"That Student Should be a Lion Tamer!" StressViz: Designing a Stress Analytics Dashboard for Teachers

RIORDAN DERVIN ALFREDO*, Monash University, Australia
LANBING NIE*, Monash University, Australia
PAUL KENNEDY, University of Technology Sydney, Australia
TAMARA POWER, University of Sydney, Australia
CAROLYN HAYES, University of Sydney, Australia
HUI CHEN, University of Technology Sydney, Australia
CAROLYN MCGREGOR, Ontario Tech University, Canada
ZACHARI SWIECKI, Monash University, Australia
DRAGAN GAŠEVIĆ, Monash University, Australia
ROBERTO MARTINEZ-MALDONADO, Monash University, Australia

Fig. 1. A stress-inducing de-escalation simulation scenario. A team of undergraduate students, enacting the roles of registered/student nurses, patient's family members and observers (not shown in the figure) all wearing Empatica E4 physiological wristbands.

In recent years, there has been a growing interest in creating multimodal learning analytics (LA) systems that automatically analyse students' states that are hard to see with the "naked eye", such as cognitive load and stress levels, but that can considerably shape their learning experience. A rich body of research has focused on detecting such aspects by capturing bodily signals from students using wearables and computer vision. Yet, little work has aimed at designing end-user interfaces that visualise physiological data to support tasks deliberately designed for students to learn from stressful situations. This paper addresses this gap by designing a stress analytics dashboard that encodes students' physiological data into stress levels during different phases of an authentic team simulation in the context of nursing education. We conducted a qualitative study with teachers to understand (i) how they made sense of the stress analytics dashboard; (ii) the extent to which they trusted the dashboard in relation to students' cortisol data; and (iii) the potential adoption of this tool to communicate insights and aid teaching practices.

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2023 Association for Computing Machinery.
Manuscript submitted to ACM

CCS Concepts: • **Applied computing** → *Computer-assisted instruction*; • **Human-centered computing** → **Information visualization**.

Additional Key Words and Phrases: Affective computing, Stress detection, Visualisation, Healthcare education, Multimodal dataset, LA dashboard

ACM Reference Format:
Riordan Dervin Alfredo, Lanbing Nie, Paul Kennedy, Tamara Power, Carolyn Hayes, Hui Chen, Carolyn McGregor, Zachari Swiecki, Dragan Gašević, and Roberto Martinez-Maldonado. 2023. "That Student Should be a Lion Tamer!" StressViz: Designing a Stress Analytics Dashboard for Teachers. In *LAK23: 13th International Learning Analytics and Knowledge Conference (LAK 2023), March 13–17, 2023, Arlington, TX, USA*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3576050.3576058

1 INTRODUCTION

While most of the current student-focused learning analytics (LA) innovations rely on the activity logs of digital learning tools [28], there is a growing interest in creating systems that automatically analyse physiological data to model students' affective and cognitive states that can considerably shape their learning experience [1, 40, 55]. Examples of these include students' emotional responses [12, 25], cognitive load [31]), arousal [15]), and stress [30] while performing learning tasks. Such states are often hard to see with the "naked eye" [23] and cannot be easily inferred from the analysis of system logs [9, 47] and conventional data sources (e.g., interviews, surveys, and observations) alone. As a result, there is an increased interest in capturing bodily signals using multimodal wearable sensors (e.g., [7]) and computer vision algorithms (e.g., [49]) to detect such states. Yet, recent reviews [22, 40] suggest that the actual use of physiological data modalities in education is, in general, still in its infancy.

A key construct that has yet to be fully explored within LA is that of *stress*. While there is debate as to whether stress negatively or positively affects learning (outcomes often termed as distress and eustress, respectively) [43], in some educational situations, educators may design learning tasks that deliberately provoke stress so students can learn from it. This approach is common in the pedagogy of high-stress domains such as healthcare [38], law enforcement, firefighting, and emergency response training [36]. The purpose of inducing stress is to provide safe and controlled learning environments where students can develop effective stress management and de-escalation skills. An approach that educators commonly use to help students develop these skills is to conduct post-hoc, reflective debriefs after challenging learning tasks [14]. Yet, the effectiveness of the debrief often depends on the educator's ephemeral observations and the extent to which students can articulate their personal experiences [2]. In these situations it may be helpful to make physiological data available to educational stakeholders for the purpose of supporting reflective practices [11]

There is existing work within LA and related research communities that has indirectly explored modelling stress from physiological data via proxy constructs such as cognitive load [31] and arousal [33]. While some LA studies have proposed ways to visualise arousal (e.g., [13, 39]), most of the work in actual stress data visualisation has been aimed at supporting people at their workplace or for general well-being (see examples in [46]) rather than supporting learners, which is a key goal in closing the LA loop. To the best of our knowledge, no previous work in LA has directly focused on providing critical insights to support reflection on students' stress levels for tasks purposely designed to provoke stress. This may partly be explained by the fact that stress detection has yet to become very accurate in the context of LA [25]. However, work beyond the LA field is increasingly showing promise in the detection of stressful states [45, 48]. To address this gap, we report on the design of *StressViz*, a dashboard that encodes students' physiological data into stress levels.

We conducted a pilot study in an authentic course in which teachers regularly deploy de-escalation nursing simulations carefully designed to deliberately increase students' stress. The study focused on simulations where teams of students visit the home of an aggressive patient, enacted by a human actor, to conduct a medical procedure (see Figure 1). Physiological and cortisol data from 20 nursing students, grouped in teams of five, were collected during the simulations. Stress levels were modelled based on a publicly available stress dataset [45]. A further qualitative study was carried out with six nursing teachers to understand (i) their sensemaking processes of the stress analytics dashboard; (ii) the extent to which they can trust the stress results with learners' cortisol data as an alternative biomarker to measure stress; and (iii) the potential usage of this tool to communicate insights and aid teaching practices.

2 BACKGROUND

2.1 Advancements in stress detection

When people face a stressful situation, the sympathetic nervous system gets activated which consequently triggers the release of the hormone cortisol by the adrenal glands [5]. As a result, physiological and physical responses often occur, such as increases in heart rate, breathing rate, muscle tightness, body temperature and blood pressure [10]. Some of these responses can be measured by several biomarkers, including heart rate variability (HRV) through electrocardiogram (ECG) or blood volume pulse (BVP), electrodermal activity (EDA), galvanic skin response (GSR), skin temperature (TEMP) [18], and salivary cortisol levels [21]. These bodily responses have been automatically detected and measured by researchers using a wide range of devices. For example, several studies have adopted medical-grade wearables [13], consumer-grade wearables [8], computer vision [49], thermal infrared [19], and multimodal sensors [32].

While different approaches to estimating stress have been explored, one practical way to classify such approaches is to divide them into personalised and generalised models. Personalised stress detection models adapt their predictions to specific individuals using subjective ratings from different affective stimuli and a pre-recorded baseline [10, 52]. These ratings are commonly collected through self-reports or questionnaires from participants [45]. In contrast, generalised stress detection models aim to predict anyone's stress level by using derived physiological data and its statistical measurements irrespective of subjects' self-reporting information and without necessarily requiring collecting an individual baseline by training models on labelled stress-related dataset [17]. In educational contexts, personalised stress detection can be impractical because acquiring each person's physiological reference data in realistic conditions would affect the scheduling and logistics of pedagogical activities.

