi, 2014; Chen et al., 2001; Hidi, 2006; Hidi & Renninger, 2019). Consequently, situational interest could be linked with features that are related to these basic concepts. For instance, electroencephalograph (EEG) band power, known as indicators of attention (Liu et al., 2013); electrodermal activity (EDA) features, often used to predict emotions (Picard et al., 2001); and photoplethysmography (PPG) features, indexing cognitive load (Lyu et al., 2015), are likely associated with situational interest. Building on this, recent studies (Babiker et al., 2019; Tan et al., 2021) have explored neurophysiological features that may directly indicate situational interest.

Beyond single-modality analyses, multimodal fusion has demonstrated better predictive performance (Poria et al., 2017). Given the composite nature of situational interest, effectively combining various types of neurophysiological features in the prediction model is suitable. Employing machine learning techniques, researchers like Gao et al. (2020) have successfully used multimodal fusion to predict student engagement in classrooms. Thus, employing a machine learning model for representing situational interest has emerged as a promising method.

The research questions of this study are as follows: Can situational interest be represented with multimodal neurophysiological data in the classroom setting? Can machine learning approaches effectively enhance multimodal fusion for this task? To answer the above questions, portable devices including head belts and wristwatches were employed to collect students' EEG, EDA, and PPG signals throughout their learning sessions. Both video stimuli and real-world lessons were incorporated to trigger students' situational interest. To the best of our knowledge, this study is among the first to explore the neurophysiological representation of situational interest in real-world scenarios. Furthermore, this study demonstrates the potential of portable biosensors in assessing the dynamics of interest and its development, highlighting their utility in educational research.

2. Methods

2.1 Data collection

This study employed a dataset collected from Grade 10 students in a typical high school in Beijing. To represent situational interest, multimodal neurophysiological data were collected in both a semi-controlled experiment and a naturalistic investigation, as shown in Figure 1. Randomized controlled experiments are widely accepted in neurophysiological studies for their effectiveness in inducing various psychological states (Dzedzickis et al., 2020; Zhou et al., 2021). However, in recent years, there has been a growing focus on naturalistic paradigms to improve the ecological validity of such research (Immordino-Yang & Gotlieb, 2017). Therefore, to improve the validity of the representation model, data were also collected during the real-world learning.

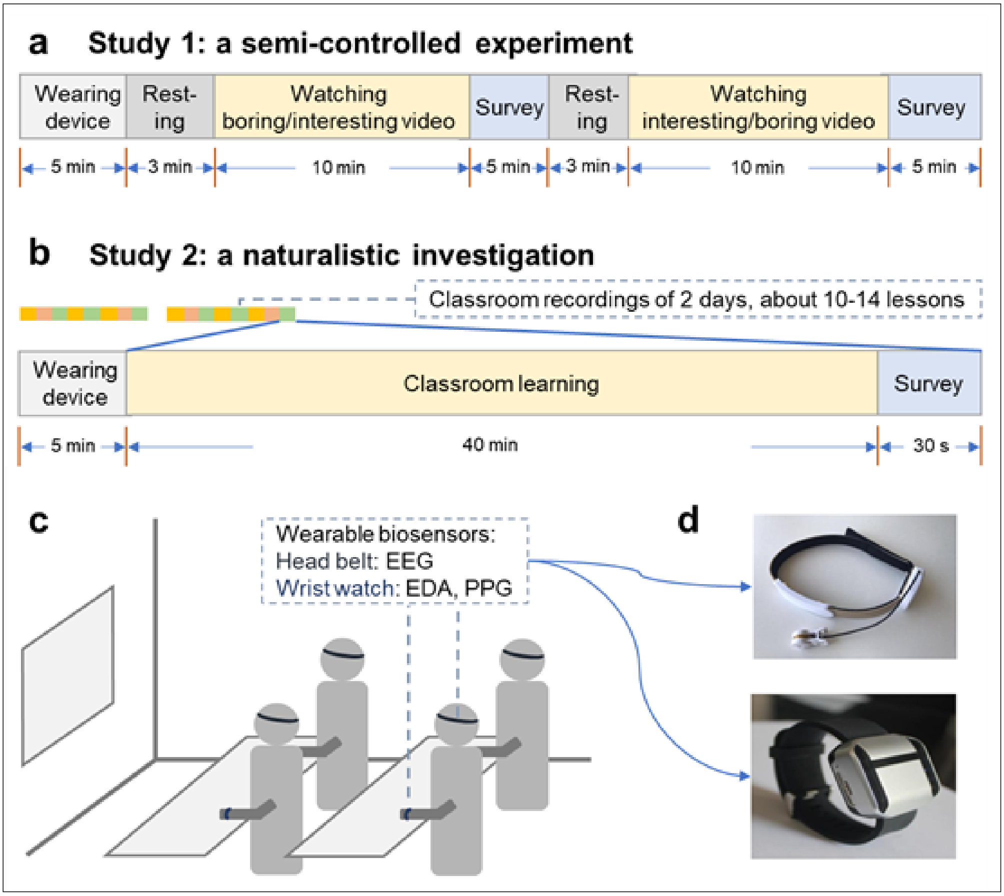

Figure 1

Research design

a b Timelines of Studies 1 and 2, which trigger students' situational interest with video stimuli and naturalistic lessons, respectively. **c** Neurophysiological data collection in the classroom setting. **d** Photos of portable devices, including a head belt and a wristwatch.

