Understanding Fun in Learning to Code: A Multi-Modal Data approach

Gabriella Tisza*
Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands
g.tisza@tue.nl

Kshitij Sharma
Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway
khitij.sharma@ntnu.nl

Sofia Papavlasopoulou
Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway
spapav@ntnu.nl

Panos Markopoulos
Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands
p.markopoulos@tue.nl

Michail Giannakos
Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway
michailg@ntnu.nl

ABSTRACT

The role of fun in learning, and specifically in learning to code, is critical but not yet fully understood. Fun is typically measured by post session questionnaires, which are coarse-grained, evaluating activities that sometimes last an hour, a day or longer. Here we examine how fun impacts learning during a coding activity, combining continuous physiological response data from wristbands and facial expressions from facial camera videos, along with self-reported measures (i.e. knowledge test and reported fun). Data were collected from primary school students (N = 53) in a single-occasion, two-hours long coding workshop, with the BBC micro:bits. We found that a) sadness, anger and stress are negatively, and arousal is positively related to students' relative learning gain (RLG), b) experienced fun is positively related to students' RLG and c) RLG and fun are related to certain physiological markers derived from the physiological response data.

CCS CONCEPTS

• **General and reference**; • **Cross-computing tools and techniques**; • **Empirical studies**; • **Computing methodologies**; • **Machine learning**;

KEYWORDS

Fun, Learning, Programming, FunQ, Multimodal Learning Analytics (MMLA)

ACM Reference Format:
Gabriella Tisza, Kshitij Sharma, Sofia Papavlasopoulou, Panos Markopoulos, and Michail Giannakos. 2022. Understanding Fun in Learning to Code: A Multi-Modal Data approach. In *Interaction Design and Children (IDC '22), June 27–30, 2022, Braga, Portugal.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3501712.3529716

*Corresponding author

IDC '22, June 27–30, 2022, Braga, Portugal

ACM ISBN 978-1-4503-9197-9/22/06.
https://doi.org/10.1145/3501712.3529716

1 INTRODUCTION

Coding skills are gaining increased attention especially as they are often considered as a core literacy skill of the 21st century [35]. Accordingly, more and more countries are introducing computer science (CS) and coding competence to their curricula[1]. Despite this ongoing momentum and development of new CS courses (e.g. [24]), currently, children's participation in out of formal education activities is the main way children obtain competence in CS and coding. Designing fun and engaging learning activities is essential to attract children as fun provides the affective coloring for all our day-to-day events and interactions.

In the field of interaction design and children, evaluation of fun has been largely focused on self-reported data from children, asking them to assess specific activities in single-item scales or to compare the experienced fun in relation to different elements of the design [38]. This pragmatic and widely used approach addresses the difficulties children have in responding to surveys but does not provide a theoretically grounded definition of fun and a corresponding psychometrically validated measurement. Such an undertaking is reported by Tisza and Markopoulos [53], where a theoretical account of the nature of fun as an affective state is proposed together with FunQ, a validated questionnaire for measuring the fun children experience during learning. According to their definition, fun is an affective state during which one feels in control, loses the perception of time and space, lets off social inhibitions, meets the appropriate level of challenge, and is dominated by positive emotions while the levels of negative emotions remain as low. This approach has arguably a better theoretical foundation and is more reliable than single item measures, but still, suffers from being coarse grained, providing a single retrospective measure for a whole activity rather than a measure that considers fun as a changing state that varies over time and in relation to the momentary activity of the child.

Despite the great promise on designing coding activities that can be both instructional and perceived as fun, there are several

[1]https://www.euractiv.com/section/digital/infographic/infographic-coding-at-school-how-do-eu-countries-compare/

challenges with this endeavor. First, there are various available methods used to measure children's affect in design research, with limited agreement among researchers about the definition and an acceptable measurement of fun. Moods and emotions, as well as human's affective preferences (i.e., what someone likes or dislikes) are complex constructs, and different methods have been developed to understand and measure them. Three broad categories are the following:

- Methods that rely on automatic affect recognition (e.g., objective signals portraying specific physiological and behavioral response patterns that represent emotions) and inspired by theories (embodiment of affect) [16, 17].
- Methods that rely on self-report (e.g., questionnaires, rankings), such items can be of verbal or pictorial scales (e.g., Smileyometer).
- Methods that rely on text or discourse analysis (can be automated via natural language processing methods or via human content analysis).

The three categories have different strengths and weaknesses but can also co-exist allowing us to capture different aspects of children's mood and emotions and understand their affective preferences comprehensively (although the third category is not relevant for this study, since there was no discourse and text). Despite the great interest in designing fun learning activities, as yet there is little known regarding the impact of fun on learning.

To further contemporary approaches and understand children's affective preferences comprehensively, we adopt the use of a multimodal approach. In particular, our approach involves the use of objective automated measures coming from children's physiological response data (collected by wristbands and facial video recordings), self-reported fun and their learning gain (via a standard test). This approach has been proven to be effective, among others, for predicting cognitive performance [44], and hence indicated that using physiological response data allows us a new level of examination. However, we know little about the nature of fun in learning, and fun in coding activities has never been examined previously from the physiological perspective. This study aims to fill the gap in the literature by investigating the relationship between the experienced fun while learning how to code (as self-reported), the learning outcomes (based on standard tests) and children's affective states derived from unobstructive subjective measurements. In particular, this study focuses on the following research questions (RQ):

RQ1: What is the relationship between children's learning and their affective states (i.e., affect from the Action Units (AUs), physiological stress and arousal) and processes during a coding activity?

RQ2: What is the relationship between children's perceived fun (as measured by FunQ) and their affective states and processes during a coding activity?

To tackle the aforementioned RQs, we designed a 2-hour-long playful coding workshop (introducing coding with BBC micro:bits) and implemented it in six primary school classes. Our findings indicate that both students' learning (i.e., relative learning gain - RLG) and the level of fun they have experienced while coding are associated with specific set of physiological predictors. On top of that, we also found a positive and significant association between fun and students' RLG. To summarize, we present the following contributions:

(1) We offer insights from a study where children, aged 8-12 years, participated in a coding workshop and their experience and learning were captured by standardized tests and physiological devices.
(2) We identify the relationship between children's learning, perceived fun and affective processes (captured using the transitions among the affective states) during the coding activity.
(3) We discuss how our approach and findings can be used to design future coding workshops.

2 RELATED RESEARCH

2.1 Affective Processes and Learning

Pekrun [37] introduced the Control-Value Theory (CVT) of achievement emotions by integrating assumptions from expectancy-value approaches to emotions, theories of perceived control, attributional theories of achievement emotions, and models that involve effects of emotions on learning and performance. More specifically, Control-Value Theory builds on the idea that experiencing emotions during learning is dependent on whether learners consider the learning activity important, and the extent to which learners have control over the achievement activities and outcomes [18]. Accordingly, emotions can be mapped on a two-dimensional plot based on their valence and activation, and thus we can distinguish positive activating (e.g. enjoyment, curiosity), negative activating (e.g. frustration, confusion), positive deactivating (e.g. relief, relaxation), and negative deactivating (e.g. boredom) emotions. In a recent systematic literature review on emotions in the technology-based learning environment, Loderer et al [29] found that research into emotions almost quadrupled in the past 20 years. Among the reviewed papers in this study, anxiety is still the most studied academic emotion, while enjoyment has become the second most frequently investigated one. They also found enjoyment to be positively related to engagement, learning strategy use, curiosity/interest and to learning outcomes as; but it was found to be negatively related to disengagement. In contrast, anxiety was negatively associated with both students' engagement, strategy use and learning outcomes. Similarly, a systematic literature review of emotions in design-based learning [58] classified emotions reported in empirical studies according to the typology of emotions introduced with the Control-Value Theory (achievement, epistemic, topic and social emotions). With very few exceptions, the studies reviewed sought for indications of enjoyment as a positive aspect of the learning activity, though the evidence on the expected positive impact of enjoyment or fun on learning engagement with the topic was found to be equivocal. It is noticeable that fun and enjoyment are terms often used interchangeably in design research, with fun being regularly adopted as an evaluation criterion for learning games (e.g. [38, 46]). In relation to the students learning experience during a coding activity, this research focuses on the four Control-Value Theory emotions - happiness, sadness, anger, and surprise together with physiological stress and arousal to capture the affective states and processes; while for capturing fun the already discussed definition was adopted [53]. These four CVT emotions were selected as other emotions (e.g., disgust, contempt, relief) accounted for less than 3% of the total interaction time.