Research beyond the LA field is increasingly showing promise in the estimation of stressful states with generalised models [4]. Yet, little LA work has aimed at automatically estimating stress to be used to support learning. The only exceptions are studies which have *indirectly* explored the potential of generalised approaches to model constructs that can be associated to stress, namely arousal [33], cognitive load [31], and eye fatigue [26]. Yet, these constructs cannot necessarily be considered as proxies of stress as they can also be associated with other factors such as increased attention [3], high task complexity [29], and over-use of screens [24], respectively. In short, although little progress has been made in creating LA innovations that automatically detect measures related to stress, for those learning tasks that deliberately provoke stress in students it may be productive to expose educators and students to physiological measures of stress. Such measures may help to facilitate reflection by students and teachers to improve learning [36, 38].

2.2 Stress data visualisation

There is an emerging proliferation of approaches to visualise people's own stress information outside educational contexts. For example, Kocielnik et al. [30] suggested the generation of stress visualisations from physiological data to help office workers reflect on their stress levels throughout the day. Semikina [46] proposed various visualisation techniques for stress data for the purpose of enabling users to reflect on their general well-being. Similar techniques can now be found in several commercial smartwatches which use proprietary software to estimate stress levels based on a combination of biomarkers (most commonly HRV, TEMP and EDA) [37] for the purpose of self-monitoring. The ultimate purpose of these works are for end-users' understanding on general well-being, but not to support learning tasks.

The closest work to ours has investigated the potential of visualising electrodermal activity (EDA) in education. This includes a prototype designed by Echeverria et al. [13] that showed the arousal peaks from students participating in a high-fidelity, team simulation as coloured dots in a timeline graph. Results from a study with teachers suggested that they could envisage potential uses of the visualisation to provoke student reflection on emotional experiences. Fernandez-Nieto et al. [15] presented also a timeline visualisation of arousal by modelling arousal into a 5-level scale visualisation by counting the arousal peaks detected from physiological EDA data. Results from a study with students inspecting their own data also showed promise in sparking reflection on their activity, but students also questioned how the arousal levels were detected. Similar studies have explored ways to visualise arousal using alternative techniques (e.g., tabular and text summaries [16]). Their findings suggested that teachers emphasised the importance of knowing more details about arousal detection to be able to trust on the visualised indicators. Indeed, exploring how to build the users trust regarding LA dashboards has been identified as a knowledge gap beyond the visualisation of physiological signals [50].

2.3 Research gap and research questions

In short, there has been a growing interest in modelling and visualising physiological information in LA and beyond. The closest work to our own has was limited to visualising changes in EDA only. However, little attention has been given to i) advancing visual methods for communicating stress information in educational situations purposely designed by teachers for students to experience stress so they can reflect on how to manage it; and ii) exploring the extent to which teachers trust stress measurement and recognise potential uses of stress visualisations to support reflective activities in stress-inducing scenarios. Against these gaps, we propose the following three research questions:

RQ1: How does a stress visualisation **contribute to the sensemaking** of a stress-inducing learning scenario?

RQ2: How is teacher **trust** influenced when models of students' stress based on physiological data are compared to an alternative stress biomarker (i.e., cortisol measures)?

RQ3: How do teachers **envisage the adoption** of a stress visualisation to support their teaching practice?

3 STRESSVIZ: THE STRESS ANALYTICS DASHBOARD

3.1 The learning situation and dataset

We designed *StressViz* to support teacher-led reflective thinking [11] in the context of an educational de-escalation simulation designed to imitate a real-life healthcare conflict event. The purpose of the simulation was to deliberately provoke tension in students so they could learn how to develop aggression management skills and stress-coping strategies. The simulation was part of the compulsory curriculum of the undergraduate Nursing program of the Monash University.

A total of 20 students, organised in four teams (simulation 1-4) of five members each, volunteered and consented to participate in the study. An actor who specialised in supporting healthcare simulations played the role of a furious patient who frequently yells at the nurses (see Figure 1). Each student in a team played one of the following roles: a registered nurse (RN); a student nurse (SN); the patient's wife; an observer of the RN (OBS1); and an observer of the patient (OBS2). RN and SN are the main participants of the simulation. They arrive at the patient's house and are required to deal with the confronting situation. Observers are responsible for recording the reactions and speech of the main participants, but they do not participate. Each simulation is around 4-5 minutes long and was facilitated by a team of three nursing teachers. Table 2 represents the learning design of each simulation created by the teachers, which includes 1) an introductory briefing, 2) the actual immersion in the simulation, and 3) the reflective debrief. Teachers divided the immersion into five phases. The actor was asked to deliberately increase the aggression according to the levels indicated in column 3 of Table 2. For example, Phases 4 and 5 were intended to bring the aggression to its climax.

Stages		Description	Intended aggression
1- Briefing		The patient and nurses prepare for the upcoming simulation session.	Moderate
2- Immersion	Phase 1	The wife welcomes the nurses and takes them to where the patient is.	High
	Phase 2	Small talk conversation.	Moderate
	Phase 3	The patient becomes irritable when was told he needs another injection.	High
	Phase 4	The nurses start to calm the patient, but the patient becomes more aggressive.	Very High
	Phase 5	The wife presents a cup of tea, which triggers the patient to escalate further.	Extreme
3- Debrief		Reflection guided by a teacher.	Moderate

Fig. 2. The learning design of the de-escalation simulation with intended aggression levels per phase to be enacted by the patient.

After receiving ethics approval for this research, students provided informed consent for their physiological and cortisol data to be captured. Physiological data were captured using Empatica E4 wristbands worn by each student in their non-dominant hand. These recorded skin temperature (TEMP) at 4 Hz, electrodermal activity (EDA) 4 Hz, blood volume pulses (BVP) at 64 Hz, and three-axis acceleration (ACC) at 32 Hz. Researchers synchronised the Empatica E4 wristbands together before the start of the simulations. All simulations were video-recorded. Salivary cortisol (ng/ml) data of all the students were collected before and after each simulation. Cortisol is a hormone produced by the adrenal glands that regulates several important functions in the human body in response to stress. Several studies showed salivary cortisol is a stress biomarker [21].

3.2 Modelling stress

3.2.1 Data pre-processing. We trained generalised stress detection models based on the multimodal, publicly available WESAD dataset [45]. The rationale was to provide a cold start model that could detect the stress and relaxation periods of a student that can be replicated without requiring a baseline collection step as it would normally be done in controlled experiments. The WESAD dataset includes physiological data recorded from a wristband (Empatica E4) worn by 15 subjects in various conditions. The dataset was labelled based on types of tasks requested to participants to complete, each designed to provoke three different affective states: *neutral*, *stress*, and *amusement*. The first label corresponded to instances when the participants were sitting or standing by a table reading some text; the *amused* label corresponded to instances when participants watched funny videos; and the *stress* label corresponded to instances reported as stressful while participants engaged in public speaking and in solving complex arithmetic tasks. The reported accuracy of models created with the WESAD dataset using only the wrist-based physiological data was 88.33% (Random Forest – RF) [45].