Participants. This study recruited 224 senior high school students, aged between 14 to 17 years. Eleven students withdrew from the experiment. Of the remaining 213 participants, there were 96 males, 107 females, and 10 participants who did not report their gender. All of these students participated in both Study 1 and Study 2. This study complied with Chinese laws and the Declaration of Helsinki and was approved by the Institutional Review Board (IRB) of Department of Psychology, Tsinghua University. All participants and their legal guardians read and signed an informed consent form. The experiments were conducted in November and December, 2020.

Apparatus and settings. Neurophysiological data were collected using head belts and wristwatches. The head belt, equipped with dry electrodes at Fp1 and Fp2 channels on the forehead, records EEG signals at a sampling rate of 250 Hz (Brainno, SOSO H&C, South Korea). This device has been previously employed in educational research to investigate disciplinary differences (Chen et al., 2023). The wristwatch gathers EDA signals at a sampling rate of 40 Hz and PPG signals at 20 Hz (Psychorus, China). Its efficacy and reliability have been validated in previous studies focusing on real-classroom environments (Liu et al., 2021; Zhang et al., 2021; Zhang et al., 2018).

Ground truth labels for situational interest. In Study 1, situational interest was measured by the scale developed by Wang and Adesope (2016), which measures the four-stage of interest according to the model proposed by Hidi and Renninger (2006). In this study, the items of triggered and maintained situational interest were added up as the overall situational interest. The scale's Cronbach's alpha was 0.960 and KMO was 0.938, implying high reliability and validity.

Given the tight schedule in the daily school time, the self-report measure of situational interest was simplified to a single question in Study 2:

Q. How did you feel about this lesson?

A. This class did not interest me at all.

B. It caught my interest once in a while, but it soon dissipated.

C. Sometimes, it interested me and lasted for a while.

D. It made me feel so interested that I wanted to continue listening.

To check the consistency of the Wang and Adesope (2016) scale and this single-item scale, a supplementary experiment using the same video stimuli from Study 1 were conducted on a different group of high school students. The result showed a high correlation between the long scale and this question (Pearson's correlation $r = 0.882$, $p < 0.001$, $n = 134$), indicating the validity of using the single-item scale to measure situational interest.

Procedure of Study 1: A semi-controlled experiment. As shown in Figure 1a, before the experiment, participants were guided to wear head belts and wristwatches, and the experimental procedure was briefly explained to them. The experimental process included watching two teaching videos: an interesting video and a boring one, which were evaluated by ten educational researchers regarding the level of interestingness before the experiment. The watching sequence of the two videos was randomized. Before each video was played, data of the resting state were collected (90 seconds of open-eye and 90 seconds of closed-eye resting, respectively). After the video session, situational interest was measured using the Wang and Adesope (2016) scale.

The interesting video used in the study was an 11-minute excerpt from a popular physics course, selected from a video website with extensive online learning resources. The boring video was an 8-minute segment focusing on physical education teaching theory, which was selected from an online learning platform. The two videos did trigger different levels of situational interest, as evidenced by the T-test results ($t = 23.9$, $p < 0.001$) of the scale.

Procedure of Study 2: A naturalistic investigation. As shown in Figure 1b, during the naturalistic investigation, data were collected during two school days, and recordings from about 10-14 class lessons were collected for each participant. Before the school day started every morning, participants wore the portable devices. At the end of each session, the researchers distributed the simplified scale to all participants, measuring self-report situational interest of the session.

2.2 Data Preprocessing

The EEG processing protocol referred to the work of Chen et al. (2023), which utilized the same portable device in similar classroom settings. The main challenges in EEG signal processing include addressing artifacts such as missing data, transient signals from lost contact, slow drifts, and ocular artifacts. Therefore, the preprocessing protocol was as follows. After identifying missing data, robust detrending (de Cheveigné & Arzounian, 2018) was applied, followed by bandpass filtering at 1 and 40 Hz, and ocular artifacts removal (Kanoga et al., 2019). Then the signal was divided into 30-second-long epochs, and those with value exceeding ± 150 μV were excluded.

For EDA data preprocessing, missing epochs were firstly determined and deleted. Then the data were downsampled to 10 Hz, and smoothed with filtering methods to remove noises (Zhang et al., 2021). Then the EDA data were decomposed into two components: the tonic component, known as skin conductance level (SCL); and the phasic component, known as skin conductance response (SCR), using the cvxEDA python toolbox (Greco et al., 2016).

The preprocessing of the PPG signal was conducted using the software developed by Psychorus, through which the heart rate was estimated. To ensure data quality, epochs exceeded the typical range (50 bpm – 100 bpm) were excluded.

Regarding missing data, for each sample, it was excluded from the analysis if there was either missing self-report situational interest data, or if all EEG and EDA epochs were excluded. The final dataset comprised 367 samples in Study 1 and 1849 samples in Study 2. Before the analysis, all the data was standardized, and the missing values were filled using mean imputation (Waljee et al., 2013).

2.3 Feature Extraction

To represent situational interest, features were extracted from various data modalities. These features, which are detailed in Table 1, are thought to be indirectly related to situational interest. As previously discussed, situational interest is associated with emotional and cognitive states, such as enjoyment, engagement, attention, and mental effort (Ainley & Hidi, 2014; Chen et al., 2001). Therefore, drawing upon these foundational states, a range of relative features were identified and extracted for analysis, as detailed below.

