# Detecting emotions in a learning environment: a multimodal exploration

Sharanya Lal, Tessa H.S. Eysink, Hannie A.H. Gijlers, Willem B. Verwey and Bernard P.Veldkamp

*University of Twente, Drienerlolaan 5, Enschede, 7522 NB, The Netherlands*

**Abstract**

Learner-emotions are intrinsically linked with learning experiences and academic outcomes. Therefore, intelligent learning environments need to be emotion-aware to bring learners to their zone of proximal development. In this paper, we describe the first steps towards such a system. In this study, we manipulated task difficulty with the aim of detecting the physiological indicators of accompanying emotions, namely boredom/anger (during an easy task), enjoyment (during a moderately challenged task) and frustration/boredom (during a difficult task). Twenty-one adults (13 females and 8 males, Mage = 24.1 years) participated in a repeated- measures quasi-experimental set-up. Data were collected via Empatica E4 wristbands and self-reports. Results indicate that varying task difficulty may be associated with changes in skin temperature, phasic and tonic skin conductance, and heart rate. Findings encourage further exploration and thoughts on study design are discussed.

**Keywords** 1

psychophysiology, wearables in education, affective computing, emotion detection

## 1. Introduction and background

### 1.1. Emotions in learning

Emotions play a significant role in learning and this is evidenced by the growing body of work on the interaction of learner emotions, well-being, and learning outcomes [1], [2], [3], [4], [5]. For example, [5] found that the induction of positive emotions in learners resulted in higher learning transfer, greater mental engagement and lower levels of reported task difficulty. In another study, [6] found that positive emotions (namely enjoyment and pride) predicted high learning achievements while the opposite was true for negative emotions (namely anger, anxiety, shame, boredom and hopelessness). Therefore, to optimise learning experiences and outcomes, it is essential that one takes learner emotions into account. In today's era of digital learning, this calls for intelligent learning systems that can detect learners' emotions to provide optimally adjusted support.

Proceedings of the Doctoral Consortium of Sixteenth European Conference on Technology Enhanced Learning, September 20–21, 2021, Bolzano, Italy (online).
EMAIL: s.lal@utwente.nl (A. 1); t.h.s.eysink@utwente.nl (A. 2); a.h.gijlers@utwente.nl (A. 3); w.b.verwey@utwente.nl (A. 4); b.p.veldkamp@utwente.nl (A. 5)
ORCID: 0000-0002-6228-8850 (A. 1); 0000-0001-5820-4469 (A. 2); 0000-0003-2406-6541 (A. 3); 0000-0003-3994-8854 (A. 4); 0000-0003-3543-2164 (A. 5)

## 1.2. Theoretical perspectives

In their meta-study that showed strong correlations between emotional, cognitive and learning processes in e-learning environments, [7] suggest fostering optimal levels of subjective control (i.e., a learner's appraisal of how much control over a task they have) and value (i.e., the value a learner places on a task). Their results align with and suggestions rely heavily on Pekrun's [4] Control-Value theory that states that the subjective appraisals of control and value are central to emotions related to learning. For example, if the learner sees positive value in a task and has high control of actions, they experience enjoyment. On the other hand, if they see no value in the task, they feel bored irrespective of whether they have high or low control. Similarly, if learners find themselves unable to control an activity, they experience frustration irrespective of the value they placed on the same. Pekrun's [4] activity related emotions draw on Csikszentmihalyi's [1] seminal work on 'flow' – a state of extreme concentration, when someone is so engaged in the task at hand that they forget the passage of time. Flow theory suggests that learners in 'flow' experience enjoyment and happiness and that this is achieved when one not only has a clear goal, a sense of purpose and immediate feedback, but also a balance of challenge and skill (with challenge and skill level being just above the average for the person) [1], [8]. This in turn shares similarities with one of the most significant concepts in learner centric education – the Zone of Proximal Development (ZPD) [9], which posits that learning is optimal when a task is just out of the learner's reach and they have available the assistance of a more skilled/knowledgeable person. Taking cue from this, in this study, we look at emotions in light of learner's perceptions of task difficulty, challenge to skill balance, absorption in a task and control-value appraisals.

## 1.3. Detecting emotions

Emotion detection has traditionally been done through learner reported data [10]. Such an approach has several limitations including the subjective nature of self-reports and the likely temporal mismatch between when an emotional state has occurred and data are collected [11]. The latter could result in the collection of data for another moment in time or even inaccuracies when recalling past experiences. Consequently, there is much interest in alternate approaches to emotion detection that can provide objective, time-specific and reliable data. One approach that is notably gaining traction is the use of physiological measures to understand underlying psychological processes. For example, [12] found that emotional valence (i.e., the extent to which an emotion is negative or positive) was positively related to blood volume pulse (i.e., a measure of the changes in blood volume flowing through one's arteries and capillaries). Skin conductance (i.e., skin's property of conducting electricity) has been found to reflect stress during a task [13], and emotional arousal [14]. In recent educational research specifically, [15] studied adolescent girls learning in maker-spaces and found that skin conductance was positively related to engagement. In another study, [16] measured average student heart rates (i.e., the number of heart beats per minute) during medical school lectures and found a steady decline from the start to the end of a lecture. They also found that heart rate significantly increased during periods of student interaction such as group-based problem solving. More recently, [17] in a study involving 67 students solving statistical exercises of varying difficulty found that heart rate and skin temperature were significantly related to self-reported cognitive load and skin temperature specifically to task performance. Studies like these suggest that these measures are useful indicators of challenge to skill balance, perceived task difficulty and task absorption and can therefore offer a glimpse into learner emotions. Physiological signals that can now be assessed with portable devices give us access to vast amounts of uninterrupted, time-specific and objective data points, thus bringing us closer to understanding a learner's emotional state in real-time. However, research is still at a nascent stage and there is value in advancing the body of literature on the same (e.g., [18], [19], [20]).