2.2 Fun and Learning

Most research into the relationship between fun and learning is found in the context of educational technology. Sim et al [46] investigated fun, usability and learning in an educational software with seven and eight year-old children and found no correlation between learning (measured by the difference between the post- and pre-test scores) and fun or between fun and usability. In line with their findings, Iten and Petko [22] studying a serious game with child participants aged between 10 to 13 also found no significant relationship between fun and learning (perceived learning and measured learning calculated as the summary score for the current answers for the pre- and post-test). Contrasting these earlier findings, Tisza, Zhu and Markopoulos [55] found when investigating a serious game with 14-15 years old students that having fun while learning had a significant effect on students' perceived learning, but not on their measured learning (when considering their pre and post outcomes). Lucardie [31] investigating fun and enjoyment in adults' learning found that both learners and their teachers perceived fun as a motivator to attend classes, and as a contributor to learning knowledge and new skills. Related to coding, Long [30] studying the influence of a programming game on learners' (mostly adults) programming skills and knowledge found that about 80% of the study participants reported a perceived increase. They also found that 87.5% of the study participants joined because of the anticipated fun of the learning game, making fun a strong motivator for participation. Therefore, having an enjoyable or amusing time (i.e. fun) while learning is a significant predictor for the learning effort. Saez-Lopez et al [40] found in a two-year long study that using Scratch - a block-based visual programming interface - to teach programming to primary school children significantly increased students' knowledge; furthermore, students reported on having fun while learning with Scratch, a finding that is also supported by the observers' reports. However, that study did not examine the relationship between fun and learning. Complementary to this, Tisza and Markopoulos [52] found that having fun while learning to program contributed significantly to the perceived learning of primary school children. As it is seen, there is a variety in earlier research in terms of the assessment of learning. While some examined perceived learning [30, 31, 52], others investigated measured learning (i.e., pre-post tests) [40, 46, 55]. To our best knowledge, there are only a few studies that report on the relationship between fun and both the perceived and the measured learning [22, 54, 55]. All of those studies report on regression analyses, indicating a slight difference between the two measures, however, without directly comparing them. Accordingly, understanding the relationship between the perceived and the measured learning, especially in reflection of the experienced fun while learning remains a topic to be addressed by future research. In sum, while earlier research appeared to be inconclusive on the role that fun plays on learning, recent empirical research results are supportive that fun contributes positively to the learning outcomes. This shift is proposed to be due to a better understanding of the notion of fun and accordingly, improved ways for the assessment of it.

2.3 Multimodal Data and Learning

Learning is a complex process and involves cooperation and coordination of several cognitive processes (e.g., information processing, creating, maintaining and updating mental schemas) and affective mechanisms (e.g., frustration, boredom, confusion, stress, arousal;[47]). These processes and mechanisms could incur an affective disequilibrium that might be detrimental for learning, when students struggle to maintain and update their existing mental models with new information [19]. Given the range of processes involved, it would make sense that a single data stream would not be able to capture all these processes. Depending on the process of interest, different data streams may be more appropriate. Some of these data streams currently used within education include video, system logs, and physiological response data such as, electrodermal activity, heart rate variability, blood volume pulse, and skin temperature. Individually, these data streams have been used to explain and predict aspects of the cognitive processes and affective mechanisms [45]. By extending these findings into interventions, researchers have used the data streams to scaffold the learning process to provide better learning support to students.

Given that a single data stream cannot capture all processes happening during learning activities as each data stream can only provide a partial view when used on its own, an upcoming field of research, multimodal learning analytics (MMLA), combines several of these data streams to serve as a virtual observer and analyst of learning activities [5, 45]. MMLA provide an unprecedented opportunity to understand students' behavior and performance during and after the learning sessions by understanding their relations with cognitive processes and affective mechanisms [12]. MMLA can provide insights into a multitude of behaviors including reasoning patterns [57], short-term memory usage [25], artefact quality [49], help-seeking and help-giving behavior [13], tentative and casual problem-solving behavior [2], and problem-solving phases [3, 48]. MMLA can be used to differentiate and identify different learning processes and behaviors [48, 57], as well as to explain the relationship between two behaviors, such as a student's physical actions and their reasoning during learning [2]. MMLA can shed light to learning processes that may be invisible to the human eye and that students cannot self-report on [13, 28, 35, 43]. Therefore, MMLA can complement our understanding on how children learn, providing more information on children's affective aspects during the coding activities.

3 METHODS

3.1 Participants

The herein introduced study was conducted in mid-February 2020 in the Netherlands. Primary school teachers across the country were approached to participate in the study. We recruited 53 children (M_{age} = 10.13 yrs, SD = 1.103, 27 boys, 26 girls) from three schools and six school classes. Participation in the activity was compulsory for the children as the workshop took place during school hours, however, participation in the study (i.e. responding to the questionnaires and allowing us to capture their screens and cameras) was voluntary. Given children's age, informed consent was obtained across the schools from both the children and their parents/guardians before the study started. The study was approved on

10 January 2020 by the ethics review board of Eindhoven University of Technology, Department of Industrial Design.

3.2 Procedure

A single-occasion, two-hours long workshop was designed in collaboration with the SkillsDojo Foundation to introduce coding with BBC micro:bits in a playful way. The workshop consisted of five main sections. First, the pre-workshop data collection section, then three distinct coding tasks and the workshop ended with the post-workshop data collection. Both the pre- and post-workshop data collection took approximately 10 minutes. Children had approximately 90 minutes to spend on the coding tasks. The coding tasks were guided by the videos provided by the SkillsDojo Foundation (see the detailed description at section 3.3). The first coding task had an introductory nature, during which children learned the basic properties of the micro:bits and thereafter they learned to program their names. In the second task, children programmed a stone-paper-scissors game. In the third task, children could create a micropet that reacted to kinetic stimuli (guided by the instructional video) or they could choose to create their own code. By their nature, the coding tasks required individual work, however, collaboration was also allowed and facilitated by the researchers.

The workshop took place in the classroom as an extra-curricular activity. During the workshop each child was equipped with a laptop to follow the instructional videos and practice coding, moreover everyone had their own micro:bits. Further materials, such as pencils, scissors, glue etc. that could be used for the micropet task were provided by the schools. During the workshop children were seated behind their laptops, however, they were allowed to move around freely and interact with each other.

3.3 The How-To Instructional Videos

The how-to instructional videos used in this study to guide children in the learning tasks were developed by the SkillsDojo Foundation according to the Creative Learning Model of the Lifelong Kindergarten research group at MIT [39]. Accordingly, the videos have a *low floor* and *high ceiling* making it easy to begin with but providing plenty of room to be constantly challenging. The videos are also designed in a way that makes the topic relevant to children by linking content to their everyday life, they are following segmentation principles, so that the videos are built up from comprehensive 'chunks' using signaling to highlight the most important parts. Additionally, the videos used working examples as part of the instructions.

3.4 Data Collection

To address the research questions, multimodal data were collected. Alongside with children's demographics, we collected their self-reported fun via questionnaires, their pre- and post- coding competence via a test, facial expressions from facial videos and physiological arousal and stress from wristband sensors.

In particular, the pre-workshop questionnaire, captured children's demographics and background information that included their perceived experience and knowledge on coding, using a 5-point Likert scale ('Do you have any idea about programming?' (1) not at all - (5) I'm a pro; 'How many programming activities have you participated before?' (1) none - (5) six or more; see Figure 1).

To assess their coding competence, we employed a pre- and post-workshop knowledge test. This allowed us to assess children's competence levels before and after their participation in the workshop, and calculate their RLG (*knowledge* level of Bloom's taxonomy [6]). The test was developed specially for the purpose of the study to cover the material of the how-to videos. It consists of seven multiple-choice questions with four response options, from which four ask about terms that are explained in the videos (e.g. What/who is a variable?) and three questions address the working of a piece of code – which are necessary to complete the programming tasks (see example in Figure 1; this piece of code is part of the stone–paper–scissors game (task 2)).

For the assessment of fun during the workshop, we employed the FunQ [53] instrument as part of the post-workshop questionnaire. FunQ is a validated instrument in several languages (including Dutch) and consists of 18 easy to understand (considering children's age) questions.

Besides using questionnaires, we collected children's physiological response data. We collected arousal data via wristbands and facial expression via facial cameras data of 11 randomly selected children in each workshop, thus from 66 children in total. However, data from 13 children was damaged or lost during recording, hence our data set used for the analysis comprises of data from 53 children. Those multimodal data were collected while the children were engaged with coding tasks. Regarding the data collection of the different data modalities, we used the Empatica E4 wristband to capture children's physiological response data consisting of 4 different variables: Heart rate variability (HRV, 1Hz), Electrodermal Activation (EDA, 64Hz), skin temperature (4Hz), and Blood Volume Pulse (BVP, 4Hz) and for the facial video we used the web camera of each laptop the children were working on. The frame rate was set to 24 frames per second.