As we did not have access to the models reported in [45], we replicated their process using the available dataset. Specifically, we segmented the sensor signals into non-overlapping windows with 10 second shifts. Then, we extracted the following 28 features for each segment from the four sensors embedded in the Empatica E4 wristbands (ACC, TEMP, BVP, EDA) as in [45]:

ACC pre-processing: From the raw 3-axis acceleration signal, we extracted the mean, standard deviation and max values for each axis (x, y, z) separately. Net acceleration $NetAcc_i$ was also computed as $\sqrt{xi^2 + yi^2 + zi^2}$ [54]

TEMP pre-processing: From the raw TEMP signal, we extracted the mean, standard deviation, min and max values. In addition, the slope of the signal $\Delta TEMP$ was used as a feature .

BVP pre-processing: From the raw BVP signal, we extracted the mean, standard deviation, min and max values. In addition, the peak frequency was computed for each frame $f_{peak,i}$

EDA pre-processing, Artifact removal: Since the electrodermal activity signal is vulnerable to several types of noisy artefacts when using wearable devices, the noise should be filtered before performing feature extractions [51]. Therefore, we used the EDA Explorer tool [51] to detect such artefacts in data. If a data segment was detected as a noisy artefact, it was excluded from the feature extraction process. After removing such artefacts, features were extracted from EDA. We extracted two kinds of skin conductance: tonic and phasic. The tonic component, skin conductance level (SCL), includes more long-term slow drifts and spontaneous fluctuations. The phasic component, skin conductance response (SCR), reflects more short-term responses to the stimulus. We employed *cvxEDA* tool[1] that decomposes the EDA signal into these two separate components using a convex optimisation approach [20]. Then, statistical features for SCL and SCR were computed (mean, standard deviation, min, max).

3.2.2 Model evaluation. For consistency with the original models trained on the WESAD data, we trained five machine learning models (SVM, AdaBoost, RF, LDA and KNN as in [45]) using the sklearn library[2]) using all the features. Results from the classifiers were analysed and compared. We note that the class distribution was unequal for the three affective states. The labels corresponding to neutral, stress and amusement states account for 53%, 30% and 17% of the dataset, respectively. This could potentially introduce bias to the modelling result in our dataset and give a poor accuracy in categories with a small number of instances (i.e., stress and amusement labels). Yet, comparing the performance of the employed classifiers, SVM, RF and KNN reached the high accuracy scores (86%, 88% and 88%, respectively) on the test data based on an 80:20 train-test split from the WESAD dataset. We thus decided to use the RF model for predicting the labels on our dataset (Section 3.1), which was also the best performing classifier by the authors of the WESAD dataset [45] and in [17]. This suggests that we replicated the results of their modelling to a great extent. Labelled segments were further grouped by each phase of the simulation to create the final visualisation of stress data.

3.3 Visualising stress

StressViz was designed as a timeline visualisation by following teachers' learning design depiction of the multiple phases of the learning task (as shown in Table 2) and related previous work [13, 15]. A timeline visualisation is typically used to depict a series of events by plotting them along a time axis at the instant or interval over which they occur [41]. Teachers who created the learning design were interested in monitoring both the stress levels within a team and also in

[1] cvxEDA algorithm - https://github.com/lciti/cvxEDA
[2] Scikit-learn - https://scikit-learn.org/

observing the stress levels for a particular role (e.g., the RN) across teams. Therefore, we designed two versions of the *StressViz* dashboard, one which showed stress data of all students on a given team (figure 3); and another that showed the stress data of all students with the same role in different teams (figure 4). In the former, the y-axis represents the roles on the team; in the latter, the y-axis represents the different teams.

In both versions, the x-axis represents each phase of the simulation as per the learning design. The phase tagged as "Before SIM" represents the students' physiological data 30 seconds before the start of the session. The squared blocks represent either the most frequent label of the data segments in that phase (if the proportion of the most frequent label was higher than 50%) or the next two most frequent labels (if the proportion of the most frequent label was lower than 50%).

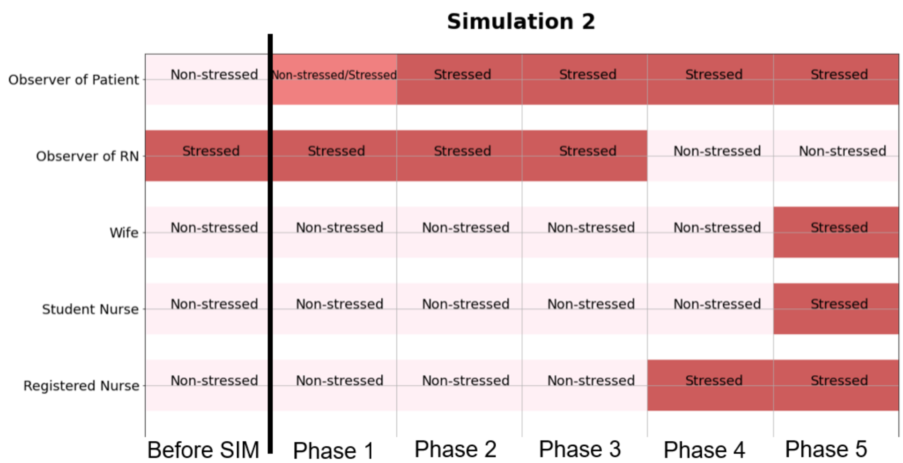

Fig. 3. StressViz of the whole team in simulation 2. Blocks represent the majority of the labelled segments encountered in each phase for each student.

The light pink "nonstressed" blocks represent the neutral condition (or no stress detected). The coral "nonstressed/ stressed" blocks represent phases with an equivalent number of segments labelled as neutral and stress. The crimson "stressed" blocks shows phases where segments were mostly identified with the stress label. The "amused" label was rarely found and only for the roles that were not active participants (e.g., the student who enacted the role of Observer in Simulation 3 in Figure. 4).

4 METHODS

4.1 Participants

The designer of the simulation study (PT1), and five other nursing teachers (PT2, PT3, PT4, PT5, and PT6 with 18, 10, 14, 12, and 17 years of teaching experience, respectively) who have led similar simulations, participated in the study. All participants were experienced teachers in nursing (5 of them identified themselves as females and 1 as male) and had signed consent forms before attending the study.