## 2. Research aims of present study

The present study is the first step in our research project that is geared towards developing an intelligent learning system that adapts to a learner's emotions so as to bring them to their ZPD. Therefore, this paper focuses on emotion detection. To this end, a repeated-measures quasi- experimental design was adopted wherein physiological data in combination with self- reported measures were used to detect emotional states. The physiological signals investigated in the study were skin conductance, skin temperature, blood volume pulse and heart rate. Emotional states were elicited primarily through the manipulation of task difficulty in a digital learning environment designed to teach programming skills. This manipulation (see Methods) was done with the expectation that it would lead to differences in learners' perceptions of challenge to skill balance, task absorption and therefore emotions. Drawing on the ideas of Csikszentmihalyi [1] and Pekrun [4] and past studies on psychophysiological measures,several conjectures were made:

1. For the task that was too easy, learners would perceive a mismatch between challenge and skills and have lowabsorption in task. Based on their appraisal of control over and value of the task, they would experience either boredom (no value, high control) or anger (negative value, high control). Boredom being a deactivating emotion (i.e., one that is associated with low arousal) would be associated with low skin conductance and heart rate. Anger on the other hand being an activating emotion (i.e., one that is associated with high arousal) would be associated with high skin conductance and heart rate.
2. For the task that was too difficult, the expectation was that learners would perceive a mismatch between challenge and skills and have low absorption in task. Based on control and value appraisal of the task, they would either experience frustration (positive/negative value, low control) or boredom (no value, low control). Unlike boredom, frustration being an activating emotion would be associated with high skin conductance and heart rate.
3. For the task that was neither too difficult nor too easy, it was expected that learners would perceive a balance between the challenge and their skills and have high task absorption. An appraisal of high control and high value of the task would be associated with a positive emotional state (i.e., enjoyment). Enjoyment being an activating emotion would be associated with high skin conductance and heart rate.We also expected blood volume pulse to be an indicator of emotional valence [12] and skin temperature to be high during the difficult task [17].

Emotional states were also elicited through a sample taken from the Open Affective Standardised Image Set (OASIS) [21] (described in Methods). The hypothesis was thatthe valence and arousal associated with the different images would be reflected in the physiological signals. Therefore, these could act as reference points when interpreting emotions during the programming tasks.

Thus, this study aimed to detect psychophysiological indicators (if any) of learner emotions associated with tasks of varying difficulty.

## 3. Methods

### 3.1. Participants

Participants consisted of 21 (13 females and 8 males, 19-32 years old, $M_{\text{age}}$ = 24.14 years) university students and working professionals based in the Netherlands. The sample consisted of persons of 6 nationalities and different educational levels (11 bachelor students, 1 bachelor's degree holder, 8master's degree holders and 1 PhD student). All participants had at least working knowledge of English and basic computer skills. Participation was voluntary and active consent had been received from all participants before the start of theexperiment.

### 3.2. Materials

#### 3.2.1. Primary stimuli set – programming tasks

In the learning environment [22], participants programmed instructions by joining blocks of code to control a red 'robot' (see Figure 1). The goal was to make the robot reach the end of its path by codingits trajectory. Paths could be 5-, 10- or 15-step, eachrequiring a longer or more sophisticated piece of code than the previous. The environment also had a free-play 'Sandbox' mode, in which participants were free to explore the environment in any way they wanted – there was no specific aim to this activity. Three tasks of varying difficulty were designed within the learning environment. The moderately challenging task was to complete a 5-, 10- and 15-step path (see Figure 2) within 10 minutes. The easy task was to do a 5-step path over and over again for 10 minutes. The difficult task wasto 'decipher the aim and rules of the Sandbox' and 'complete it successfully' in 10 minutes. This was considered 'difficult' because the Sandbox mode does not actually have a tangible goal or rules, thus making the task a wild goose chase (however, participants were not aware of this fact). User responses during pilot testing of the environment and tasks concurred with these expectations.