3.5 Measurements

Relative Learning Gain (RLG): From the pre and post knowledge acquisition test, we calculated children's RLG that has been used previously in similar studies [34]. This measure is more accurate than typical learning gain, since it considers children's initial knowledge when assessing learning gain and avoids potential floor effects. RLG captures how much students learn beyond what they knew prior to the intervention.

$$RLG = \begin{cases} \frac{posttest - pretest}{Max.pretest - \Pr estest}, \; if\ Posttest \geq \Pr estest \\ \frac{Posttest - \Pr etest}{\Pr etest}, \; if\ Posttest < \Pr etest \end{cases}$$

Fun Dimensions: FunQ instrument [53] was employed to measure the experienced fun along its six dimensions, FunQ has eighteen questions (items), and it uses a 5-point Likert scale. The six dimensions are Autonomy (perceived control over participation and the activity itself), Challenge (experienced challenge), Delight (perceived positive emotions and related desires), Immersion (perceived loss of time and space), Loss of Social Barriers (perceived social connectivity), and Stress (perceived negative emotions). More details about the instrument and its evaluation with non-adult population can be found here [72].

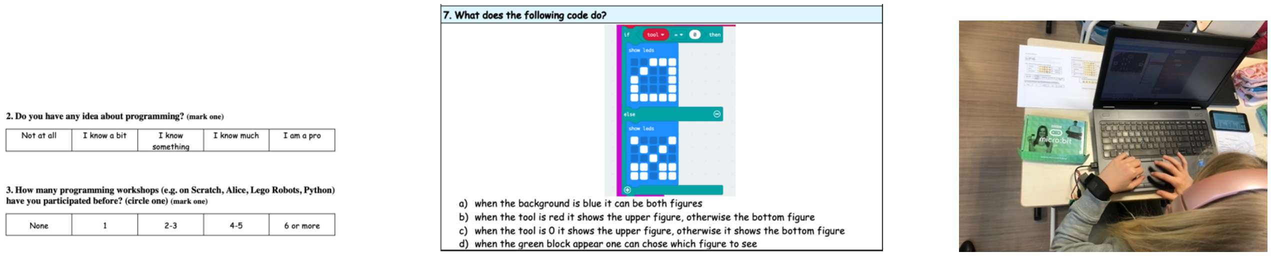

Figure 1: *Left*: Example of the questionnaire. *Middle*: Example of the knowledge test. *Right*: Setup during the workshop.

Table 1: Defining emotions from Action Units.

Affective state	Action units	Affective state	Action units
Happiness	AU6, AU10	Anger	AU1, AU2, AU5, AU26
Sadness	AU1, AU4, AU15	Surprise	AU4, AU5, AU7, AU23

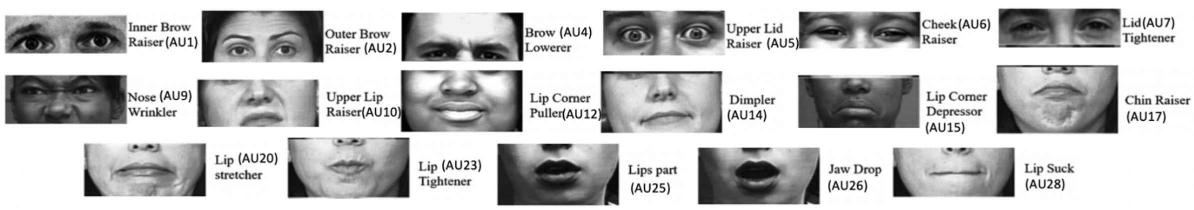

Figure 2: Action Unites detected for this paper. The facial images are taken from https://www.cs.cmu.edu/~face/facs.htm The action unit number are mentioned in parentheses next to the action unit names to have a mapping with Table 1.

Affect: We used the face images coming from the videos to extract the facial Action Units (AUs, [15]) using the OpenFace framework [1]. Facial Action Coding System (FACS) is a taxonomy for human facial movements as they appear on the face. Movements of individual facial muscles are encoded by FACS from slight instant changes in facial appearance. Using FACS it is possible to code nearly any anatomically possible emotion, deconstructing it into the specific AU that produced the facial expression. FACS is an established scheme for coding facial expressions, which is supported by multiple studies that have evaluated FACS with positive results with adults [11, 41]. Additionally, studies used the scheme in the previous years with children with positive results as well [35, 36]. Furthermore, it is a common standard to objectively describe emotions from facial expressions using such techniques [56]. Figure 4 shows the AUs detected for this paper and Table 1 shows how to define emotions from the AUs. In this paper, we are using the proportion of each emotion during the coding activity. We define happiness, sadness, anger, and surprise from the action units (shown in table 1). These affective states are a subset of achievement emotions included in Control-Value Theory [37]. These four CVT emotions are use because other emotions (e.g., disgust, contempt, relief) make less than 3% of the total interaction time. Therefore, we discarded the emotions that are not detected with a significant proportion of the interaction time. The interpretation of facial expressions can change from one situation to other however, the coding is well-evaluated and the qualitative interpretation in the context we studied will be done in our future analysis. This study focused more (being the first of its kind, to the best of our knowledge) on finding the relationships between CVT-based emotions and sensor and facial data.

Affective states transition: the second set of measurements were the transition probabilities between two affective states. These transitions capture the affective process during the coding activity. We did not consider the self-loops in this paper, because we are already using the proportion of the duration of each individual emotion as the first set of measurements.

Physiological Stress: This is computed as the heart rate's increasing slope. The more positive the slope of the heart rate is in a given time window, the higher the stress is [51]. Heart rate has been used to measure stress in educational [45] and problem-solving [32] contexts. In the rest of the paper, physiological stress is referred to as stress among the physio-affective states and processes.

Physiological arousal: EDA signal is comprised of two parts: the tonic and phasic components. The tonic component of the EDA signal is the one with slow evolving patterns. The phasic component of the EDA signal is the one with rapid changes and is found to be related to physiological arousal [26]. In this paper, we consider only the mean phasic EDA component as a measure of physiological arousal. In the rest of the paper, physiological arousal is referred to as arousal among the physio-affective states and processes.

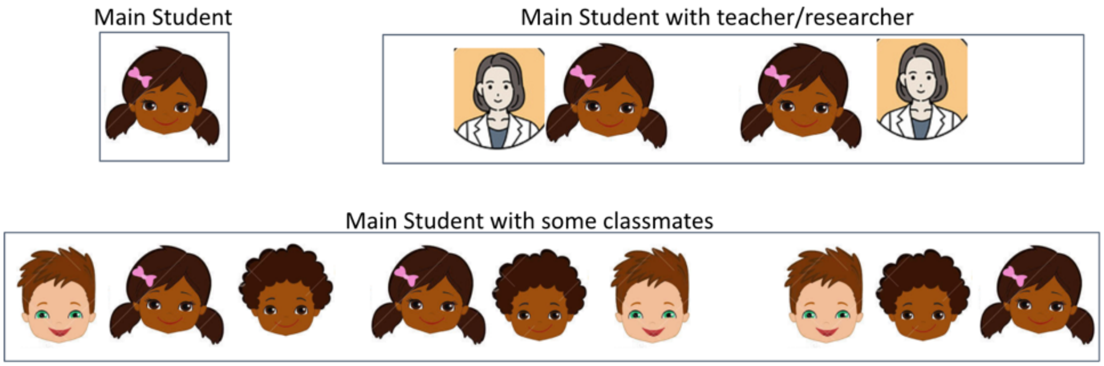

Figure 3: Facial data pre-processing. Detecting the main student.

3.6 Data Pre-processing

To remove noise, and potential conditional biases from the sensor data, the following pre-processing was conducted.

Wristband data: A simple smoothing function was used to remove any unwanted spikes in the time series in the 4 data streams originating from the E4 wristband (HRV, EDA, Skin Temperature, and BVP). This was a simple running average with a moving window of 100 samples, and an overlap of 50 samples between two consecutive windows. Physiological response data, such as HRV, BVP, and skin temperature, are susceptible to many subjective and contextual biases. These biases include the time of the day, physical health condition, gender, age, overnight sleep, and others. All 4 data streams were normalized using the first 30 seconds of the data to remove the subjective and contextual biases from the data.