EEG features. EEG is a widely-applied technique in cognitive neuroscience for detecting brain activity. In this study, EEG signals were measured in the prefrontal area, corresponding to the prefrontal cortex, a crucial region for learning and cognitive processes. Band power in this area has been used in previous studies to represent attention (Liu et al., 2013), and cognitive control (Cavanagh & Frank, 2014). Therefore, this study involved calculating both absolute and relative powers for various EEG frequency bands: delta (1 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 13 Hz), beta (13 – 30 Hz), and gamma (30 – 40 Hz). Additionally, the ratios of EEG alpha to beta and theta to beta were extracted, based on the their relation to attention (Putman et al., 2014).

EDA and PPG features. EDA is a commonly-used physiological measure in affective computing (Greco et al., 2017), which has been employed to predict arousal (Ahuja et al., 2003), academic performance (Zhang et al., 2018), and engagement (Zhang et al., 2021) in daily scenarios. EDA signals were first decomposed into SCL (tonic component) and SCR (phasic component). SCL reflects the continuous, slow-changing aspect, while SCR indicates the fast-changing responses to stimuli (Roy et al., 2012). Then average values, standard deviation, first and second differences were calculated for both SCL and SCR components. For PPG data, heart rate was calculated, which is a metric commonly employed in educational research (AL-Ayash et al., 2016).

Synchrony features. In addition to calculating features at the individual level, there is an increasing focus on analyzing synchrony at the group level. When the students' physiological signals exhibit high inter-subject correlation with other students, their learning activities are assumed to be more aligned with others, and it is often described as "shared attention" at the group level (Dikker et al., 2017). High situational interest typically leads to increased engagement in class activities (Ainley, 2012), which may in turn influence group-level synchrony. For EEG signals, Total Independence (TI) was utilized to compute the synchrony feature (Dikker et al., 2017). Additionally, synchrony was assessed in specific power bands, namely delta, theta, alpha, low beta (13 – 18 Hz), high beta (18 – 30 Hz), and gamma bands (Chen et al., 2023). In addition to coherence measures, the Pearson's correlation of the signals with those of other students was also calculated. For EDA signals, synchrony was calculated using two methods: Pearson's correlation and dynamic time warping (DTW) distance, which were applied to both the SCL and SCR components. These features have been utilized in predicting student engagement in classroom settings (Gao et al., 2020).

Extracting time dynamics. Learning activities are dynamic and subject to fluctuations over time, as discussed by theoretical models and neurophysiological evidence (D'Mello & Graesser, 2012; Qu et al., 2020; Sung et al., 2023). To capture this variability, the study not only computed average and median values across epochs, but also included standard deviation (std), quartiles (Q1 and Q3), minimum, and maximum values, as depicted in Figure 2. Consequently, a total of 33 features, each analyzed with 7 different statistical measures on epochs, were extracted.

Table 1

Neurophysiological features representing situational interest

Feature type	Feature name	Description
EEG frequency domain features	delta	EEG absolute delta (1 – 4 Hz) spectral power
	delta_rel	EEG relative delta spectral power
	theta	EEG absolute theta (4 – 8 Hz) spectral power
	theta_rel	EEG relative theta spectral power
	alpha	EEG absolute alpha (8 – 13 Hz) spectral power
	alpha_rel	EEG relative alpha spectral power
	beta	EEG absolute beta (13 – 30 Hz) spectral power
	beta_rel	EEG relative beta spectral power
	gamma	EEG absolute gamma (30 – 40 Hz) spectral power
	gamma_rel	EEG relative gamma spectral power
	alpha_beta_ratio	Power ratio of alpha to beta
	theta_beta_ratio	Power ratio of theta to beta
EEG synchrony	TI	Total Independence (TI) of the EEG data with other students
	delta_TI	TI of the EEG data with other students at delta power band
	theta_TI	TI of the EEG data with other students at theta power band
	alpha_TI	TI of the EEG data with other students at alpha power band
	lowbeta_TI	TI of the EEG data with other students at low beta power band
	highbeta_TI	TI of the EEG data with other students at high beta power band
	gamma_TI	TI of the EEG data with other students at gamma power band
	correlation	Pearson's correlation coefficient of the EEG data with other students
EDA features	SCL_mean	Mean value of the SCL data
	SCL_std	Standard deviation of the SCL data
	SCL_delta	Mean of the first difference of the SCL data
	SCL_delta2	Mean of the second difference of the SCL data
	SCR_mean	Mean value of the SCR data
	SCR_std	Standard deviation of the SCR data
	SCR_delta	Mean of the first difference of the SCR data
	SCR_delta2	Mean of the second difference of the SCR data
EDA synchrony	SCL_correlation	Pearson's correlation of the SCL data with other students
	SCL_dtw	Dynamic time warping (DTW) distance of the SCL data with other students
	SCR_correlation	Pearson's correlation of the SCR data with other students
	SCR_dtw	DTW distance of the SCR data with other students
PPG feature	hr	Heart rate (beat per minute)