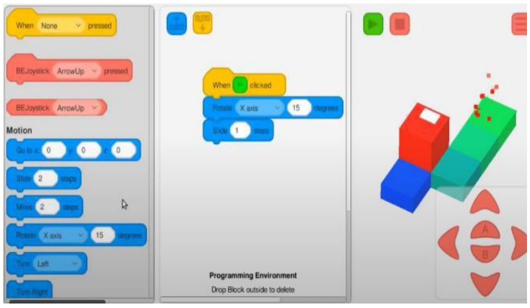

**Figure 1:** In the digital learning environment, participants selected blocks of code (left pane), edited and joined them to form a piece of code (center pane) that would move the red robot to the end of its path (right pane)

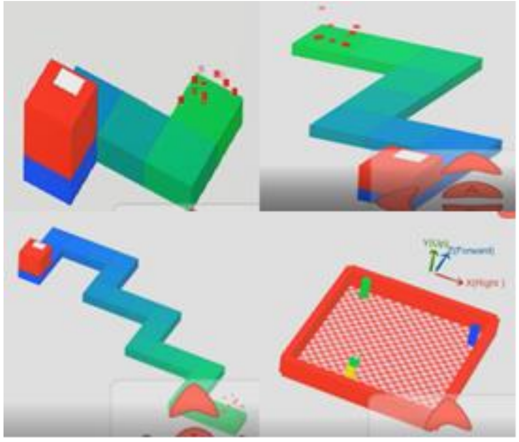

**Figure 2:** In the learning environment, one could either code to make the red robot reach the end of its 5- (Top-Left), 10- (Top-Right) or 15-step (Bottom-Left) path, or explore freely in the Sandbox mode (Bottom-Right)

### 3.2.2. Baseline-measurement stimulus

A video with the instructions, "Sit still and relax" was displayed for 5 minutes. At the 4 m 50s mark, an audio signal indicated the end of the rest period. At this point, the phrase "I feel: " followed by a smiley meter (described in a subsequent sub-section) appeared on the screenfor 10 seconds.

### 3.2.3. Secondary stimuli set – images

A set of 35 500x400 pixel images – 13 positive (for example, a puppy in a teacup), 10 negative (for example, garbage) and 12 neutral (for example, a tiled roof) were sampled from OASIS [21]. The value given to these images was based on participant-reported valence in the original study. While sampling, graphic and sexually explicit images were excluded. The images were presented one after the other with intermittent5 s pauses wherein a blank screen was inserted. Each image was displayed for 10 seconds. On the 6th second, a smiley meter (described in a subsequent paragraph) along with the phrase "This photo makes me feel…" appeared below the image and stayed visible till the end of the 10th second.

### 3.2.4. Hardware and software set up

Physiological data were collected using the biosensing wristband E4. The E4 makes use of an electrodermal activity sensor that measures sympathetic nervous system arousal via stainless steel electrodes that are placed on the ventral wrist. This arousal is quantified in terms of skin conductance which is measured in microSiemens (μS) and sampled at 4 Hz (i.e., 4 readings per second). Skin temperature was collected in degree Celsius (°C) via the E4's infrared thermopile sensor at a sampling frequency of 4 Hz. Blood volume pulse was collected from the E4's photoplethysmography (PPG) sensor placed on the dorsal wrist and was sampled at 64 Hz. Heart rate (calculated per 10 s) was derived from blood volume pulse. In addition to this, acceleration data (indicating movement) from the E4's accelerometerwere collected at 32 Hz. All data were streamed to Empatica's cloud-based repository via an android application set up on a mobile phone which in turn was connected via Bluetooth to the E4. The internal clock of the E4 was synchronised with that of the computer on which the stimuli were loaded. A screen recorder was set up on the computer so as to capture timestamps of the different stimuli and digital behaviour during the programming tasks. A handheld timer was used to facilitate and keep trackof the different activities in the study.

### 3.2.5. Self-reports

Self-reported data were collected using several tools:

**Smiley meter:** A five point smiley meter [23] was used to collect participants' perception of different stimuli during the study. Participants were expected to reflect on how the stimulus (a programming task, a baseline activity or an image) made them feel and point to the smiley that best represented their emotional state. The scale was used unmarked to avoid putting specific affect-related words into the participant's head.

**Short flow scale (SFS) and task difficulty scale:** A 20-item short flow scale [24] was used asa self-report of experiences during the three programming tasks. The SFS has 2 sub-scales, 'Challenge to skill balance' (Chal2Skill) (11 items) and 'Task Absorption' (Task_Absorption) (9 items) [24]. Since the twostatements in the scale , "It was boring for me" and "My attention was not engrossed at all by the activity" were negatively framed, they were recoded. Testing for reliability, we foundCronbach's α = .92, α = .79 and α = .91 of the SFS for the moderately challenging, easy and difficult task respectively. Reliability tests were also performed for each subscale 'challenge to skill balance' ('Chal2Skill') and 'task absorption' ('Task_Absorption'). We found that the sub-scales Chal2Skill and Task_Absorption had a) Cronbach's α = .95 and α = .74 respectively, for the moderately challenging task, b) α = .91 and α = .93 respectively, for the easy task, and c) α = .88 and α = .90 respectively, for the difficult task. Consequently, new variables valued as the mean of each subscale were computed to be used for further analyses. It is important to note that lowand high Chal2Skill ratings denote an imbalanceof challenge and skill (i.e. a task is too difficult or a task is too easy, respectively) and a moderate Chal2Skill rating denotes a balance of challenge and skill. Another self-report measure used after the programming tasks was a one- item scale on perceived task difficulty(henceforth referred to as the Task_Difficulty scale). The scale consisted of the following item – 'Was this task 1) Too easy 2) Easy 3) Just right 4) Difficult 5) Very difficult?'