Facial data: For most of the frames in the video recordings, only one face was visible. However, sometimes the researcher overseeing the activity appeared in the field of view of the camera. For some other frames there were a few other children in the frame as well (visualized in the Figure 3). First, we used the OpenFace [1] library in the videos, in order to detect the faces for every frame. Thus, each face is given a label starting from left to right (1 to N, where N is the number of faces in each frame). There are three cases where the left-to-right labeling of faces fails as shown in Figure 3. First, when students are with the teacher and/or the researcher. Second, when classmates join the student for a short time. We need to keep the face to which the recording belongs. To achieve this, we used a pre-trained deep neural network, INCEPTION-v4 [50], to extract features from the individual face images and used a k-nearest neighbor prediction algorithm to recognize the original student in every recording. Figure 3 shows the example for all the three cases. The first few minutes are used to create the feature vectors for the original student in each recording.

3.7 Data Analysis

First, to get a better idea of the results of our study, a descriptive and correlational analysis has been conducted on the main variables. Then, to address our RQs, appropriate data analyses have been conducted.

To examine possible predictors of children's learning from their physio-affective states (i.e. affect from AUs and physiological stress and arousal; RQ1), multiple regression equations were calculated. We use the RLG as the dependent variable and all the measurements from the facial and wristband data as the regressors in a regression model. The adjusted R-square value of the models shows the variance of the RLG explained by the physio-affective variables. Second, we use a t-test to find which individual coefficients from the regression model contribute significantly to the dependent variables and explain the relationship between the RLG and the physio-affective states. For multiple t-tests, the p-values are corrected using a Bonferroni correction.

Similarly, to examine possible predictors of children's perceived fun (the dimensions measured by FunQ) from their physio-affective states (RQ2), a series of multiple regression equations were calculated (Table 3-5).

To check the gender bias in the data, we use one-way ANOVA with the physio-affective measurements, FunQ dimensions and RLG as dependent variable and gender as the independent variable. Regarding the age bias in the data, we use Pearson Correlation between age and the other measurements (i.e., physio-affective measurements, FunQ dimensions and RLG).

4 RESULTS

From the descriptive analysis, we found that just above half of the children in our study were novices to coding. For the question 'Do you have any idea about programming?', 22.6% of the children reported on having no idea about coding, and 35.8% of the children reported knowing a bit. The mean for this question is 2.38 (with 5 being the highest) (SD = 1.105) which also translates into 'knowing a bit'. As for the responses for the question 'How many coding workshops have you participated in before?', 43.4% of the children reported never having participated in a coding activity, and an additional 13.2% reported having participated in 1 coding activity only. The mean of the responses for this question is 2.25 (with 5 being the highest) (SD = 1.356). Therefore, some of the children who participated in our study had some previous experience with coding, and most of them had none or very limited. When it comes to children's relative learning gain it was on an average 0.61 (SD = 0.22, minimum = 0, maximum = 1).

Table 3: The model for RLG with Control-Value Theory affective states, transitions among them, arousal, and stress. This table shows only the significant terms.

	β	Error	T-value	P-value
Intercept	0.27	0.69	0.52	> 0.05
Sadness (sad)	**-1.37**	**0.009**	**-2.38**	**0.01**
Anger (ang)	**-1.56**	**0.003**	**-3.12**	**0.001**
Trans:hap <-> sup	**1.13**	**0.04**	**2.03**	**0.02**
Trans:sad <-> ang	**-1.88**	**0.004**	**-3.42**	**0.0006**
Stress	**-1.03**	**0.01**	**-1.78**	**0.04**
Arousal	**1.19**	**0.02**	**1.71**	**0.04**

Children's average FunQ score is 71.55 (SD = 9.756; Cronbach's alpha = 0.819, min: 50, max: 89), which is quite high if you consider that the possible scores range from 18 (lowest fun) to 90 (highest fun).

We also checked for any age- or gender-related biases for the RLG and the different FunQ dimensions (i.e., autonomy, challenge, delight, immersion, social barrier, stress). There was no correlation between the age of the children and their RLG or any of the FunQ dimensions. However, there was one exception. The social barrier was higher for boys than that for girls ($F[1,37] = 4.63$, $p = 0.03$, nine children had missing values). As we show in the main analysis that we did not find any significant relationship between the social barrier and physio-affective states, this bias will not be discussed in the light of the results reported in this paper.

4.1 Results from Modeling the Relative Learning Gain (RQ1)

We modeled the relative learning gain (RLG) using the proportions of emotions, the transition among them, stress, and arousal. The overall model was significant ($F (10, 37) = 10.41$, $p < 0.001$, $R^2 = 0.72$), accounting for 72% of explained variance in children's RLG. We found *arousal* and the *transitions between happiness and surprise* to be positive predictor for RGL, while *sadness*, *anger*, *stress*, and *transition between sadness and anger* contributed negatively to students' RLG. The coefficients of the significantly contributing predictors are in Table 3, the complete model is to see in Appendix 1.

4.2 Results from Modeling the FunQ Dimensions (RQ2)

We modelled the FunQ Total Score and all the dimensions using the proportions of emotions, the transition among them, stress, and arousal. Below are the details for each of the dependent variables.

FunQ Total Score: The overall model for the total score of FunQ was not significant ($F (10, 37) = 1.44$, $p = 0.20$, $R^2 = 0.26$). We have provided the model details in Appendix 1. In other words, from the used physio-affective states we could not predict the total score of the FunQ.

FunQ Autonomy: The overall model for the Autonomy dimension of FunQ was not significant ($F (10, 37) = 1.47$, $p= 0.19$, R-sq = .32). We have provided the model details in Appendix 1. In other words, from the investigated physio-affective states we could not predict the FunQ Autonomy scores.

FunQ Challenge: The overall model was significant ($F (10, 37) = 10.19$, $p < 0.001$, $R^2 = 0.71$), accounting for the 71% of explained variance in children's FunQ Challenge. H*appiness*, *anger*, *arousal*, and the *transitions between happiness and sadness* predict FunQ Challenge positively. On the other hand, *sadness*, *surprise*, and *transition between sadness and surprise* contribute negatively to FunQ Challenge. The coefficients of the significantly contributing predictors are shown in Table 4.

FunQ Delight: We found the overall model to be significant ($F (10, 37) = 9.93$, $p < 0.001$, $R^2 = 0.65$). The predictor model accounts for 65% of the explained variance in children's FunQ Delight. In details, *happiness* and *surprise* are positive predictors for FunQ Delight whereas, *stress*, and *transition between happiness and anger* contribute negatively to FunQ Delight. The coefficients of the significant predictors are shown in Table 4.

FunQ Immersion: The overall model was significant ($F (10, 37) = 8.16$, $p < 0.0001$, $R^2 = 0.63$), accounting for the 63% of explained variance in children's FunQ Immersion. H*appiness* and *arousal* are positive predictors for FunQ Immersion while the *transition between sadness and anger* contribute negatively to FunQ Immersion. The coefficients of the significant predictors are shown in Table 5.

FunQ Social Barrier: The overall model for the social barrier dimension of FunQ was not significant ($F (10, 37) = 1.09$, $p= 0.39$, $R^2 = 0.17$). We have provided the model details in Appendix 1. In other words, from the investigated physio-affective states we could not predict the Social Barrier dimension of FunQ. As we mentioned earlier, there was a gender bias for this sub-construct. Boys (mean = 9.00, SD = 3.22) reported a higher social barrier than girls (mean = 7.11, SD = 1.99). However, because there is no relationship between this construct and the RLG or any other physio-affective measurements, we will not explore this bias in this contribution.

FunQ Stress: We found the overall model to be significant ($F (10, 37) = 10.02$, $p < 0.001$, $R^2 = 0.70$), accounting for the 70% of explained variance in children's FunQ Stress. We found that *sadness*, *anger*, *stress*, and *transitions between sadness and anger* predict FunQ Stress positively, while the *transition between happiness and surprise* contribute negatively to FunQ Stress. The coefficients of the significant predictors are shown in Table 5.

Table 4: The models for FunQ Challenge and Delight with control-value theoretic affective states, transitions among them, arousal and stress. This table shows only the significant terms (p < 0.05).

Model Challenge	β	Error	t	***Model Delight***	β	Error	t
intercept	0.31	0.12	0.43	intercept	0.44	0.24	0.54
Happiness	1.41	0.005	2.92	Happiness	1.34	0.051	2.13
Anger	1.02	0.004	3.13	Surprise	1.93	0.009	3.19
Sadness	-1.32	0.012	-2.10	Trans. Hap-Ang	-0.93	0.001	-4.18
Surprise	-1.35	0.003	-2.44	Stress	-1.34	0.003	-3.53
Trans. Hap-Sad	0.99	0.001	3.38				
Trans. Sad-Sup	-1.45	0.001	-3.04				
Arousal	1.23	0.014	2.23				

Table 5: The model for FunQ Immersion and Stress with control-value theoretic affective states, transitions among them, arousal, and stress. This table shows only the significant terms (p < 0.05).