4.2 Study design

Participants were interviewed with the purpose of documenting their perspectives and responses to our research questions. Each interview was recorded using an online video conferencing platform (i.e., Zoom) and had an approximate

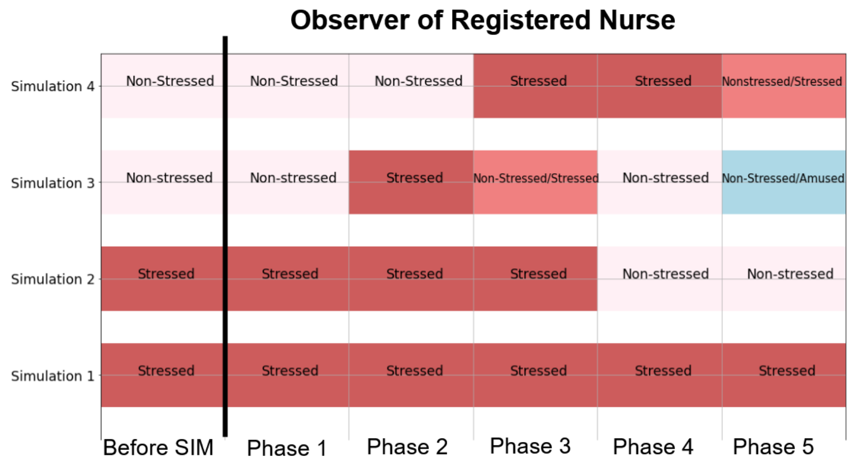

Fig. 4. StressViz for all students who played the role of 'Observer of Registered Nurse' across simulations

duration of 45 – 60 minutes. Each interview was fully transcribed using a professional service. Following a semi-structured format, the interview had three parts, each addressing one research question (RQs 1, 2, and 3, respectively):

Part 1: A think-aloud protocol was followed to document teachers' perspectives on the visualised stress data (RQ1: the sensemaking of stress visualisations). The visualisations were presented to teachers in two ways: (i) stress data of students in the same randomly chosen team; and (ii) stress data of students with the same role in different teams. For the case of (i), we defined prompt questions that required teachers to compare stress data of different student roles in particular phases of the simulation (namely, 1) during the briefing; 2) phases 1-3 –*aggression is escalating*–, and phases 4 and 5 –*aggression gets to its climax*–). For the case of (ii), we defined prompt questions that required teachers to verbalise their observations on stress levels of the same roles (RN, SN and Observer of the RN) across all simulations.

Part 2: This part focused on eliciting participants' insights on the *trustworthiness* of the stress visualisations with reference to the changes in students' cortisol before and after each simulation (RQ2). Cortisol levels are common indicators of stress although the brain can be desensitised to it and drop the physiological responses to stress [56]. This means that a perfect correlation between cortisol and the physiological markers used to model affective states for *StressViz* is not expected [18]. Instead, we aimed at letting teachers observe how the cortisol data, as an alternative biomarker, either supported or contradicted the modelling results in some instances in order to prompt discussion about the trustworthiness of the system. This was operationalised by enhancing the stress data of students with the same role shown in Part 1 of RNs and SNs (the primary roles) with the corresponding cortisol data (shown in Figure 5).

Part 3: Finally, teachers were asked to respond to some open-ended questions to understand how teachers might **envisage the adoption** of the stress-awareness dashboard (RQ3). The questions were carefully selected based on the LA-TEP protocol for designing human-centred LA tools [34]. Some of the questions included the following: (i) Can you think about any potential usage of these visualisations?; (ii) Who do you think should look at those visualisations (i.e., students and/or teachers); (iii) To what extent can such visualisations be incorporated into the teachers' feedback system to reflect on stress?; (iv) To what extent do you think showing such visualisations to students could improve or negatively affect their learning experience?; (v) To what extent do you think using *StressViz* will influence your teaching practice over time (as opposed to short term impact)?

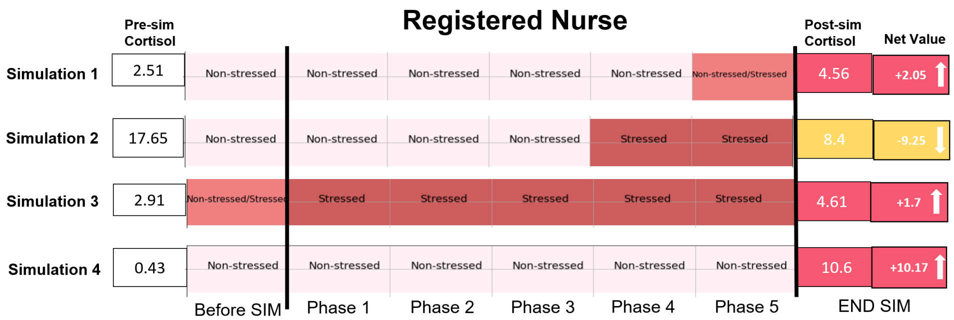

Fig. 5. StressViz for all students who played the role of ‘Registered Nurse‘ (RN) across simulations with cortisol data shown at the end of the timeline. Red and yellow blocks and an upwards/downwards arrow represent an increase or decrease in cortisol, respectively.

4.3 Analysis

We conducted inductive thematic analyses [6] of the interview data to address each of the three questions. The video data and the transcripts were first triangulated and screened by one researcher. Then, two more researchers participated in the analysis of the video recordings. Since all decisions were made simultaneously and through consensus, inter-rater reliability was not necessarily required [35]. The analysis of RQ1 consisted in a vignette analysis of Part 1 of the interview. The three researchers sought instances illustrating how teachers talked about stress based on the dashboard. The researchers selected two vignettes that could potentially point to insights and contradictions across teachers inspecting the same visualisations. For RQ2, the same researchers analysed Part 2, seeking instances that illustrated how participants talked about their trust on the modelling and the cortisol data. For RQ3, statements of interest from Part 3 were jointly identified by two researchers. The resulting statements were reviewed by another researcher until full agreement was reached. Next, all researchers had several discussions to identify and group emerging themes in alignment with the research questions.

5 RESULTS

5.1 RQ1 – Sensemaking of the stress-inducing learning scenario

5.1.1 Vignette 1: sensemaking of a team’s stress data. This vignette illustrates how teachers explored one instance of *StressViz* that showed the stress levels of the team in Simulation 2 (shown in Figure 3). It started by asking participants to reflect on the preparation phase (i.e., before the sim). Most participants thought that the results from the modelling were believable, but they also expected some stress to be experienced by students before the simulation. For example, the simulation designer (PT1) explained this as follows: *“In anticipation, I would have thought they’d have been just mildly stressed. It may be indicative that they felt prepared”*. PT2 also believed that observing some stress being flagged by the dashboard would be positive. This was explained as follows: *“They should have some form of stress because of the unknown of what’s going to happen in the next few minutes”*. For simulation 2 though, the model detected a high level of stress for Observer of RN before the simulation even started. PT3 explored potential explanations for this as follows: *“maybe she didn’t prepare in advance, or she didn’t know too much about the simulation, or had personal issues.”*

Regarding phases 1, 2 and 3, in which the aggression was steadily increasing, all teachers expected the stress level of SN and RN to increase as they directly dealt with the aggressive patient. While this increase was observed in some teams, in others, such as the one in simulation 2, these roles did not show many segments labelled as stressed. PT3 reflected on this as follows *“I am kind of surprised that their stress [RN, SN, wife] was just normal, you know?”*. PT2 mentioned that it may be possible for some nursing students to start managing the situation because they were trying to enact active roles: *"The more experience the students have in dealing with different types of people, the better they are at managing their stress. Maybe she (RN) was in control of the situation. The student nurse (SN) depended on how well the RN was dealing with the situation. If the RN was managing the situation the SN would actually feel nonstressed"*.