**Interview:** An audio-recorded face-to-face semi- structured interview was conducted at the end of the study to glean participants' experiences during the experiment. Participants were asked how they were feeling at the start and end of the study, if they could describe their experiences during the different programming tasks and baselines, and their rationale for selecting a particular smiley on corresponding smiley meters.

## 3.3. Procedure

This study took place during the Covid-19 pandemic. Consequently, participants received hygiene and safety guidelines by e-mail and the experimental space and all equipment were sanitized before each use. On the day of the study, participants were individually seated in a closed lab space set up to minimize external distractions. Demographic data of participants namely age, sex, nationality, handedness, prior knowledge in programming and educational level were collected. Participants then received a general outline of the experimental set-up, procedure, tools and expected code of conduct. Once ready, they were fitted with the Empatica E4 on their non-dominant hand to mitigate the effects of hand movements, making sure that the wristband's sensors made complete skin contact and the electrodes for skin conductance detection were in line with the gap between the middle and ring finger. The E4 was then switched on, and readings were checked to see that a stable connection had been established. Participants then faced a computer screen with their non-dominant hand either on their lap or on the table. Participants first watched an instructional video outlining the components of the learning environment and how to navigate it. They were then guided by the baseline video during which they sat still and could either look at the computer screen orthe white wall behind it, or keep their eyes closed. Then participants proceeded to do the three programming tasks one after the other. The completion of the tasks was followed by another baseline reading, then a viewing of the images and a third and final baseline reading. After each baseline, programming task and image, participants indicated their emotional state on the smiley meter. Thus for each participant, a total of 41 smiley meter ratings were collected. Meanwhile, the researcher kept time, took notes and checked that the wristband was collecting a continuous stream of data. Participants then filled three copies of the SFS and Task_Difficulty scale, once for each programming task, were interviewed and finally debriefed about the purpose of the study. Figure 3 shows the experimental procedure.

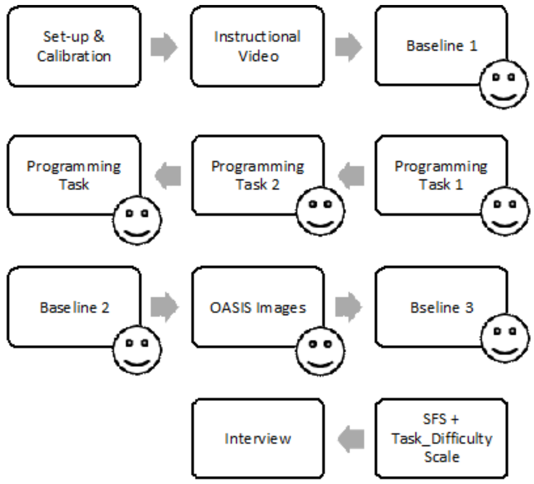


**Figure 3:** Study procedure

## 3.4. Data pre-processing

Blood volume pulse, heart rate, skin conductance and skin temperature readings were obtained as separate files. These were combined using a Python program that took the earliest and latest time stamps and interpolated all readings between the two. This involved bringing all data capturing times to a 0.25 second temporal resolution (in keeping with the 4 Hz sampling rate of the electrodermal activity sensor). Timestamps for various user actions and events (i.e., start and end of a stimulus) were obtained from screen recordings and added to these data. These were used to determine the duration of time windows to be analysed. Baselines were computed as the start of the baseline video to the reading just before the appearance of the smiley meter. The duration of an image stimulus was coded as the moment the image was displayed to the moment just before the appearance of the smiley meter. Task duration was 10 minutes unless a participant took less time to complete a task. All continuous physiological readings falling within a time window were averaged. These were then standardised by subtracting from them the average of all the baseline readings. Further analyses were performed using these standardised values.

Skin conductance was pre-processed using the MATLAB (The MathWorks, Inc., Natick, MA, U.S.A. ) software package 'Ledalab' (version 3.4.9 http://www.ledalab.de). Signal pre-processing included decomposition to its two components, phasic skin conductance (rapidly changing signal) and tonic skin conductance level (slow-moving signal), using the continuous decomposition analysis method [25] and feature extraction. Feature extraction was done using a threshold of 0. 01 μS. Phasic signal features that were extracted were namely onset and amplitude of non-specific significant skin conductance responses (nSCRs). These were used to compute nSCR frequency (nSCR/min) for each programming task. Baseline nSCR frequency was computed as the average of all three baselines. Taking cue from Pijeira-Díaz et al. (2018), phasic skin conductance was computed as a categorical variable with 3 values: 0 (low nSCR frequency – 0 to 3 SCR/min), 1 or (medium nSCR frequency – 4 to 20 nSCR/min) and 2 (high nSCR frequency – 21 and above nSCR/min). Tonic skin conductance data was extracted as a continuous variable.