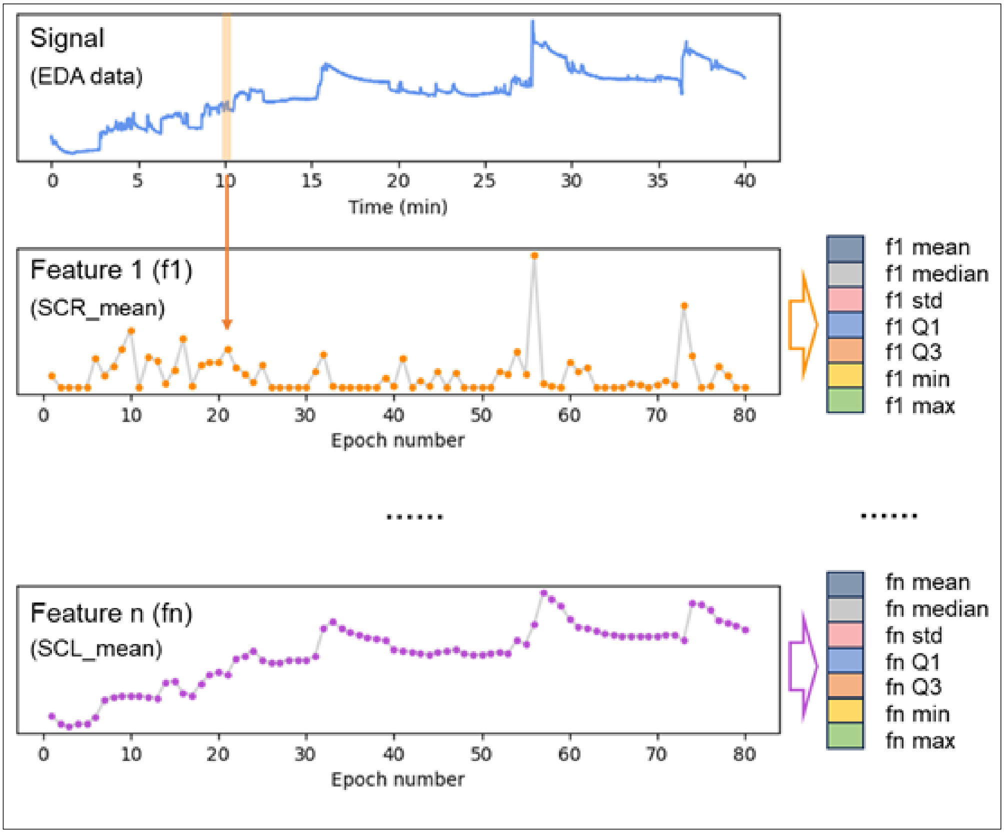

Figure 2

Schematic illustration of feature extraction

This figure illustrates a sample of the EDA signal from one participant during a lesson. The signal was first segmented into 30-second-long epochs. Then features were extracted from these epochs, and finally analyzed using various statistical measures.

2.4 Machine learning approach

This study adopted a regression model as the dependent variable is a continuous variable, which is a commonly-used approach in the prediction of psychological states (Gao et al., 2020; Lan et al., 2016; Sabbagh et al., 2020). The prediction pipeline is described below.

Prediction pipeline. In addressing irrelevant and redundant features within the machine learning pipeline, the following approach was adopted: i). *Feature Selection*: The k best features were selected based on their correlation with self-report situational interest. ii). *Dimensionality Reduction*: Given the presence of high collinearity among features (e.g., a strong correlation between feature_x_mean and feature_x_median), Principal Component Analysis (PCA) was applied. iii). *Model Training*: Subsequently, various regression models, including linear regression, ridge regression, support vector regression (Awad et al., 2015), and LightGBM regression (Ke et al., 2017), were employed to train the model. Following this pipeline, the study aimed to create a model that effectively captures the important patterns of situational interest while minimizing the influence of less relevant or redundant information.

Model evaluation. These models were compared with two baselines: *Random*: This baseline generates predictions based on a normal distribution estimated from the training set labels; *Average*: This baseline approach produces predictions using the average value derived from the training set. To compare the performance of the models, cross-validation, a widely used technique (Browne, 2000), was employed. Following 10-fold cross-validation, metrics such as mean absolute error (MAE), and mean squared error (MSE) were calculated to assess the models' effectiveness.

3. Results

3.1 Correlation analysis

Pearson's correlation coefficients (PCC) were computed between self-report situational interest and extracted features. In Study 1, out of $33 \times 7 = 231$ features analyzed, 47 showed a level of significance with $p < 0.05$, while 25 features demonstrated significance at $p < 0.01$, and 10 features at $p < 0.001$. In Study 2, out of the evaluated features, 49 displayed a level of significance at $p < 0.05$, 31 features showed significance at $p < 0.01$, and 19 features reached a level of significance at $p < 0.001$.

Figure 3 depicts the significant correlations between multimodal features and situational interest, with the majority of significant features identified within the EEG frequency domain. In Study 1, significant features were identified not only in EEG signals, but also in EDA and PPG data. However, correlation analysis in Study 2 revealed only two significant EDA or PPG features, which was probably due to the complex environment.