Model Immersion	β	Error	t	***Model Stress***	β	Error	t
intercept	0.19	0.89	0.52	intercept	0.16	0.21	0.89
Happiness	1.46	0.001	4.34	Sadness	1.28	0.005	2.48
Trans. Sad-Ang	-1.73	0.003	-3.28	Anger	0.94	0.001	3.32
Arousal	2.01	0.001	4.26	Trans. sad-ang	1.27	0.017	2.70

5 DISCUSSION

In this study we set out to investigate the relationship between children's coding learning, the experienced fun, and their physio-affective states during a coding activity. We collected data from a questionnaire, and physiological response data collected by wristbands and facial video recordings. Using data from different modalities and analyzing them we provide a novel approach as earlier research has been limited to either the investigation of affective states (e.g., by interviews or surveys (e.g., [52]) or physiological measures (e.g., [45]). By combining these we extended our current body of knowledge by adding a new, physiological level of understanding of learning procedures. One can argue that the two measurements are not exactly the same, which is evident by the results reported in the paper. We have shown that there is a significant overlap between the retrospective measurement of fun (through questionnaire) and the spontaneous measurement of affect (through sensor data). Both measurements have been evaluated separately [11, 41, 53]. This study is an attempt to find a relationship between the two measurements to have more real-time information about the semantic beliefs and memories using the sensor data. Accordingly, we found that RLG and most of the FunQ dimensions can be explained by the CVT affective states (i.e., happiness, sadness, anger, surprise, and the transition between these). Therefore, the introduced results indicate that there is a link between learners' affective states, their learning outcomes, and the fun they have experienced while learning.

More specifically, regarding students' learning and their affective states during coding activity (RQ1), we found that sadness, anger, and stress contribute negatively on students' learning, while arousal positively on it. This finding is in line with previous research that has investigated this relationship with traditional methods (i.e. questionnaires and observations) [29]. However, it also goes beyond those by applying MMLA and physiological measures.

Our research results indicate that from physiological data we could not predict the level of fun - measured as the total score on FunQ – that children experienced while learning to code. Nevertheless, we found that the total FunQ score significantly correlates with the RLG (Pearson correlation = 0.33, $p < 0.05$). This finding is in line with earlier research, which suggests that having fun while learning contributes to the learning outcomes [30, 31, 40, 52, 55]. Although some dimensions of fun could be predicted from the physio-affective states of the child, we found that the physio-affective states do not predict fun comprehensibly. This aligns with previous works in physiological response measures that indicate challenges with achieving perfect one-to-one relationship between physiological response measures and psychological constructs [9].

Concerning our findings about the dimensions of FunQ, we conclude that as just mentioned, not all its dimensions could be predicted from physiological response data. Accordingly, neither the Autonomy nor the Loss of Social Barriers dimensions could be modelled by the affective states. We believe that these results rather reflect the characteristics of the activity rather than general tendencies. Namely, despite children were provided with some freedom and attributes that are atypical in a formal learning environment (e.g., they could decide whether they wanted to follow the activity, they were allowed to move around freely and ask the instructors whenever they wanted), the workshop was still scripted as it followed a fixed sequence of tasks. Hence, children might have not felt a sufficient level of autonomy (or not frequently enough) to be able to relate it to physiological response data. Regarding the Loss of Social Barrier dimension, on top of the aforementioned possible explanations, the scripted structure of the workshop might not have provided enough space for social interactions that would have led to an increase in social connectedness. Hence, it appeared not to be possible to link this dimension to physiological response data. Regarding both dimensions, further research is required to establish general tendencies as our findings might be a consequence of the activity design and be activity specific.

The Challenge dimension of FunQ could be predicted from the CVT affective states happiness, anger and arousal positively, and the transitions between happiness and sadness. These transitions' positive contribution to FunQ challenge can perhaps be because children were in a constant loop of succeeding and failing, as not everything worked at their first attempt. On the other hand, sadness, surprise, and transition between sadness and surprise, contribute negatively to FunQ Challenge. Connected to the previous finding, when children were failing the task, it could have been a sign that the task was (momentarily) too difficult for them and it can explain of why the transition between sadness and surprise contributed negatively to FunQ Challenge. During a coding activity, the children need to deal with different aspects of the tasks, like debugging, problem solving and reflecting iteratively on the needed actions, and this process can be difficult and challenging [33].

As for the Delight dimension, happiness and surprise appeared to be a positive contributor, whereas stress, and the transition between happiness and anger contributed negatively. We propose that Delight can be seen as an emotion related to solving or understanding a problem or even having a desired outcome in a given possible task [14] and this can be triggered from positive emotions or an unexpected outcome. Regarding Immersion, happiness and arousal are found to contribute positively to it, while transition between sadness and anger turned out to be a negative contributor to Immersion.

Concerning the Stress dimension, which is a contra-indicative dimension of FunQ with reversed items, we found that the physio-affective states sadness, anger, stress, and the transition between sadness and anger contributed positively, while the transition between happiness and surprise contributed negatively. In other words, the physio-affective states sadness, anger and stress are inducing stress, while changing from happy to surprised, and vice-versa is a contra-indicative signal, indicating low levels of stress.

For all the constructs that we have used in this contribution (i.e., RLG and the FunQ dimensions), stress and/or arousal have been a significant predictor. We found that arousal is positively associated with RLG and Challenge, Delight, and Immersion dimensions of FunQ; stress is negatively associated with RLG and positively associated with the Stress dimension of FunQ. The positive association of arousal and the negative association of stress with the RLG (or in other words learning or cognitive performance) is consistent with various other studies. For example, in game-based learning settings with children Lee-Cultura et al [27] and Sharma et al [42] found physiological stress to be negatively associated with learning performance and experiences. Similarly, Joëls et al [23] showed that the memory-based learning performances decrease under stress. These studies are also in line with the finding that higher levels of stress are negatively associated with the RLG extends the consensus from these studies. Furthermore, the physiological response measurement of stress being positively associated with the self-reported stress is indicative of the measurements' validity in the context of children coding.

On the other hand, physiological arousal provides us with a reliable proxy of engaged behavior [7, 26, 28]. The high levels of engagement have been shown to be positively associated with learning [8, 10, 20]. In our case, higher levels of physiological arousal indicate high levels of engagement which in turn increases the probability of children with high physiological arousal also having a high RLG. Moreover, with high levels of engagement, children might also feel immersed and challenged at appropriate levels, which in turn might increase their ratings for the delight dimension of FunQ.

5.1 Implications

Our findings support endeavors of educators, designers, and researchers to make learning to code a fun experience, as we found a positive relationship between those. Further research studies could aim to improve the applicability of physiological measure devices (e.g. wristbands) for children. Beyond research purposes, such improvements in the physiological measure devices could pave the way for everyday (classroom) use. Our research also opens ways for at-the-moment measurement of fun that will allow us a precise insight into the activity, in contrast with the post-hoc tests and get a more holistic understanding. This way, micro-level investigations and interventions are enabled for supporting fun, leading to increased learning outcomes – a finding introduced by this study and supported by previous research indicating a clear relationship between children's perceived fun while learning and their learning outcomes [52]. Additionally, it can be particularly relevant for the development of different systems for educational purposes, to use multimodal data to support both teachers and students in their everyday learning activities. This can happen for example by providing systems with affordances for reflective purposes, indicating students' disengagement or other features to support better classroom management. This can be helpful from the students' perspective because they will be able to signal when they are in need for more support from the teacher/instructor. Future systems with different functionalities can also exploit multimodal data, to, for example, automatically adjust the difficulty level of a learning task, providing personalized learning to students on a given task. Personalization should then also take into account the learning setup, whether the task requires individual work, or collaboration is possible. In the latter case adjusting the timing of the scaffolding would be required in order to support social learning. Knowing the affective state of the students, can be powerful information helping them to overcome affective states that may hinder their learning or fun during coding.

5.2 Limitations and Future Work

Besides the applicable findings and the new approach introduced, the limitations of this study should be mentioned. First, we highlight the practical difficulties involved in collecting of physiological response data from children as these technologies are designed for adults. One example is the difficulty we faced in attaching the wristbands to some of the children's wrists. This led to some uncontrolled data loss. To resolve this issue, further research could assist the design of the wristbands to be more suitable for child users. In the same lines, the use of sensing devices increases children's curiosity, therefore researchers need to spend time to explain in simple words how each device works, what data we collect and why, letting the children interact with them. One example from our study is that we observed that some children wanted to see on the mobile device connected with the wristband how their heart rate is shown or how it changes. If this cannot be controlled, it can lead to some data removal.