## 4. Results

To answer the exploratory question of whether we could detect psychophysiological indicators (if any) of learner emotions associated with tasks of varying difficulty, we made comparisons across the three tasks and deviations from the baseline. We used linear mixed models while controlling for acceleration and demographic data. Pairwise comparisons were computed having applied Bonferroni correction. Across tasks, we found a significant variation in skin conductance [$F(3, 60) = 15.09$, $p = 0.00$] , heart rate [$F(3, 60) = 9.61$, $p = 0.00$] and temperature [$F(3, 60) = 3.13$, $p = 0.03$]. Please refer to Figures 4, 5 and 6 for more details.

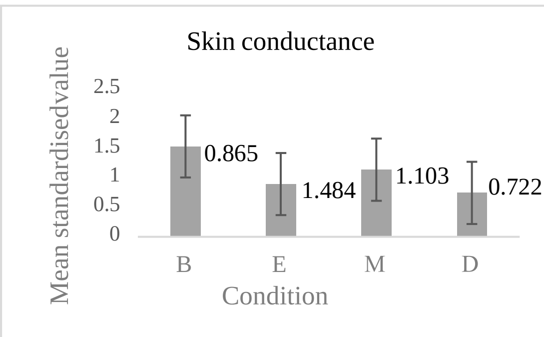


**Figure 4:** SC at baseline (B) was significantly higher than that during the easy (E) [mean difference = 0.62, $p$ = 0.00], moderately challenging (M) [mean difference = 0.38, $p$ = 0.02] and difficult (D) tasks [mean difference = 0.76, $p$ = 0.00]. SC during the moderately challenging task (M) was significantly higher than that during the difficult (D) task [mean difference = 0.38, $p$ = 0.02].

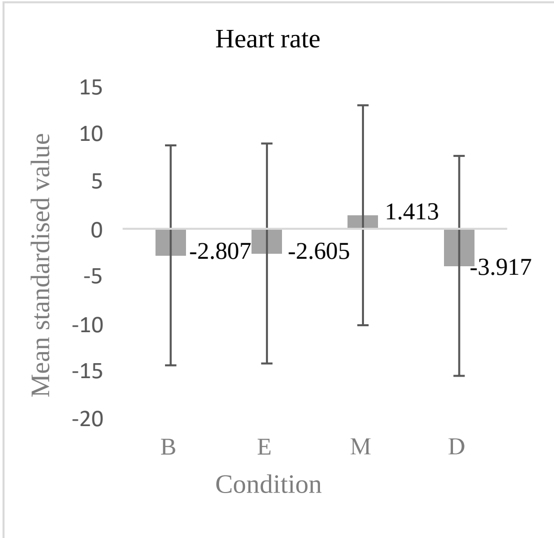


**Figure 5:** Heart rate during the moderately challenging task (M) was significantly greater than that during the baseline (B) [mean difference = 4.22, $p$ = 0.00], easy (E) [mean difference = 4.02, $p$ = 0.00] and difficult (D) tasks [mean difference = 5.33, $p$ = 0.00].

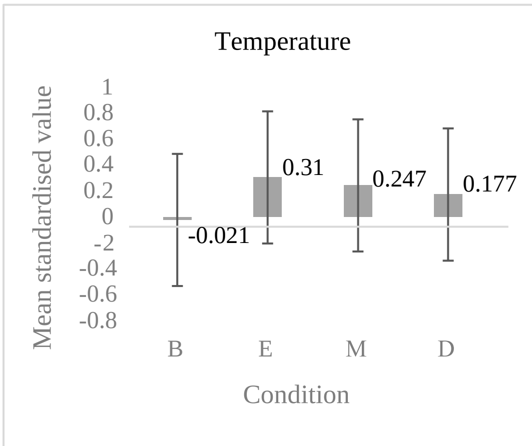


**Figure 6:** Temperature during the easy (E) task was significantly higher than that during the baseline (B), [mean difference = 0.33, $p$ = 0.03]. No significant changes during the moderately challenging (M) and difficult (D) tasks were observed.

Results indicated no significant changes in blood volume, $F(3, 60) = 1.20$, $p = 0.32$ and tonic skin conductance, $F(3, 66) = 1.46$, $p = 0.23$.