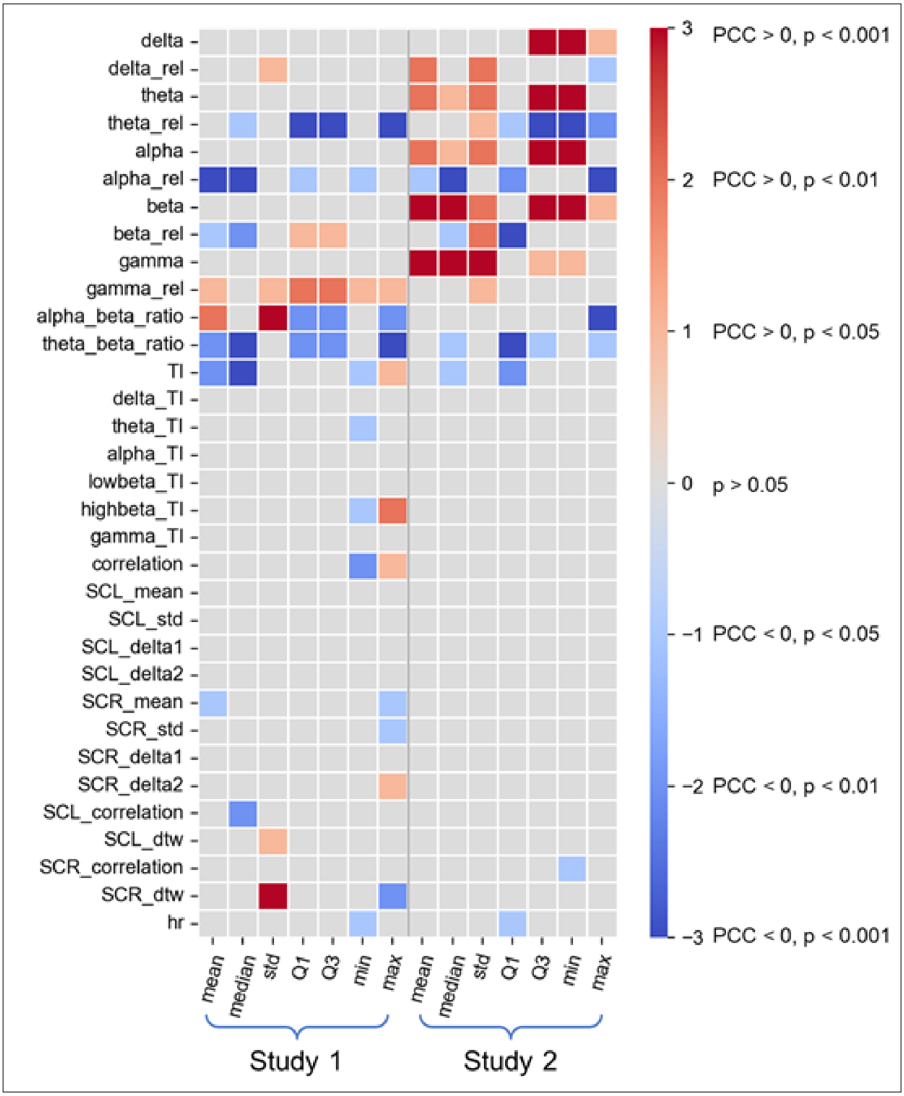

Figure 3

Influential features on situational interest

Note: PCC: Pearson's correlation coefficient

3.2 Machine learning results

The correlation analysis determined the hyperparameter in feature selection to be 10, 30, or 50. These values correspond to the approximate number of features significant at the 0.001, 0.01, and 0.05 levels, respectively.

As shown in Table 2, in Study 1, after 10-fold cross-validation, all models outperformed the two baseline models. Specifically, the model that selected the 30 best features and utilized Support Vector Regression (SVR) achieved the highest performance, with a mean absolute error (MAE) of 0.772 and a mean squared error (MSE) of 0.883. This represents a 30.3% reduction in MAE and a 57.4% reduction in MSE compared to the random baseline.

In Study 2, all regression models surpassed the performance of the random baseline model. Notably, the model that selected the top 30 features and employed the LightGBM regressor (LGBM) achieved the most impressive result: 0.798 in MAE and 0.993 in MSE. This represents a 31.0% reduction in MAE and a 49.7% reduction in MSE compared to the random baseline. However, it's important to note that the MAE for this model, at 0.798, did not surpass the average baseline, which had an MAE of 0.790. This outcome suggests that while the model has strengths, its generalizability and predictive accuracy in a naturalistic classroom environment may not be as robust as desired.

Table 2

Prediction performance for situational interest

Metrics	Study 1		Study 2	
	MAE	MSE	MAE	MSE
Random	1.109	2.072	1.157	1.974
Average	0.845	1.007	**0.790**	1.003
k = 10				
Linear	0.817	0.995	0.802	0.996
Ridge	0.817	0.994	0.802	0.996
SVR	0.810	0.971	0.800	1.004
LGBM	0.818	0.961	0.800	0.998
k = 30				
Linear	0.814	0.998	0.802	0.997
Ridge	0.814	0.995	0.802	0.997
SVR	**0.772**	**0.883**	0.803	1.003
LGBM	0.803	0.934	0.798	**0.993**
k = 50				
Linear	0.808	1.024	0.816	1.018
Ridge	0.807	1.020	0.816	1.018
SVR	0.783	0.913	0.811	1.016
LGBM	0.802	0.933	0.801	0.996

4. Discussion

The current study's findings offer contributions to the understanding and measurement of situational interest in educational settings. Correlation analysis and machine learning results indicate that situational interest can be predicted using multimodal neurophysiological data in a semi-controlled experiment (Study 1). While the data in a naturalistic setting (Study 2) identified several features significantly correlated with situational interest, the machine learning model demonstrated limited generalizability in this context.