Second, the coding activity was designed as a non-curricular activity, but in a classroom setting, aiming to provide participating children with autonomy over their participation and the activity itself. Since we did not find physiological response correlates for the Autonomy and Loss of Social Barriers dimension of FunQ, we speculate that given the activity was scripted (i.e. three tasks were given to be followed), children might not have felt the desired level of autonomy, and in relation to this, they also might not felt enough freedom to connect to each other more than usual. Future studies, hence, should examine the physiological response correlates of the FunQ dimensions in relation to a broad range of learning activities, including possibly informal learning setups as well.

Another limitation of this study is the use of only quantitative data, future research should also include qualitative data such as interviews or observations to do a triangulation of the findings and get deeper into children's behavior. Although facial expressions have been used in many studies to extract emotions, this method comes with some limitations [4, 21]. Lastly, more studies are needed to better understand the cognitive and affective states of children during coding and to monitor how they may shift naturally or not with the ultimate goal to offer more effective and efficient learning experiences.

While this study is the first to connect FunQ and sensor data focusing on a quantitative exploration, we call on further research, including both qualitative, quantitative, and especially mixed-method approaches to provide more insights into this relationship, and to triangulate the results of the herein introduced study. Hence, in future mixed-method studies participants could be closely monitored on the individual level in order to pinpoint moments that are detrimental or helpful for learning and/or fun, and those moments could be studied in depth to extend our understanding on the topic.

6 CONCLUSION

We contribute to the literature regarding the role of fun in how children learn to code in a number of ways. First, we investigated fun, a construct, which is frequently in the focus of evaluation in design and educational research, however, our knowledge is still limited about its nature. By using multimodal data, we went a step further than earlier research as it either pertained to surveys or to physiological response data only. Using the combination of the two allowed us a deeper understanding on how fun occurs during learning to program, and which physio-affective states can be used as a predictor of fun. Being able to predict fun from physiological signals can help assess different learning activities, but potentially can be developed further to support timely interventions to get disengaged children on track again, ultimately leading to better learning outcomes. In contrast to surveying children about their level of fun, using physiological response data by its unobstructive nature can provide us with immediate feedback, without disrupting the learning experience (surveying several times during an activity) and without inducing recency bias (surveying once at the end). Developing new tools or further improving existing ones that address the potential of unobstructive physiological response data could support both teachers and students in their everyday life, by providing systems with affordances for reflective purposes, indicating students' disengagement or other features to support better classroom management.

7 SELECTION AND PARTICIPATION OF CHILDREN

Before the start of the study, both children and their parents were informed about the study, and informed consent was obtained accordingly. The consent form contained information on the purpose of the study, the explanation of the procedures, the potential risks and benefits, the data handling and confidentiality, and the withdrawing of participation. At the beginning of each workshop, 11 children were randomly selected for participation in the study, i.e. for collect physiological response data and facial videos.

ACKNOWLEDGMENTS

We wish to express our gratitude for the participating children and their class teachers for contributing to our research. We are also grateful for Jeroen Storm for his collaboration with the videos.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement NO 787476. This paper reflects only the authors' views. The Research Executive Agency (REA) and the European Commission are not responsible for any use that may be made of the information it contains.

REFERENCES

[1] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. 2016. OpenFace: A general-purpose face recognition library with mobile applications. Technical Report. CMU-CS-16-118, CMU School of Computer Science.
[2] Alejandro Andrade, Joshua A. Danish, and Adam V. Maltese. 2017. A Measurement Model of Gestures in an Embodied Learning Environment: Accounting for Temporal Dependencies. *Journal of Learning Analytics* 4, 3. https://doi.org/10.18608/jla.2017.43.3
[3] Alejandro Andrade, Ginette Delandshere, and Joshua A. Danish. 2016. Using Multimodal Learning Analytics to Model Student Behavior: A Systematic Analysis