Relationships between physiological data and appraisals of challenge to skill balance, task difficulty and task absorption were explored. To do this, demographic data were included as fixed factors and participant was a random factor in the linear mixed model. We found no effect of Chal2Skill ($F$(1, 38.04) = 0.79, $p$ = 0.38), Task_Absorption ($F$(1, 37.61) = 0.06, $p$ = 0.81) and Task_Difficulty ($F$(4, 36.06) = 0.10, $p$ = 0.98) on phasic skin conductance. We also found no effect of Chal2Skill ($F$(1, 33.69) = 1.97, $p$ = 0.17), Task_Absorption $F$(1, 33.58) = 0.46, $p$ = 0.50) and Task_Difficulty ($F$(4, 33.35) = 0.96, $p$ = 0.44) on heart rate. No significant effect of Chal2Skill ($F$(1, 34.88) = 1.45, $p$ = 0.24), Task_Absorption ($F$(1, 34.61) = 1.45, $p$ = 0.24) and Task_Difficulty ($F$(4, 33.96) = 0.93, $p$ = 0.46) was found on blood volume pulse. Chal2Skill ($F$(1, 38.60) = 1.29, $p$ = 0.26), Task_Absorption ($F$(1, 38.22) = 0.80, $p$ = 0.38) and Task_Difficulty ($F$(4, 36.37) = 2.09, $p$ = 0.10) had no significant effects on temperature. Chal2Skill was found to have a positive effect on tonic skin conductance ($\beta$ = 0.43, $t$(36.96) = 2.93, $p$ = 0.00, 95% CI [0.13, 0.73] and Task_Absorption was found to have a negative effect ($\beta$ = -0.37, $t$(37.38) = -3.56, $p$ = 0.00, 95% CI [-0.59, -0.16]). There are some indications that Task_Difficulty ratings negatively affect tonic skin conductance: For Task_Difficulty = 1, $\beta$ = -2.27, $t$(34.67) = -4.33, $p$ = 0.00, 95% CI [-3.34, -1.21], for Task_Difficulty = 2, $\beta$ = -1.20, $t$(33.77) = -2.48, $p$ = 0.02, 95% CI [- 2.19, -0.22], for Task_Difficulty = 3, $\beta$ = -0.88, $t$(34.70) = -2.09, $p$ = 0.04, 95% CI [-1.74, -0.03] and for Task_Difficulty = 4, $\beta$ = 0.036, $t$(34.91) = 0.12, $p$ = 0.90, 95% CI [-0.56, 0.63].

Next, to examine whether the valence of (OASIS image-induced) emotions would be reflected in physiological data, relationships between the latter and smiley meter ratings for images were analysed. We found no significant relation between smiley meter ratings and tonic skin conductance levels [$F$(4, 677.72) = 1.63 , $p$ = 0.17 ] , blood volume pulse [$F$(4, 654.02) = 0.97 , $p$ = 0.42 ], heart rate [$F$(4, 683.29) = 1.66 , $p$ = 0.16 ] and skin temperature [F(4, 676.20) = 1.37 , $p$ = 0.24 ]. Feature extraction from phasic skin conductance data corresponding to the image stimuli resulted in no significant SCRs for practically the whole dataset (except 1 to 2 images of few participants).

Finally, we also evaluated the stimuli, i.e., examined whether participants perceived the programming tasks as they were intended to be (namely, task 1 – moderately challenging and positive-emotion inducing, task 2 – too easy, negative-emotion inducing, and task 3 – too difficult, negative-emotion inducing). We used linear mixed models while controlling for demographic data. Results indicated significant differences in Chal2Skill ratings [$F$(2, 40) = 43.59, $p$ = 0.00]. The average Chal2Skill rating for the moderately challenging task exceeded that of the difficult task (mean difference = 1.43, $p$ = 0.00), while that of the easy task was greater than that of the moderately challenging (mean difference = 0.76, $p$ = 0.01) and difficult task (mean difference = 2.20, $p$ = 0.00). We found significant differences in Task_Difficulty ratings [$F$(2, 39) = 40.97, $p$ = 0.00]. As expected, Task_Difficulty ratings for the difficult task were greater than those of the moderately challenging task (mean difference = 1.86 , $p$ = 0.00) and easy task (mean difference = 2.60, $p$ = 0.00), while ratings for the moderately challenging task were higher than those for the easy task (mean difference = 0.75, $p$ = 0.05). No significant differences in Task_Absorption ratings were found [$F$(2, 40) = 2.15, $p$ = 0.13]. We also found no significant differences in smiley meter ratings for the different tasks, $F$(2,46) = 1.14, $p$ = 0.33. This is corroborated by the interviews in which several participants exhibit recall bias at the time of responding to the smiley meters. For example, one participant provided a low smiley meter rating despite having enjoyed the task simply because they felt disappointed at not being able to complete it on time. In another case, a participant displayed agitation through most of the task period but gave a high rating because they managed to understand the task towards the end. Consequently, smiley meter ratings for the tasks were not included in any other analyses. During the interviews, some words used to describe experiences during the moderately challenging task were "confused", "challenging", "enjoyable" and "fun". Some participants (n = 5) described feeling slightly stressed or frustrated when they could not find a solution at the beginning, but feeling better afterwards. Some (n = 4) displayed disappointment at not being able to complete the task. Talking about the easy task, most participants (n = 13) mentioned its repetitive nature or described being bored at some point during the task. While describing their experience during the difficult task, most participants (n = 11) mentioned frustration, annoyance, a sense of hopelessness or incompetence.