In contrast, the data for both observers (of the RN and the patient) were labelled as stressed for most of the simulation (first two rows in Figure 3). Here, PT4 expressed surprise: *"It is very interesting that people who were just looking at the situation, were experiencing heightened stress levels!"*. PT2 also interpreted the data as follows: *"This is a perfect depiction of how an observer of a patient would probably feel if [they] haven't encountered abusive or agitated patients before."* PT1 provided a potential explanation regarding why observers may be more stressed than the actual active participants: *"They want to rescue or help someone and they're not able to do so. They would feel very stressed if they could not help"*.

For the aggression climax (phases 4 and 5), all participants agreed that this was the point where all the students were stressed. PT1 explained the possible reasons why students were stressed, as follows: *"This, I imagine, happened because what they're doing was not working to deescalate the scenario. They were becoming stressed because they didn't know what they could do next"*. PT4 questioned why there was not a middle point for stress detection in the dashboard: *"Interestingly, for the RN, stress goes from nothing to a load of stress, rather than that middle range stress option"*. For the observer of RN, PT2 mentioned that *"The strangest part is the observer of the RN not being stressed in the phases 4 and 5 when that's the time when the other students were getting more and more anxious and angry. That's kind of very interesting, it tells us a lot that we are all different and we managed stress differently"*.

5.1.2 Vignette 2: sensemaking of the same roles across different teams. The first visualisation of this vignette focuses on teachers' reflections on how students with the same role (namely the RN) behave similarly or differently across simulation sessions. When the teachers explored the visualisations of the RNs (similar to the on in Figure 5 – without cortisol data), all agreed that results were mixed. PT2 mentioned that the stress levels of the RN in simulation 3 (see row 3 in Figure 5) matched what was expected in the learning design, as follows: *"if you consolidate the information and if you look at the different phases in this scenario, it brings to the mind that simulation 3 was the most stressful and normal situation"*. For the result of simulation 4, where the RN did not have any blocks with the majority of segments labelled as 'Stressed', PT4 thought the student was extraordinary and explained: *That's really someone that's walked through that whole situation and did not feel any stress, that's quite incredible. That student should be a lion tamer or something!"*.

Next, when participants looked at the second visualisation of this vignette, which shows observers of the RN across teams (see Figure 4), all participants agreed that it was interesting to see that the observers were overall more stressed than the nurses who were actually dealing with the patient. PT4 argued though that this pattern was not consistent for all students: *"It's reflecting really different patterns. There's not a consistent pattern, we can't draw an assumption about what part of the simulation was more stressful for everyone, because everyone is experiencing it very differently"*. Moreover, for the "Nonstressed/Amused" label in simulation 3 (see row 2 in Figure 4), PT3 believed that the student was not engaged in the simulation scenario, and PT5 suggested that observers could also detach themselves from the situation, as follows: *"at that particular point in time, this observer realised it is not real"*.

In short, Vignettes 1 and 2 presented in this section illustrate how teachers made sense of the stress labels by triangulating the data with the learning design and their own experience as nursing educators. Teachers could articulate potential explanations to trends and variations in student modelled data. Yet, we learned that: i) teachers managed to provide explanations about how students may have felt; and ii) in part one of the study, teachers did not question the validity of the interface. The latter is further explored in the next section.

5.2 RQ2 – Teachers' trust on the stress modelling outputs

In this part, we focus on the alignment between the stress visualisation and cortisol data on the main roles (RN and SN) during different simulation sessions to understand to what extent teachers trusted the dashboard. From a quantitative point of view, the sample was too small to perform a valid correlation analysis. In any case, the modelling results and the cortisol data did not correlate in this small sample and this was made visible to teachers in Figure 5. Indeed, all teachers could also not see any relationship between the stress data and the cortisol data. As a result, they questioned the validity of the stress data or cortisol data. For example, PT1 argued: *"I just don't understand why that one would have dropped so significantly in cortisol when they've got a high level of stress there [detected by the algorithm]. Would it be the accuracy of the cortisol collection? It shouldn't be dropping."*

PT2 also questioned stress models as follows: *"But this doesn't always add up, isn't it? Because when you're stressed out, your cortisol levels need to go high. But here in simulation 4, the dashboard says students were not stressed at all. But they had the maximum amount of cortisol level. [...] Is it accurate enough for us to actually come up with any sort of concrete results there?"*. PT1 and PT6 also gave some possible explanations regarding the potential issues with using physical sensors. PT1 explained this as follows: *"Maybe one of the wristbands was not as sensitive and perhaps because of this, it was not that accurate?"* and PT6 explained: *"In Sim 3, whether it's not picked up on the cortisol levels, or whether the automation isn't picking up, it's missing some cues, or misinterpreting cues to say that they were stressed. [...] Maybe, it's a manageable stress because it's not a big peak of stress in cortisol levels"*.

In sum, the introduction of cortisol data as a potential benchmark for teachers to compare the physiological data prompted teachers to wonder whether the cortisol or the physiological data was more or less accurate for indicating stress levels. This shifted their thinking from making strong statements about students' experienced stress to recognise the complexity of detecting stress reliably.

5.3 RQ3- Envisaged adoption of StressViz in teaching practice

Under this theme, we summarise the teachers' views on how the stress visualisation could be integrated into their existing professional teaching practice. The first three sub-themes address how the stress data could be potentially used to facilitate the process of professionalisation for students, teachers and researchers, whereas the last sub-theme summarises the long-term impacts if those kinds of visuals could be widely validated and used.

Supporting nursing students. All teachers agreed that *StressViz* could provoke reflection during the debriefing and could be used as an additional resource to consider when developing stress-coping strategies. More specifically PT1 and PT6 explained the potential role of this tool in supporting the debrief. For instance, PT1 said: *"When you go back to debriefing, you can then reflect on using that data and say, 'look, at this point, you seem very stressed, what was happening for you at that moment?' and 'Can you explain to me what was going on in your thinking processes?'"*. Regarding the potential role of the dashboard in developing stress-coping strategies for other situations, PT1 explained *"So not only did I worry about the de-escalation of the scenario. You could also extrapolate for stress [experienced] in any scenario. It could be cardiovascular resuscitation, or a major trauma [situation]. You're going to get stressed during those situations too and if you can get to understand that yourself better, you can then employ some stress relieving activities, like breathing.*

Supporting nursing teachers. All the participants agreed that *StressViz* can also potentially be used to support teachers in investigating extreme cases of student stress and, in turn, develop better teaching strategies. PT3 explained this as follows: *"If a facilitator has this tool and realises that some students are super stressed, she or he can work on that, going deep on that person, to know what is going on with him or her, like asking more questions, addressing any concerns,* *and helping them have a meaningful learning experience. Otherwise, that student may leave the room, maybe thinking, 'oh, it was horrible'"*. PT5 and PT6 further added that *StressViz* can also be used in the transition of graduate nurses to professional practise context. PT6 said: *"I would be tempted to use this in transition to practice. Because as a graduate nurse, you're going to experience stress over and over again, and then we can use [the dashboard] as a way to talk about strategies for dealing with it before they have to actually do it"*.

Regarding the potential use of the tool to support learning design, PT3 and PT4 added that these stress visualisations could be used by teachers to assess how different students react to the same simulation design to calibrate this design. For example, PT3 explained this as follows: *"I think it would be helpful for managing expectations that we have of nursing students, but also helping students understand what happens when they're under stress in different situations, and how people respond differently in different contexts to the same scenario"*.