of Epistemological Framing. *Journal of Learning Analytics* 3, 2: 282–306. https://doi.org/10.18608/jla.2016.32.14
[4] L.F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak. 2019. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest* 20, 3: 165–166. https://doi.org/10.1177/1529100619889954
[5] Paulo Blikstein and Marcelo Worsley. 2016. Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics* 3, 2: 220–238. https://doi.org/10.18608/jla.2016.32.11
[6] Benjamin Bloom. 1956. *Taxonomy of educational objectives: The classification of educational goals*. David McKay Co Inc, New York, NY, US.
[7] Boyan Bontchev and Dessislava Vassileva. 2016. Assessing engagement in an emotionally-adaptive applied game. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 747–754. https://doi.org/10.1145/3012430.3012602
[8] Chris A. Boulton, Carmel Kent, and Hywel T.P. Williams. 2018. Virtual learning environment engagement and learning outcomes at a 'bricks-and-mortar' university. *Computers & Education* 126: 129–142. https://doi.org/10.1016/j.compedu.2018.06.031
[9] John T. Cacioppo and Louis G. Tassinary. 1990. Inferring psychological significance from physiological signals. *American Psychologist* 45, 1: 16–28. https://doi.org/10.1037/0003-066X.45.1.16
[10] Michelene T. H. Chi and Ruth Wylie. 2014. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist* 49, 4: 219–243. https://doi.org/10.1080/00461520.2014.965823
[11] J. F. Cohn, Z. Ambadar, and P. Ekman. 2007. Observer-based measurement of facial expression with the Facial Action Coding System. e. In *The handbook of emotion elicitation and assessment,*. 203–221.
[12] Mutlu Cukurova, Michail Giannakos, and Roberto Martinez-Maldonado. 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
[13] Mutlu Cukurova, Carmel Kent, and Rosemary Luckin. 2019. Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring. *British Journal of Educational Technology* 50, 6: 3032–3046. https://doi.org/10.1111/bjet.12829
[14] Sidney D'Mello and Art Graesser. 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
[15] P. Ekman, W. V. Friesen, and J. C. Hager. 2002. *Facial Action Coding System. The Manual On CD ROM.*
[16] Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion* 6, 3–4: 169–200. https://doi.org/10.1080/02699939208411068
[17] Paul Ekman, Robert W Levenson, and Wallace V. Friesen. 1983. Autonomic Nervous System Activity Distinguishes Among Emotions. *Science* 221, 4616: 1208–1210. https://doi.org/10.1126/science.6612338
[18] Arthur C. Graesser. 2019. Emotions are the experiential glue of learning environments in the 21st century. *Learning and Instruction*, May: 101212. https://doi.org/10.1016/j.learninstruc.2019.05.009
[19] Arthur C. Graesser, Shulan Lu, Brent A. Olde, Elisa Cooper-Pye, and Shannon Whitten. 2005. Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory & Cognition* 33, 7: 1235–1247. https://doi.org/10.3758/BF03193225
[20] Jianpeng Guo. 2018. Building bridges to student learning: Perceptions of the learning environment, engagement, and learning outcomes among Chinese undergraduates. *Studies in Educational Evaluation* 59: 195–208. https://doi.org/10.1016/j.stueduc.2018.08.002
[21] Douglas Heaven. 2020. Why faces don't always tell the truth about feelings. *Nature* 578, 7796: 502–504. https://doi.org/10.1038/d41586-020-00507-5
[22] Nina Iten and Dominik Petko. 2016. Learning with serious games: Is fun playing the game a predictor of learning success? *British Journal of Educational Technology* 47, 1: 151–163. https://doi.org/10.1111/bjet.12226
[23] Marian Joëls, Zhenwei Pu, Olof Wiegert, Melly S. Oitzl, and Harm J. Krugers. 2006. Learning under stress: how does it work? *Trends in Cognitive Sciences* 10, 4: 152–158. https://doi.org/10.1016/j.tics.2006.02.002
[24] Heather Killen, David Weintrop, and Megean Garvin. 2019. AP Computer Science Principles' Impact on the Landscape of High School Computer Science using Maryland as a Model. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 1060–1066. https://doi.org/10.1145/3287324.3287356
[25] Panagiotis Kosmas, Andri Ioannou, and Symeon Retalis. 2018. Moving Bodies to Moving Minds: A Study of the Use of Motion-Based Games in Special Education. *TechTrends* 62, 6: 594–601. https://doi.org/10.1007/s11528-018-0294-5
[26] Elena Di Lascio, Shkurta Gashi, and Silvia Santini. 2018. Unobtrusive Assessment of Students' Emotional Engagement during Lectures Using Electrodermal Activity Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3: 1–21. https://doi.org/10.1145/3264913
[27] Serena Lee-Cultura, Kshitij Sharma, and Michail Giannakos. 2021. Children's play and problem-solving in motion-based learning technologies using a multi-modal mixed methods approach. *International Journal of Child-Computer Interaction*: 100355. https://doi.org/10.1016/j.ijcci.2021.100355
[28] Serena Lee-Cultura, Kshitij Sharma, Sofia Papavlasopoulou, Symeon Retalis, and Michail Giannakos. 2020. Using sensing technologies to explain children's self-representation in motion-based educational games. *Proceedings of the Interaction Design and Children Conference, IDC 2020*, June: 541–555. https://doi.org/10.1145/3392063.3394419
[29] Kristina Loderer, Reinhard Pekrun, and James C. Lester. 2020. Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction* 70: 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002
[30] Ju Long. 2007. Just For Fun: Using Programming Games in Software Programming Training and Education - A Field Study of IBM Robocode Community. *Journal of Information Technology Education* 6: 279–290. https://doi.org/10.28945/216
[31] Dorothy Lucardie. 2014. The Impact of Fun and Enjoyment on Adult's Learning. *Procedia - Social and Behavioral Sciences* 142: 439–446. https://doi.org/10.1016/j.sbspro.2014.07.696
[32] Shayan Mirjafari, Kizito Masaba, Ted Grover, Weichen Wang, Pino Audia, Andrew T. Campbell, Nitesh V. Chawla, Vedant Das Swain, Munmun De Choudhury, Anind K. Dey, Sidney K. D'Mello, Ge Gao, Julie M. Gregg, Krithika Jagannath, Kaifeng Jiang, Suwen Lin, Qiang Liu, Gloria Mark, Gonzalo J. Martinez, Stephen M. Mattingly, Edward Moskal, Raghu Mulukutla, Subigya Nepal, Kari Nies, Manikanta D. Reddy, Pablo Robles-Granda, Koustuv Saha, Anusha Sirigiri, and Aaron Striegel. 2019. Differentiating Higher and Lower Job Performers in the Workplace Using Mobile Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2: 1–24. https://doi.org/10.1145/3328908
[33] Sofia Papavlasopoulou, Michail N. Giannakos, and Letizia Jaccheri. 2019. Exploring children's learning experience in constructionism-based coding activities through design-based research. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2019.01.008
[34] Sofia Papavlasopoulou, Kshitij Sharma, Michail Giannakos, and Letizia Jaccheri. 2017. Using Eye-Tracking to Unveil Differences Between Kids and Teens in Coding Activities. In *Proceedings of the 2017 Conference on Interaction Design and Children*, 171–181. https://doi.org/10.1145/3078072.3079740
[35] Sofia Papavlasopoulou, Kshitij Sharma, and Michail N. Giannakos. 2018. How do you feel about learning to code? Investigating the effect of children's attitudes towards coding using eye-tracking. *International Journal of Child-Computer Interaction* 17: 50–60. https://doi.org/10.1016/j.ijcci.2018.01.004
[36] Sofia Papavlasopoulou, Kshitij Sharma, and Michail N. Giannakos. 2019. Coding activities for children: Coupling eye-tracking with qualitative data to investigate gender differences. *Computers in Human Behavior*, 7491: 1–11. https://doi.org/10.1016/j.chb.2019.03.003
[37] Reinhard Pekrun. 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
[38] Janet C. Read. 2008. Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. *Cognition, Technology & Work* 10, 2: 119–128. https://doi.org/10.1007/s10111-007-0069-9
[39] Mitchel Resnick. 1998. Technologies for lifelong kindergarten. *Educational Technology Research and Development* 46, 4: 43–55. https://doi.org/10.1007/BF02299672
[40] José Manuel Sáez-López, Marcos Román-González, and Esteban Vázquez-Cano. 2016. Visual programming languages integrated across the curriculum in elementary school: A two year case study using "scratch" in five schools. *Computers and Education* 97: 129–141. https://doi.org/10.1016/j.compedu.2016.03.003
[41] M. A. Sayette, J. F. Cohn, J. M. Wertz, M. A. Perrott, and D. J. Parrott. 2001. A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *Journal of Nonverbal Behavior* 25, 3: 167–185.
[42] K. Sharma, S. Lee-Cultura, and M. Giannakos. 2022. Keep Calm and Don't Carry-Forward: Towards sensor-data driven AI agent to enhance human learning. *Frontiers in Artificial Intelligence* 198.
[43] Kshitij Sharma and Michail Giannakos. 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
[44] Kshitij Sharma, Evangelos Niforatos, Michail Giannakos, and Vassilis Kostakos. 2020. Assessing Cognitive Performance Using Physiological and Facial Features. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3. https://doi.org/10.1145/3411811
[45] Kshitij Sharma, Zacharoula Papamitsiou, and Michail Giannakos. 2019. Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach. *British Journal of Educational Technology*, June: bjet.12854. https://doi.org/10.1111/bjet.12854
[46] G. Sim, S. MacFarlane, and J. C. Read. 2006. All work and no play: Measuring fun, usability, and learning in software for children. *Computers & Education* 46, 3: 235–248. https://doi.org/10.1016/j.compedu.2005.11.021
[47] R.E. Snow, L. Corno, and D.N. Jackson. 1996. Individual differences in affective and conative functions. In *Handbook of Educational Psychology.*

[48] Daniel Spikol, Katerina Avramides, and Mutlu Cukurova. 2016. Exploring the interplay between human and machine annotated multimodal analytics in hands-on STEM activities. In *Proceedings of the Sixth International Learning Analytics and Knowledge Conference*, 522–523. https://doi.org/10.1145/1235

[49] Daniel Spikol, Emanuele Ruffaldi, Giacomo Dabisias, and Mutlu Cukurova. 2018. Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning* 34, 4: 366–377. https://doi.org/10.1111/jcal.12263

[50] Christian Szegedy, Sergey Ioffe, and Vincent Vanhoucke. 2016. *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv: 1602.07261*. Retrieved from http://arxiv.org/abs/1602.07261

[51] Joachim Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel. 2009. Influence of Mental Stress on Heart Rate and Heart Rate Variability. . 1366–1369. https://doi.org/10.1007/978-3-540-89208-3_324

[52] Gabriella Tisza and Panos Markopoulos. 2021. Understanding the role of fun in learning to code. *International Journal of Child-Computer Interaction* 28: 100270. https://doi.org/10.1016/j.ijcci.2021.100270

[53] Gabriella Tisza and Panos Markopoulos. 2021. FunQ: Measuring the fun experience of a learning activity with adolescents. *Current Psychology*. https://doi.org/10.1007/s12144-021-01484-2

[54] Gabriella Tisza, Panos Markopoulos, and Tilde M. Bekker. 2020. Learning to code: Interplay of attitude, emotions and fun. *Manuscript submitted for publication.*

[55] Gabriella Tisza, Sijie Zhu, and Panos Markopoulos. 2021. Fun to Enhance Learning, Motivation, Self-efficacy, and Intention to Play in DGBL. In *IFIP ICEC 2021*. 28–45. https://doi.org/10.1007/978-3-030-89394-1_3

[56] Tzu-Wei Tsai, Hsiao Yu Lo, and Kai-Shao Chen. 2012. An affective computing approach to develop the game-based adaptive learning material for the elementary students. In *Proceedings of the 2012 Joint International Conference on Human-Centered Computer Environments - HCCE '12*, 8. https://doi.org/10.1145/2160749.2160752

[57] Marcelo Worsley and Paulo Blikstein. 2018. A Multimodal Analysis of Making. *International Journal of Artificial Intelligence in Education* 28, 3: 385–419. https://doi.org/10.1007/s40593-017-0160-1

[58] Feiran Zhang, Panos Markopoulos, and Tilde Bekker. 2020. Children's Emotions in Design-Based Learning: a Systematic Review. *Journal of Science Education and Technology* 29, 4: 459–481. https://doi.org/10.1007/s10956-020-09830-y

A APPENDDIX 1

The full model details are provided in the following Appendix.

Table 6: Detailed results for Model predicting the RLG using affective states, the transitions among them and the physiological measurements. The boldface coefficients are significant. The predictor model accounts for 72% variance in children's RLG.