## 5. Discussion

In this study, we attempted to detect physiological indicators of learning related emotions by using multimodal data from a biosensing wristband and self-reports. To this end, we presented participants with an easy, moderately challenging and difficult task with the expectation that these would be associated with different emotions. It was expected that during the easy and difficult tasks, participants would experience negative emotions (boredom/frustration/anger). This negative emotional state would be associated with a combination of low blood volume pulse and either low skin conductance and heart rate, or high skin conductance and low heart rate. We also expected that during the moderately challengingtask, participants would experience a positive emotional state (i.e., enjoyment), which in turn would be associated with high blood volume pulse, skin conductance and heart rate. Results show that participants in general had lower phasic skin conductance and heart rate during the difficult task as compared to the moderately challenging task. In fact, heart rate during the moderately challenging task was also higher than that during baseline and the easy task. On the other hand, no significant differences in blood volume pulse were found. Thefindings of high heart rate and phasic skin conductance during the moderately challenging task align with our expectation of indicators of enjoyment. Similarly, low phasic skin conductance, tonic skin conductance and heart rate during the difficult task could indicate boredom. While we did not see high skin temperatures during the difficult task as expected, indications of high skin temperature and tonic skin conductance levels during the easy task could indicate anger [26], [27].These indications of enjoyment, boredom and anger also align with our expectations based on the control-value theory [6]. However, a comparison with self-reports and certain limitations of the study (discussed below) suggest that more evidence is required to ascertain whether all these physiological changes are indeed due to the emotional stimuli.

The biggest limitations of this study are the fixed order of the programming tasks and a lack of sufficient evidence to ascertain clear relationships between all the physiological signals and self- reports. Therefore, we cannot write off order-effects and there is a great likelihood that the changes in physiological signals are simply due to the passage of time. Also, there is the issue of obtaining clear self-reports on emotions. In this study, data fromthe smiley meters did not add value to the analysis. The decision to use a smiley meter was to ensure that we did not put words into participants' heads. However, this resulted in not having direct measures of learner emotions and having to make inferences based only on learner appraisals of task difficulty, challenge to skill balance and task absorption. We also gathered that the 10 minute intervals between smiley meter ratings on the programming tasks were likely too long as several participants displayed recall bias. Since these limitations warrant further research, in our next study, we will tweak our design to ensure increased reliability of our findings. Firstly, we plan to randomise the order of tasks for each participant. And secondly, we will collect regular and intermittent reports during the task (for example, every 3 to4 minutes) on a more sophisticated scale such as the Affect Grid [28]or Self-Assessment Manikin [29].

The use of physiological measures of emotion detection has important theoretical and practical implications. As mentioned earlier, the vast majority of studies in learner emotion have utilized self-reported data [10]. These include the building of significant educational theories such as [6]. An approach utilizing multimodal data including physiological data (such as whatwe do in this study) opens up the possibility to test such theories in a more robust manner and advance our knowledge base on learner psychology. Additionally, such studies take us closer towards realizing intelligent systems that can detect and therefore cater to the emotions of learners. The results of the present study thus contribute towards the field of emotions in learning.

## 6. Conclusion

In the present study, we found indications that certain learner emotions related to different task difficulties may possibly be characterised by a combination of phasic and tonic skin conductance, heart rate, and skin temperature. Such a psychophysiological approach to emotion detection can open the doors to real- time adaptive support that can bring learners to their zone of proximal development and consequently greatly improve learning outcomes. Therefore, though the results of the present study are far from definitive, we see value in advancing research in this area. Our next steps include a) furthering our exploration of signals collected from the E4 after including design changes derived from this study, b) exploring other nonintrusive measures of learner engagement such as camera based eye tracking and screen activity, c) developing a multimodal system of emotion detection, d) prototyping an adaptive system based on affective feedback.

## 7. Acknowledgements

This study was funded by the BMS Signature PhD grant at the University of Twente and made possible with the technical assistance of the university's BMSLab. We would like to thank André Bester and Lucia M. Rabago Mayer for the Python code, setting up the Unity platform and other technical guidance. We would also like to thank Johannes Steinrücke for his guidance on statistical methods, especially linear mixed models.

## 8. References

[1] B. Mihaly Csikszentmihalyi, Flow: The Psychology of Optimal Experience, Harper & Row, New York, 1990.

[2] Kort, Barry, Rob Reilly and Rosalind W. Picard, An affective model of interplay between emotions and learning: engineering educational pedagogy-building a learning companion, in: Proceedings IEEE International Conference on Advanced Learning Technologies, 2001, pp 43-46. doi: 10.1109/ICALT.2001.943837

[3] R. E. Mayer, Searching for the role of emotions in e-learning, Learn. Instr. 70 (2019). doi: 10.1016/j.learninstruc.2019.05.010.

[4] R. Pekrun, The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice, Educ. Psychol. Rev. 18 (2006), pp. 315–341. doi: 10.1007/s10648-006-9029-9.

[5] E. R. Um, J. L. Plass, E. O. Hayward, B. D. Homer, Emotional design in multimedia learning, J. Educ. Psychol. 104 (2012), pp. 485–498. doi: 10.1037/a0026609.

[6] R. Pekrun, S. Lichtenfeld, H. W. Marsh, K. Murayama, T. Goetz, Achievement Emotions and Academic Performance: Longitudinal Models of Reciprocal Effects, Child Dev. 88 (2017), pp. 1653–1670. doi: 10.1111/cdev.12704.

[7] K. Loderer, R. Pekrun, J. C. Lester, Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments, Learn. Instr. 70 (2020). doi: 10.1016/j.learninstruc.2018.08.002.