Finally, PT2 added that using *StressViz* would be beneficial if appropriate training and data literacy required are considered in their deployment: *"Most of us haven't learned basic statistics, we haven't learned how to read basic charts, how to basically understand a table with information. So, if you're able to explain it to educators and teachers, then the nursing students would really benefit from it. I definitely love for students to start thinking out of the square rather than just follow certain principles and focus only on anatomy and physiology. They would benefit a lot., but they need to be educated."*

Supporting educational and nursing researchers. All participants suggested *StressViz* could help researchers in doing nursing research, and they all had varied perspectives. For example, PT1 focused more on the potential of measuring how nurses' stress levels can relate to the patient's outcome: *"I want to see something that research can do to enhance patient outcomes. So that's why it's important that you can do this work with healthcare researchers"*. PT2 explained that researchers would be able to interpret the data, learn from the results and find out better ways to do future research: *"These visualisations can help nursing researchers understand how different people feel in each phase of a simulation: compare, contrast and analyse the data, and learn from the research that we do in nursing education"*.

Long-term impact. All teachers agreed that once the visualisation process is fine-tuned and validated over time, it could be useful for both professional teaching and learning, as well as patient care. PT1 and PT4 highlighted that educators and students could develop some stress-coping strategies in the long run: *"When you're doing stress inducing simulations, having these stress visualisations as a component of it would enrich the teaching and the student educational experience. The educator and the student can work on building strategies to manage stress. I think in the near future we will be able to measure stress levels in this way with much more accuracy and clarity"*. PT3 and PT5 believed that LA tools will become the future trend for nursing practices if used appropriately and this innovation can contribute to this. PT3 further explained as follows: *"Teachers need to know how to use these analytics. It is a new concept, so it's going to take a little while, but it's going to become a part of the curriculum in the future. So, if we are going to deal with visualisations, artificial intelligence and machine language, we need to be able to incorporate it and teach the teachers appropriately so that they can teach the students well"*.

5.4 Challenges, Risks, and Limitations

This theme collates the teachers' opinions about the potential challenges, risks, and limitations of using stress visualisations in practice. PT2 explained that the teachers who are using *StressViz* need to understand the basic mechanisms behind it: *"My only concern is the teachers who are providing the feedback need to actually understand the essence behind it, they need to understand each and every block of what you're trying to explain to the student. If this is understood then it would be really good"*. Similarly, PT4 explained that: *"[the effectiveness of StressViz] could also depend on what explanation* *teachers give, what support teachers give around understanding the feedback. It's not just about giving the visualisation to students, but it's also about the scaffolding of how we help them."*

PT2 raised ethical concerns about the potential misuse of incorporating the visualised stress data into the teachers' feedback strategy: *"As long as you're objective, as long as you take the visualisations about a person's behaviour and use that to aid them in understanding their own behaviour, it's really good. But if you're going to use that as subjective and try not to promote that person, then it can become detrimental to that person"*. PT5 further added: "I think you need more qualitative data such as real-time feedback responses from students, for sure." PT1 was also concerned about the situation or scenario where the tool can be applied with integrity: *"I wouldn't be doing it for exams. I think it's too intrusive. I think this really is for the domain of simulation personally. Because I think in tutorials [or regular classroom teaching] you want to have a more conversational type of experience with students, and you don't want to be stopping and looking at data the whole time. It's a very different dynamic in a tutorial versus a simulation"*.

6 DISCUSSION

6.1 Research questions

Dewey [11] explained that deeper reflective thinking allows for continuous interpretation and investigation while employing the information gained from past experiences to inform and guide new actions. In terms of RQ1, our results suggest that *StressViz* can provide teachers with evidence that provokes reflection. Teachers interpreted the visualisations based on the expected increase in aggression as per the simulation script (the learning design). While most teachers suggested that further actions would involve to more deeply investigate student perceptions of experienced stress (i.e., scaffolding student reflection), some teachers made some strong assumptions about the stress levels of students based on *StressViz* alone, suggesting they initially trusted the modelling.

Yet, regarding RQ2, as in other fields [4], we are still far from creating trustworthy and reliable stress detectors in LA [25]. When comparing the physiological modelling with cortisol levels as an alternative biomarker, teachers questioned the validity of either the modelling or the cortisol data. Indeed, although cortisol is a common indicator of stress, the physiological reactions to stress can be subject to intentional control [18] and cortisol levels may also be affected throughout the day by multiple factors such as food, coffee or the time of the day [56]. Yet, using this biomarker as a benchmark enabled us to unpack potential future directions in end-user interface design for physiological data. First, there is a risk of over interpretation as we observed in contrasting the statements from teachers in the vignettes (RQ1) and their concerns about validity when presented with the cortisol data (RQ2). Teachers need to be properly trained so they understand the essence behind the LA tools in order to trust them and make informed decisions [42]. Second, there is a lack of large open datasets that could be used to train generalised models. In turn, further research is needed to generate these data so that the detection of stress can become more valid and reliable.

In terms of RQ3, nursing educators foresee the adoption of the *StressViz* to assist their teaching practice. Specifically, they envisaged StressViz as a *companion tool* to help them in supporting 1) nursing students who may have experienced extreme levels of stress in the debriefing to develop better stress-coping strategies; 2) graduate nurses transitioning to professional placements; 3) the establishment of teaching strategies to focus on students' feelings after stress-inducing scenarios; and 4) the calibration of the learning design and the intensity of such scenarios. Teachers also identified the potential contribution to nursing research which can translate into actual patient outcomes. These envisaged uses have wider implications for simulation-based learning. In fact, there is an increasing interest in making use of physiological data healthcare simulation research [27] and in other simulation-based training disciplines such as flight simulation training [44]. The study presented in this paper is thus critical since people are already increasingly using smartwatches that provide indicators of stress based on black box algorithms [37]. This paper can be seen as an initial attempt to understand potential future avenues of research that are needed to develop ways to visualise physiological data in ways that are aligned for education and training purposes with integrity.

6.2 Ethical issues and responsible use

Ethical issues may arise because the data collected to measure stress includes students' (*under-the-skin*) physiological data which would not be normally accessible to teachers and other stakeholders [9]. Some teachers expressed concerns about unintended misuses of those data for decision making that can adversely affect students, teachers, and broader educational systems. Therefore, we suggest that the collection and usage of such data should be limited to educational scenarios that deliberately aim at helping students experience stressful situations, experience the stress, and reflect upon it to develop coping strategies. In line with what the teachers in our study suggested and contemporary LA dashboard literature [53], we discourage using stress modelling and visualisation for surveillance in scenarios such as exams and regular classrooms, or for measuring students' performance.