The correlation analysis in this study provides neurophysiological support for different conceptualizations of situational interest. Hidi and Renninger (2019) described triggered situational interest as a “psychological state resulting from short-term changes in cognitive and affective processing associated with a particular class of content”, and maintained situational interest as a “psychological state that involves focused attention to a particular class of content that reoccurs and/or persists over time”. Chen et al. (2001), from a different perspective, identified five dimensions of situational interest: novelty, challenge, attention demand, exploration intention, and instant enjoyment. In the current study, an increase in situational interest was associated with a decrease in relative alpha power (alpha_rel), suggesting an increase in attention (Klimesch, 2012). Additionally, an increase in situational interest was accompanied by a decrease in the theta to beta ratio (theta_beta_ratio). This change could result from an increase in beta power or a decrease in theta power, pointing to associated cognitive processes.

While the correlation analysis successfully links features to situational interest as per these concepts, it’s important to note the inconsistency in significant features observed between the two studies. In Study 1, several EDA features showed significant correlation with situational interest. Contrastingly, in Study 2, only one EDA feature was significant. This difference in results might be attributed to the motion artifacts induced by diverse class activities, resulting in noisier electrodermal activity signals and consequently affecting the analysis results.

In addition to mean and median values, notable associations were also observed in the quartiles and extremes of the features across epochs (see Figure 3). This suggests that neurophysiological features at certain moments may reflect the overall situational interest during the stimuli, which aligns with Azevedo’s (2018) conceptualization of situational interest as a short-term spike in a specific activity. Among the significant features, TI presents an interesting pattern: while the mean and median of TI (TI_mean and TI_median) were negatively correlated with situational interest, suggesting that students with high situational interest generally exhibited lower synchrony with their peers, key moments showed a different trend. The positive association of TI’s maximum value (TI_max) indicates that at these critical moments, synchrony among these students was higher. These findings highlight the potential of leveraging dynamic information through various descriptive statistics.

In the machine learning analysis, 30 features were found to offer the most accurate predictions, effectively balancing the risks of overfitting and underfitting. Notably, the most effective model varied between the two datasets. In Study 1, the Support Vector Regressor (SVR) had the best prediction performance, which is likely attributed to the SVR’s pattern extraction capabilities facilitated by kernel method. In contrast, Study 2 found the LightGBM regressor to be the most effective. LightGBM utilizes the benefits of decision trees for pattern extraction, which is likely more fitted to the data of Study 2.

In Study 1, the prediction outcomes were comparable with similar studies that have focused on representing diverse psychological concepts in educational contexts (Gao et al., 2020; Sharma et al., 2020). However, the findings of Study 2 revealed that the model's generalizability was less than ideal. This could be due to the influence of various factors on situational interest, such as the subject matter (Kunter et al., 2007), and teaching style (Dever & Karabenick, 2011). These factors were less controllable in the more naturalistic setting of Study 2, potentially impacting the model's effectiveness in this context. The difference of two datasets underscores the challenge of accurately predicting situational interest in real-world educational environments, which can be affected by various unknown factors.

In conclusion, the study's findings provide empirical support for the conceptual framework of situational interest, demonstrate the potential of neurophysiological data in educational assessments, and also underscore the challenges in naturalistic paradigm. Although cumulative experimental evidence increasingly supports the feasibility of using neurophysiological biomarkers to predict psychological states (Goswami, 2009), stable biomarkers for situational interest still need verification by further experiments. To fully exploit the potential of the neuroscientific approach in educational settings, a more precise conceptualization and additional experimental evidence are essential for a comprehensive neurophysiological interpretation of situational interest.

Acknowledgement

This work was supported by the Beijing Educational Science Foundation of the Fourteenth 5-year Planning (BGEA23019). The authors appreciate Xinqiao Gao, Baosong Li, Zhilin Qu, Fei Qin, Guannan Yao, Xinya Liu, Kun Wang, Liyi Yang, Jianhua Zhang, Ziyan Xu, Zheng Dong, Wenhui He, Xiaomeng Xu, and Yingyao Fu for their assistance during data collection.

References

Ahuja, N. D., Agarwal, A. K., Mahajan, N. M., Mehta, N. H., & Kapadia, H. N. (2003). GSR and HRV: its application in clinical diagnosis. *16th IEEE Symposium Computer-Based Medical Systems, 2003. Proceedings*., New York, NY, USA, 2003, pp. 279-283, https://doi.org/10.1109/CBMS.2003.1212802

Ainley, M. (2012). Students' interest and engagement in classroom activities. In S. L. Christenson, A. L. Reschly & C.Wylie(Eds.), *Handbook of research on student engagement* (pp. 283-302). Springer. https://doi.org/10.1007/978-1-4614-2018-7_13

Ainley, M., & Hidi, S. (2014). Interest and enjoyment. In R. Pekrun & L. Linnenbrink-GarciaInternational (Eds.), *International handbook of emotions in education* (pp 205-227). Routledge/Taylor & Francis Group.