[8] K. S. Beard, Theoretically Speaking: An Interview with Mihaly Csikszentmihalyi on Flow Theory Development and Its Usefulness in Addressing Contemporary Challenges in Education, Educ. Psychol. Rev. 27 (2015). doi: 10.1007/s10648- 014-9291-1.

[9] L.S. Vygotsky, Mind in society: The development of higher psychological processes, Harvard University Press, Massachusetts, 1978.

[10] C. H. Wu, Y. M. Huang, J. P. Hwang, Review of affective computing in education/learning: Trends and challenges, Br. J. Educ. Technol. 47 (2016). doi: 10.1111/bjet.12324.

[11] E. Yadegaridehkordi, N. F. B. M. Noor, M. N. Bin Ayub, H. B. Affal, N. B. Hussin, Affective computing in education: A systematic review and

future research, Comput. Educ. 142 (2019). doi: 10.1016/j.compedu.2019.103649.

[12] E. H. Jang, M. S. Park, B. J. Park, S. H. Kim, M. A. Chung, J. H. Sohn, Relationship between affective dimensions and physiological responses induced by emotional stimuli: Base on affective dimensions: Arousal, valence, intensity and approach, in: Proceedings of PhyCS 2014 - Proc. Int. Conf. Physiol. Comput. Syst., 2014, pp. 254–259. doi: 10.5220/0004728302540259.

[13] Brouwer, A., Van Beurden, M., Nijboer, L., Derikx, L., Binsch, O., Gjaltema, C., Noordzij, M., A comparison of different Electrodermal variables in response to an acute social stressor, Symbiotic Interaction 7 (2018). doi: https://doi.org/10.1007/978-3-319-91593-7_2.

[14] M. Jindrová, M. Kocourek, P. Telenský, Skin conductance rise time and amplitude discern between different degrees of emotional arousal induced by affective pictures presented on a computer screen, bioRxiv (2020). doi: 10.1101/2020.05.12.090829.

[15] V. R. Lee, L. Fischback, R. Cain, A wearables-based approach to detect and identify momentary engagement in afterschool Makerspace programs, Contemp. Educ. Psychol. 59 (2019). doi: 10.1016/j.cedpsych.2019.101789.

[16] D. K. Darnell, P. A. Krieg, Student engagement, assessed using heart rate, shows no reset following active learning sessions in lectures, PLoS One 14 (2019). doi: 10.1371/journal.pone.0225709.

[17] C. Larmuseau, J. Cornelis, L. Lancieri, P. Desmet, F. Depaepe, Multimodal learning analytics to investigate cognitive load during online problem solving, Br. J. Educ. Technol. (2020). doi: 10.1111/bjet.12958.

[18] I. Arroyo, B. P. Woolf, W. Burelson, K. Muldner, D. Rai, M. Tai, A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect, Int. J. Artif. Intell. Educ. 24 (2014) 387–426. doi: 10.1007/s40593- 014-0023-y.

[19] M. Dindar, J. Malmberg, S. Järvelä, E. Haataja, P. A. Kirschner, Matching self-reports with electrodermal activity data: Investigating temporal changes in self- regulated learning, Educ. Inf. Technol. 25 (2020) 1785–1802. doi: 10.1007/s10639-019-10059-5.

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

[21] B. Kurdi, S. Lozano, M. R. Banaji, Introducing the Open Affective Standardized Image Set (OASIS), Behav. Res. Methods 49 (2017) 457–470. doi: 10.3758/s13428-016-0715-3.

[22] D.C. Menezes, 2019. Play Mode Blocks Engine. URL: https://assetstore.unity.com/packages/te mplates/systems/play-mode-blocks-engine-158224.

[23] Read JC, MacFarlane SJ, Casey C, Endurability, engagement and expectations: measuring children's fun, in : Proceedings of the International Workshop Interaction Design and Children, Shaker Publishing, Eindhoven, 2002.

[24] T. Magyaródi, H. Nagy, P. Soltész, T. Mózes, A. Oláh, Psychometric properties of a newly established flow state questionnaire, 2013.

[25] Benedek, M., & Kaernbach, C., Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology 47* (2010) 647 – 658. doi: 10.1111/j.1469-8986.2009.00972.x.

[26] V. Jha, N. Prakash, S. Sagar, Wearable anger-monitoring system, ICT Express 4 (2018) 194–198. doi: 10.1016/J.ICTE.2017.07.002.

[27] G. Stemmler, Physiological processes during emotion, in: P. Philippot, R. S. Feldman (Eds.), The regulation of

emotion, 1st ed., Lawrence Erlbaum Associates Publishers, 2004, pp. 33–70. doi: 10.4324/9781410610898
[28] J. A. Russell, A. Weiss, G. A. Mendelsohn, Affect Grid: A single-item scale of pleasure and arousal., J. Pers. Soc. Psychol. 57 (1989) 493–502. doi: 10.1037/0022-3514.57.3.493.
[29] M. M. Bradley, P. J. Lang, Measuring emotion: The self-assessment manikin and the semantic differential, J. Behav. Ther. Exp. Psychiatry 25 (1994) 49–59. doi: 10.1016/0005-7916(94)90063-9.