6.3 Limitations and Future Work

Our study has some limitations. First, the WESAD dataset used to model students' stress is relatively small and it was collected under experimental conditions which differ from those of a nursing situation. The modelling conducted using WESAD should thus be seen as a device to investigate important end-user interface questions about teachers' perceptions of stress visualisation. Although using cortisol data in regular classes is not feasible and it is heavily discouraged, in this study it enabled investigating important questions about trustworthiness when teachers are presented with an additional biomarker. Future work could explore using other datasets or develop more advanced methodologies for stress detection, and validate how stress can be measured accurately at scale. Second, Empatica E4 wristbands are research-grade sensors that are expensive and have shortcomings when they are deployed in settings where students can move freely. Given that sensors of various kinds are rapidly evolving, future research could consider other kinds of sensors that are more affordable and that can provide with similar data granularity. Third, since this study was embedded into an actual class, even though the patient was expected to escalate their aggression equally across all sessions, some sessions may have been more intense than others. Further work that focus on stress modelling needs to empirically investigate potential strategies to improve the accuracy of stress models, within similar or other learning contexts. Finally, whilst teachers' perspective is important to identify potential pedagogical uses of the tool, future work should investigate the perspective of the students on the visualisation of their own data.

7 CONCLUSION

Very few studies have addressed the challenge of visualising data to support reflection on students' stress. This paper presented the *StressViz* dashboard that decodes students' physiological data into stress labels for different phases of simulation in an authentic nursing teamwork context; and a qualitative study with teachers that sparked ideas about the potential of visualising stress as well as current limitations in the validity of stress detection, ethical concerns and potential misuses of such interfaces. We anticipate that this paper can serve as an inspiration for conducting further research in the development of stress visualisation tools to support stress-inducing learning situations.

ACKNOWLEDGMENTS

Riordan Dervin Alfredo gratefully acknowledges the Monash University for his PhD scholarship. This research was at least in part funded by the Australian Research Council (DP210100060) and Jacobs Foundation (Research Fellowship).

REFERENCES

[1] H. Alwahaby, M. Cukurova, Z. Papamitsiou, and M. Giannakos. 2021. The evidence of impact and ethical considerations of Multimodal Learning Analytics: A Systematic Literature Review. *EdArXiv* (2021). https://doi.org/10.35542/osf.io/sd23y

[2] C. Aucott and A. Soni. 2016. Reflections on the use of Critical Incident Stress Debriefing in schools. *Education Psychology in Practice* 32, 1 (2016), 85–99. https://doi.org/10.1080/02667363.2015.1112257

[3] R. L. Bailey. 2017. Electrodermal activity (EDA). *The international encyclopedia of communication research methods* (2017), 1–15.

[4] M. B. Bin Heyat, F. Akhtar, S. J. Abbas, M. Al-Sarem, A. Alqarafi, A. Stalin, R. Abbasi, A. Y. Muaad, D. Lai, and K. Wu. 2022. Wearable Flexible Electronics Based Cardiac Electrode for Researcher Mental Stress Detection System Using Machine Learning Models on Single Lead Electrocardiogram Signal. *Biosensors* 12, 6 (2022), 427. https://doi.org/10.3390/bios12060427

[5] C. L. Bong, K. Fraser, and D. Oriot. 2016. *Cognitive Load and Stress in Simulation.* Springer. 3–17 pages. https://doi.org/10.1007/978-3-319-24187-6_1

[6] V. Braun and V. Clarke. 2012. Thematic Analysis. APA, Washington, DC, 57–71. https://doi.org/10.1037/13620-004

[7] M. Bustos-López, N. Cruz-Ramírez, A. Guerra-Hernández, L. N. Sánchez-Morales, N. A. Cruz-Ramos, and G. Alor-Hernández. 2022. Wearables for Engagement Detection in Learning Environments: A Review. *Biosensors* 12, 7 (2022). https://doi.org/10.3390/bios12070509

[8] J. A. Castro-García, A. J. Molina-Cantero, I. M. Gómez-González, S. Lafuente-Arroyo, and M. Merino-Monge. 2022. Towards Human Stress and Activity Recognition: A Review and a First Approach Based on Low-Cost Wearables. *Electronics* 11, 1 (Jan. 2022), 155.

[9] M. Cukurova, M. Giannakos, and R. Martinez-Maldonado. 2020. The promise and challenges of multimodal learning analytics. *Br J Educ Technol* 51, 5 (2020), 1441–1449. https://doi.org/10.1111/bjet.13015

[10] F. de Arriba-Pérez, J. M. Santos-Gago, M. Caeiro-Rodríguez, and M. Ramos-Merino. 2019. Study of Stress Detection and Proposal of Stress-Related Features Using Commercial-off-the-Shelf Wrist Wearables. *Journal of Ambient Intelligence and Humanized Computing* 10, 12 (2019), 4925–4945.

[11] J. Dewey. 1910. *How We Think.* D C Heath, Lexington. https://doi.org/10.1037/10903-000

[12] S. D'Mello. 2017. Emotional Learning Analytics. (2017), 115–127. http://solaresearch.org/hla-17/hla17-chapter10

[13] V. Echeverria, R. Martinez-Maldonado, and S. Buckingham Shum. 2019. Towards collaboration translucence: Giving meaning to multimodal group data. In *Proc. of the 2019 CHI conference on human factors in computing systems*. ACM, 1–16. https://doi.org/10.1145/3290605.3300269

[14] R. M. Fanning and D. M. Gaba. 2007. The Role of Debriefing in Simulation-Based Learning. *Simulation in Healthcare* 2, 2 (2007).

[15] G. M. Fernandez-Nieto, V. Echeverria, S. B. Shum, K. Mangaroska, K. Kitto, E. Palominos, C. Axisa, and R. Martinez-Maldonado. 2021. Storytelling With Learner Data: Guiding Student Reflection on Multimodal Team Data. *IEEE Transactions on Learning Technologies* 14, 5 (2021), 695–708.

[16] G. M. Fernandez-Nieto, K. Kitto, S. Buckingham Shum, and R. Martinez-Maldonado. 2022. Beyond the Learning Analytics Dashboard: Alternative Ways to Communicate Student Data Insights Combining Visualisation, Narrative and Storytelling. In *LAK22: 12th Int. Learning Analytics and Knowledge Conf.* (Online USA). ACM, 219–229. https://doi.org/10.1145/3506860.3506895

[17] P. Garg, J. Santhosh, A. Dengel, and S. Ishimaru. 2021. Stress Detection by Machine Learning and Wearable Sensors. In *26th Int. Conf. on Intelligent User Interfaces - Companion (IUI '21 Companion)*. ACM, 43–45. https://doi.org/10.1145/3397482.3450732

[18] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis. 2019. Review on Psychological Stress Detection Using Biosignals. *IEEE Trans. Affective Comput.* 13, 1 (July 2019), 440–460. https://doi.org/10.1109/TAFFC.2019.2927337

[19] F. Gioia, A. Greco, A. L. Callara, and E. P. Scilingo. 2022. Towards a Contactless Stress Classification Using Thermal Imaging. *Sensors* 22, 3 (Jan. 2022), 976. https://doi.org/10.3390/s22030976

[20] A. Greco, G. Valenza, A. Lanata, E. Scilingo, and L. Citi. 2016. cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing. *IEEE Transactions on Biomedical Engineering* (2016), 1–1. https://doi.org/10.1109/TBME.2015.2474131

[21] D. H. Hellhammer, S. Wüst, and B. M. Kudielka. 2009. Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology* 34, 2 (Feb. 2009), 163–171. https://doi.org/10.1016/j.psyneuen.2008.10.026

[22] A. Horvers, N. Tombeng, T. Bosse, A. W. Lazonder, and I. Molenaar. 2021. Detecting Emotions through Electrodermal Activity in Learning Contexts: A Systematic Review. *Sensors* 21, 23 (2021). https://doi.org/10.3390/s21237869

[23] S. Järvelä, D. Gašević, T. Seppänen, M. Pechenizkiy, and P. A. Kirschner. 2020. Bridging learning sciences, machine learning and affective computing for understanding cognition and affect in collaborative learning. *Br J Educ Technol* 51, 6 (2020), 2391–2406.