AL-Ayash, A., Kane, R. T., Smith, D., & Green-Armytage, P. (2016). The influence of color on student emotion, heart rate, and performance in learning environments. *Color Research & Application, 41*(2), 196-205. https://doi.org/10.1002/col.21949

Awad, M., & Khanna, R. (2015). Support vector regression. In *Efficient learning machines: Theories, concepts, and applications for engineers and system designers* (pp 67-80). Apress. https://doi.org/10.1007/978-1-4302-5990-9_4

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in psychology, 12,* 702538. https://doi.org/10.3389/fpsyg.2021.702538

Azevedo, F. S. (2018). An inquiry into the structure of situational interests. *Science Education, 102*(1), 108-127. https://doi.

org/10.1002/sce.21319

Babiker, A., Faye, I., Mumtaz, W., Malik, A. S., & Sato, H. (2019). EEG in classroom: EMD features to detect situational interest of students during learning. *Multimedia Tools and Applications, 78*(12), 16261-16281. https://doi.org/10.1007/s11042-018-7016-z

Browne, M. W. (2000). Cross-validation methods. *Journal of mathematical psychology, 44*(1), 108-132. https://doi.org/10.1006/jmps.1999.1279

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences, 18*(8), 414-421. https://doi.org/10.1016/j.tics.2014.04.012

Chen, A., Darst, P. W., & Pangrazi, R. P. (2001). An examination of situational interest and its sources. *British journal of educational psychology, 71*(3), 383-400. https://doi.org/10.1348/000709901158578

Chen, J., Qian, P., Gao, X., Li, B., Zhang, Y., & Zhang, D. (2023). Inter-brain coupling reflects disciplinary differences in real-world classroom learning. *npj Science of Learning, 8*(1), 11. https://doi.org/10.1038/s41539-023-00162-1

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

de Cheveigné, A., & Arzounian, D. (2018). Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data. *NeuroImage, 172*, 903-912. https://doi.org/10.1016/j.neuroimage.2018.01.035

Dever, B. V., & Karabenick, S. A. (2011). Is authoritative teaching beneficial for all students? A multi-level model of the effects of teaching style on interest and achievement. *School Psychology Quarterly, 26*(2), 131–144.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., & Ding, M. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current biology, 27*(9), 1375-1380. https://doi.org/10.1016/j.cub.2017.04.002

Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. *sensors, 20*(3), 592. https://doi.org/10.3390/s20030592

Gao, N., Shao, W., Rahaman, M. S., & Salim, F. D. (2020). n-Gage: Predicting in-class emotional, behavioural and cognitive engagement in the wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4*(3), 1-26. https://doi.org/10.1145/3411813

Goswami, U. (2009). Mind, brain, and literacy: Biomarkers as usable knowledge for education. *Mind, Brain, and Education, 3*(3), 176-184. https://doi.org/10.1111/j.1751-228X.2009.01068.x

Greco, A., Valenza, G., Citi, L., & Scilingo, E. P. (2017). Arousal and valence recognition of affective sounds based on electrodermal activity. *IEEE Sensors Journal, 17*(3), 716-725. https://doi.org/10.1109/JSEN.2016.2623677

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2016). cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering, 63*(4), 797-804. https://doi.org/10.1109/TBME.2015.2474131

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of educational psychology, 89*(1), 92-102. https://doi.org/10.1037/0022-

0663.89.1.92

Hidi, S. (2000). An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance: The Search for Optimal Motivation and Performance* (pp. 309–339). Academic Press. https://doi.org/10.1016/B978-012619070-0/50033-7

Hidi, S. (2006). Interest: A unique motivational variable. *Educational research review, 1*(2), 69-82. https://doi.org/10.1016/j.edurev.2006.09.001

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist, 41*(2), 111-127. https://doi.org/10.1207/s15326985ep4102_4

Hidi, S. E., & Renninger, K. A. (2019). Interest development and its relation to curiosity: Needed neuroscientific research. *Educational Psychology Review, 31*, 833-852. https://doi.org/10.1007/s10648-019-09491-3

Immordino-Yang, M. H., & Gotlieb, R. (2017). Embodied brains, social minds, cultural meaning: Integrating neuroscientific and educational research on social-affective development. *American Educational Research Journal, 54*(1, Suppl), 344S–367S. https://doi.org/10.3102/0002831216669780

Kanoga, S., Kanemura, A., & Asoh, H. (2019). Multi-scale dictionary learning for ocular artifact reduction from single-channel electroencephalograms. *Neurocomputing, 347,* 240-250. https://doi.org/10.1016/j.neucom.2019.02.060

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. GarnettAdvances (Eds.), *Advances in neural information processing systems 30* (pp 3149–3157). Curran Associates Inc.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences, 16*(12), 606-617. https://doi.org/10.1016/j.tics.2012.10.007

Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and instruction, 17*(5), 494-509. https://doi.org/10.1016/j.learninstruc.2007.09.002

Lan, Z., Müller-Putz, G. R., Wang, L., Liu, Y., Sourina, O., & Scherer, R. (2016). Using support vector regression to estimate valence level from EEG. *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* 002558 - 002563. https://doi.org/10.1109/SMC.2016.7844624

Liu, N.-H., Chiang, C.-Y., & Chu, H.-C. (2013). Recognizing the degree of human attention using EEG signals from mobile sensors. *sensors, 13*(8), 10273-10286. https://doi.org/10.3390/s130810273

Liu, Y., Wang, T., Wang, K., & Zhang, Y. (2021). Collaborative learning quality classification through physiological synchrony recorded by wearable biosensors. *Frontiers in Psychology, 12*, 674369. https://doi.org/10.3389/fpsyg.2021.674369

Lyu, Y., Luo, X., Zhou, J., Yu, C., Miao, C., Wang, T., Shi, Y., & Kameyama, K.-i. (2015). Measuring photoplethysmogram-based stress-induced vascular response index to assess cognitive load and stress. *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 857-866. https://doi.org/10.1145/2702123.2702399

Picard, R. W., Vyzas, E., & Healey, J. (2001).

Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence, 23*(10), 1175-1191. https://doi.org/10.1109/34.954607

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion, 37*, 98-125. https://doi.org/10.1016/j.inffus.2017.02.003

Putman, P., Verkuil, B., Arias-Garcia, E., Pantazi, I., & van Schie, C. (2014). EEG theta/beta ratio as a potential biomarker for attentional control and resilience against deleterious effects of stress on attention. *Cognitive, Affective, & Behavioral Neuroscience, 14*, 782-791. https://doi.org/10.3758/s13415-013-0238-7

Qu, Z., Chen, J., Li, B., Tan, J., Zhang, D., & Zhang, Y. (2020). Measurement of high-school students' trait math anxiety using neurophysiological recordings during math exam. *IEEE Access, 8*, 57460-57471. https://doi.org/10.1109/ACCESS.2020.2982198

Renninger, K. A. (1992). Individual interest and development: Implications for theory and practice. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 361–395). Lawrence Erlbaum Associates, Inc.

Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 373–404). Academic Press. https://doi.org/10.1016/B978-012619070-0/50035-0

Renninger, K. A., & Hidi, S. (2015). *The power of interest for motivation and engagement.* Routledge.

Rotgans, J. I., & Schmidt, H. G. (2011). The role of teachers in facilitating situational interest in an active-learning classroom. *Teaching and teacher Education, 27*(1), 37-42. https://doi.org/10.1016/j.tate.2010.06.025

Roy, J.-C., Boucsein, W., Fowles, D. C., & Gruzelier, J. (2012). *Progress in electrodermal research* (Vol. 249). Springer Science & Business Media.

Sabbagh, D., Ablin, P., Varoquaux, G., Gramfort, A., & Engemann, D. A. (2020). Predictive regression modeling with MEG/EEG: From source power to signals and cognitive states. *NeuroImage, 222*, 116893. https://doi.org/10.1016/j.neuroimage.2020.116893

Schraw, G., Flowerday, T., & Lehman, S. (2001). Increasing situational interest in the classroom. *Educational Psychology Review, 13*(3), 211-224. https://doi.org/10.1023/A:1016619705184

Sharma, K., Niforatos, E., Giannakos, M., & Kostakos, V. (2020). Assessing cognitive performance using physiological and facial features: Generalizing across contexts. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4*(3), 1-41. https://doi.org/10.1145/3411811

Shui, X., Zhang, M., Li, Z., Hu, X., Wang, F., & Zhang, D. (2021). A dataset of daily ambulatory psychological and physiological recording for emotion research. *Scientific Data, 8*(1), 161. https://doi.org/10.1038/s41597-021-00945-4

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

Tan, A. L., Gillies, R., & Jamaludin, A. (2021). A case study: Using a neuro-physiological measure to monitor students' interest and learning during a micro: bit activity. *Education sciences, 11*(8), 379. https://doi.org/10.3390/educsci11080379

Tsai, Y.-M., Kunter, M., Lüdtke, O., Trautwein, U., & Ryan, R. M. (2008). What makes lessons interesting? The role of situational and individual factors in three school subjects. *Journal of Educational Psychology, 100*(2), 460-472. https://doi.org/10.1037/0022-0663.100.2.460

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Zhu, J., & Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ open, 3*(8), e002847. https://doi.org/10.1136/bmjopen-2013-002847

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., Gao, S., Sun, Y., Ge, W., & Zhang, W. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion, 83*, 19-52. https://doi.org/10.1016/j.inffus.2022.03.009

Wang, Z., & Adesope, O. (2016). Exploring the effects of seductive details with the 4-phase model of interest. *Learning and Motivation, 55*, 65-77. https://doi.org/10.1016/j.lmot.2016.06.003

Zhang, J., Wang, K., & Zhang, Y. (2021). Physiological Characterization of Student Engagement in the Naturalistic Classroom: A Mixed-Methods Approach. *Mind, Brain, and Education, 15*(4), 322-343. https://doi.org/10.1111/mbe.12300

Zhang, Y., Qin, F., Liu, B., Qi, X., Zhao, Y., & Zhang, D. (2018). Wearable neurophysiological recordings in middle-school classroom correlate with students' academic performance. *Frontiers in Human Neuroscience, 12*, 457. https://doi.org/10.3389/fnhum.2018.00457

Zhou, Y., Huang, S., Xu, Z., Wang, P., Wu, X., & Zhang, D. (2021). Cognitive workload recognition using EEG signals and machine learning: A review. *IEEE Transactions on Cognitive and Developmental Systems, 14*(3). https://doi.org/10.1109/TCDS.2021.3090217