Variable	β	Error	T-value	P-value
Intercept	0.27	0.69	0.52	> 0.05
Happiness (hap)	0.59	0.88	0.49	> 0.05
Sadness (sad)	**-1.37**	**0.009**	**-2.38**	**0.01**
Anger (ang)	**-1.56**	**0.003**	**-3.12**	**0.001**
Surprise (sup)	0.42	0.94	0.36	> 0.05
Trans:hap <-> sad	0.64	0.74	0.23	> 0.05
Trans:hap <-> ang	-0.48	0.84	-0.52	> 0.05
Trans:hap <-> sup	**1.13**	**0.04**	**2.03**	**0.02**
Trans:sad <-> ang	**-1.88**	**0.004**	**-3.42**	**0.0006**
Trans:sad <-> sup	0.11	0.88	0.16	> 0.05
Trans:ang <-> sup	0.08	0.63	0.12	> 0.05
Stress	**-1.03**	**0.01**	**-1.78**	**0.04**
Arousal	**1.19**	**0.02**	**1.71**	**0.04**

Table 7: Model for the FunQ Total Score using the affective states, the transitions among them and the physiological measurements. The boldface coefficients are significant. The predictor model accounts for 26% variance in children's FunQ Total score.

	β	Error	T-value	P-value
Intercept	0.57	0.56	0.75	> 0.05
Happiness (hap)	0.58	0.90	0.12	> 0.05
Sadness (sad)	-0.79	0.57	-0.61	> 0.05
Anger (ang)	-0.67	0.49	-0.46	> 0.05
Surprise (sup)	0.77	0.43	0.75	> 0.05
Trans:hap <-> sad	0.65	0.70	0.69	> 0.05
Trans:hap <-> ang	0.81	1.01	0.48	> 0.05
Trans:hap <-> sup	0.72	0.92	0.33	> 0.05
Trans:sad <-> ang	-0.71	0.46	-0.32	> 0.05
Trans:sad <-> sup	-0.58	0.68	-0.39	> 0.05
Trans:ang <-> sup	0.63	0.46	0.47	> 0.05
Stress	-0.51	0.78	-0.40	> 0.05
Arousal	0.70	0.84	0.37	> 0.05

Table 8: Model for the FunQ Autonomy using the affective states, the transitions among them and the physiological measurements. The boldface coefficients are significant. The predictor model accounts for 32% variance in children's FunQ Autonomy.

	β	Error	T-value	P-value
Intercept	0.04	0.02	0.34	> 0.05
Happiness (hap)	-0.24	0.17	-0.82	> 0.05
Sadness (sad)	-0.72	0.58	-0.74	> 0.05
Anger (ang)	0.88	0.75	0.35	> 0.05
Surprise (sup)	0.96	0.45	0.34	> 0.05
Trans:hap <-> sad	0.44	0.35	0.06	> 0.05
Trans:hap <-> ang	0.89	0.80	0.06	> 0.05
Trans:hap <-> sup	0.29	0.18	0.61	> 0.05
Trans:sad <-> ang	-0.35	0.34	-0.30	> 0.05
Trans:sad <-> sup	-0.79	0.64	-0.37	> 0.05
Trans:ang <-> sup	0.95	0.87	0.99	> 0.05
Stress	0.28	0.32	0.73	> 0.05
Arousal	0.10	0.21	0.17	> 0.05

Table 9: Model for the FunQ Challenge using the affective states, the transitions among them and the physiological measurements. The boldface coefficients are significant. The predictor model accounts for 71% variance in children's FunQ Challenge.

	β	Error	T-value	P-value
Intercept	0.31	0.12	0.43	> 0.05
Happiness (hap)	**1.41**	**0.005**	**2.92**	**0.002**
Sadness (sad)	**-1.32**	**0.012**	**-2.10**	**0.02**
Anger (ang)	**1.02**	**0.004**	**3.13**	**0.001**
Surprise (sup)	**-1.35**	**0.003**	**-2.44**	**0.009**
Trans:hap <-> sad	**0.99**	**0.001**	**3.38**	**0.0007**
Trans:hap <-> ang	0.31	0.30	0.35	> 0.05
Trans:hap <-> sup	0.42	0.33	0.33	> 0.05
Trans:sad <-> ang	0.47	0.43	0.48	> 0.05
Trans:sad <-> sup	**-1.45**	**0.001**	**-3.04**	**0.001**
Trans:ang <-> sup	0.33	0.46	0.39	> 0.05
Stress	0.32	0.33	0.34	> 0.05
Arousal	**1.23**	**0.014**	**2.23**	**0.01**

Table 10: Model for the FunQ Delight using the affective states, the transitions among them and the physiological measurements. The boldface coefficients are significant. The predictor model accounts for 65% variance in children's FunQ Delight.

	β	Error	T-value	P-value
Intercept	0.44	0.24	0.54	> 0.05
Happiness (hap)	**1.34**	**0.051**	**2.13**	**0.02**
Sadness (sad)	-0.42	0.11	-0.97	> 0.05
Anger (ang)	-0.33	0.09	-1.44	> 0.05
Surprise (sup)	**1.93**	**0.009**	**3.19**	**0.001**
Trans:hap <-> sad	-0.39	0.42	-0.66	> 0.05
Trans:hap <-> ang	**-0.93**	**0.001**	**-4.18**	**0.00006**
Trans:hap <-> sup	0.53	1.12	0.20	> 0.05
Trans:sad <-> ang	-0.99	0.12	-1.35	> 0.05
Trans:sad <-> sup	0.33	0.23	0.44	> 0.05
Trans:ang <-> sup	-0.59	0.62	-0.43	> 0.05
Stress	0.43	0.91	0.12	> 0.05
Arousal	**-1.34**	**0.003**	**-3.53**	**0.0004**

Table 11: Model for the FunQ Immersion using the affective states, the transitions among them and the physiological measurements. The boldface coefficients are significant. The predictor model accounts for 63% variance in children's FunQ Immersion.

	β	Error	T-value	P-value
Intercept	0.19	0.89	0.52	> 0.05
Happiness (hap)	**1.46**	**0.001**	**4.34**	**0.00003**
Sadness (sad)	0.22	0.45	0.38	> 0.05
Anger (ang)	-0.62	0.26	-0.34	> 0.05
Surprise (sup)	0.44	0.61	0.22	> 0.05
Trans:hap <-> sad	0.27	0.77	0.40	> 0.05
Trans:hap <-> ang	-0.83	0.49	-0.36	> 0.05
Trans:hap <-> sup	0.32	0.59	0.22	> 0.05
Trans:sad <-> ang	**-1.73**	**0.003**	**-3.28**	**0.009**
Trans:sad <-> sup	0.31	0.82	0.37	> 0.05
Trans:ang <-> sup	0.62	0.77	0.20	> 0.05
Stress	0.83	0.65	0.20	> 0.05
Arousal	**2.01**	**0.001**	**4.26**	**0.00005**

Table 12: Model for the FunQ Social Barrier using the affective states, the transitions among them and the physiological measurements. The boldface coefficients are significant. The predictor model accounts for 17% variance in children's FunQ Social barrier.

	β	Error	T-value	P-value
Intercept	0.47	0.66	0.40	> 0.05
Happiness (hap)	0.57	0.48	0.45	> 0.05
Sadness (sad)	-0.77	0.67	-0.33	> 0.05
Anger (ang)	-0.80	0.49	-0.51	> 0.05
Surprise (sup)	0.50	0.92	0.30	> 0.05
Trans:hap <-> sad	-0.55	0.90	-0.56	> 0.05
Trans:hap <-> ang	-0.47	0.73	-0.57	> 0.05
Trans:hap <-> sup	0.44	0.54	0.48	> 0.05
Trans:sad <-> ang	-0.77	0.91	-0.44	> 0.05
Trans:sad <-> sup	-0.39	0.69	-0.56	> 0.05
Trans:ang <-> sup	-0.81	0.69	-0.55	> 0.05
Stress	0.62	0.90	0.47	> 0.05
Arousal	0.59	0.91	0.31	> 0.05

Table 13: Model for the FunQ Stress using the affective states, the transitions among them and the physiological measurements. The boldface coefficients are significant. The predictor model accounts for 70% variance in children's FunQ Stress.

	β	Error	T-value	P-value
Intercept	0.16	0.21	0.89	> 0.05
Happiness (hap)	-0.70	0.78	-0.25	> 0.05
Sadness (sad)	**1.28**	**0.005**	**2.48**	**0.008**
Anger (ang)	**0.94**	**0.001**	**3.32**	**0.0009**
Surprise (sup)	-0.87	0.86	-0.21	> 0.05
Trans:hap <-> sad	0.58	0.72	0.21	> 0.05
Trans:hap <-> ang	0.63	0.83	0.34	> 0.05
Trans:hap <-> sup	**-0.89**	**0.006**	**3.29**	**0.0009**
Trans:sad <-> ang	**1.27**	**0.017**	**2.70**	**0.005**
Trans:sad <-> sup	-0.88	0.74	-0.26	> 0.05
Trans:ang <-> sup	0.59	0.82	0.34	> 0.05
Stress	**1.39**	**0.001**	**3.88**	**0.0001**
Arousal	0.70	0.62	0.32	> 0.05