[24] H. Jeong. 2012. A comparison of the influence of electronic books and paper books on reading comprehension, eye fatigue, and perception. *Electronic Library* 30, 3 (June 2012), 390–408. https://doi.org/10.1108/02640471211241663

[25] S. Joksimović, S. Dawson, A. Barthakur, O. Poquet, Y. E. Wang, F. Marmolejo-Ramos, and G. Siemens. 2022. Mapping the Landscape of Social and Emotional Learning Analytics. In *Social and Emotional Learning and Complex Skills Assessment.* Springer, 27–47.

[26] C. Jyotsna and J. Amudha. 2018. Eye Gaze as an Indicator for Stress Level Analysis in Students. In *Int. Conf. on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 1588–1593. https://doi.org/10.1109/ICACCI.2018.8554715

[27] J. Kacer, V. Krivanek, L. Cicmanec, P. Kutilek, J. Farlik, J. Hejda, S. Viteckova, P. Volf, K. Hana, and P. Smrcka. 2019. Physiological data monitoring of members of air forces during training on simulators. In *World Congress on Medical Physics and Biomedical Engineering 2018*. Springer, 855–860.

[28] S. N. Kew and Z. Tasir. 2022. Learning analytics in online learning environment: a systematic review on the focuses and the types of student-related analytics data. 27, 2 (June 2022), 405–427. https://doi.org/10.1007/s10758-021-09541-2

[29] F. Kirschner, F. Paas, and P. A. Kirschner. 2009. A cognitive load approach to collaborative learning: United brains for complex tasks. *Educ. Psychol. Rev.* 21, 1 (2009), 31–42. https://doi.org/10.1007/s10648-008-9095-2

[30] R. Kocielnik, M. Pechenizkiy, and N. Sidorova. 2012. Stress analytics in education. In *Educational Data Mining 2012*.

[31] C. Larmuseau, J. Cornelis, L. Lancieri, P. Desmet, and F. Depaepe. 2020. Multimodal learning analytics to investigate cognitive load during online problem solving. *Br. J. Educ. Technol.* 51, 5 (2020), 1548–1562. https://doi.org/10.1111/bjet.12958

[32] K. Mangaroska, K. Sharma, D. Gašević, and M. Giannakos. 2022. Exploring students' cognitive and affective states during problem solving through multimodal data: Lessons learned from a programming activity. *Journal of Computer Assisted Learning* 38, 1 (2022), 40–59.

[33] R. Martinez-Maldonado, V. Echeverria, O. C. Santos, A. D. P. D. Santos, and K. Yacef. 2018. Physical learning analytics: A multimodal perspective. In *Proc. of the 8th Int. Conf. on Learning Analytics and Knowledge (LAK '18)*. ACM, 375–379. https://doi.org/10.1145/3170358.3170379

[34] R. Martinez-Maldonado, D. Elliott, C. Axisa, T. Power, V. Echeverria, and S. Buckingham Shum. 2022. Designing translucent learning analytics with teachers: An elicitation process. *Interactive Learning Environments* 30, 6 (2022), 1077–1091. https://doi.org/10.1080/10494820.2019.1710541

[35] N. McDonald, S. Schoenebeck, and A. Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, 72 (2019), 1–23. https://doi.org/10.1145/3359174

[36] J. T. Mitchell and A. Dyregrov. 1993. Traumatic stress in disaster workers and emergency personnel. In *Int. handbook of traumatic stress syndromes*. Springer, 905–914. https://doi.org/10.1007/978-1-4615-2820-3_76

[37] A. Morales, M. Barbosa, L. Morás, S. C. Cazella, L. F. Sgobbi, I. Sene, and G. Marques. 2022. Occupational Stress Monitoring Using Biomarkers and Smartwatches: A Systematic Review. *Sensors* 22, 17 (2022), 6633. https://doi.org/10.3390/s22176633

[38] N. Nakayama, N. Arakawa, H. Ejiri, R. Matsuda, and T. Makino. 2018. Heart rate variability can clarify students' level of stress during nursing simulation. *PLoS One* 13, 4 (2018), e0195280. https://doi.org/10.1371/journal.pone.0195280

[39] O. Noroozi, I. Alikhani, S. Järvelä, P. A. Kirschner, I. Juuso, and T. Seppänen. 2019. Multimodal Data to Design Visual Learning Analytics for Understanding Regulation of Learning. *Computers in Human Behavior* 100 (2019), 298–304. https://doi.org/10.1016/j.chb.2018.12.019

[40] O. Noroozi, H. J. Pijeira-Díaz, M. Sobocinski, M. Dindar, S. Järvelä, and P. A. Kirschner. 2020. Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. *Education and Information Technologies* 25, 6 (2020), 5499–5547.

[41] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. 1996. LifeLines: visualizing personal histories. In *Proc. of the SIGCHI conference on Human factors in computing systems*. 221–227. https://doi.org/10.1145/238386.238493

[42] M. J. Reiss. 2021. The Use of al in Education: Practicalities and Ethical Considerations. *London Review of Education* 19, 1 (2021).

[43] J. R. Rudland, C. Golding, and T. J. Wilkinson. 2020. The stress paradox: how stress can be good for learning. *Medical education* 54, 1 (2020), 40–45.

[44] J. S. Sanko and A. Battista. 2019. Making Use of Diverse Data Sources in Healthcare Simulation Research. In *Healthcare Simulation Research*. Springer, 239–247. https://doi.org/10.1007/978-3-030-26837-4_32

[45] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven. 2018. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In *Proc. of the 20th ACM Int. Conf. on Multimodal Interaction*. 400–408. https://doi.org/10.1145/3242969.3242985

[46] S. Semikina. 2014. Stress Data Visualization. Eindhoven University of Technology. https://research.tue.nl/en/studentTheses/stress-data-visualization

[47] K. Sharma and M. Giannakos. 2020. Multimodal data capabilities for learning: What can multimodal data tell us about learning? *Br J Educ Technol* 51, 5 (2020), 1450–1484. https://doi.org/10.1111/bjet.12993

[48] E. Smets, E. Rios Velazquez, G. Schiavone, I. Chakroun, E. D'Hondt, W. De Raedt, J. Cornelis, O. Janssens, S. Van Hoecke, S. Claes, et al. 2018. *NPJ digital medicine* 1 (2018), 67. https://doi.org/10.1038/s41746-018-0074-9

[49] Ö. Sümer, P. Goldberg, S. D'Mello, P. Gerjets, U. Trautwein, and E. Kasneci. 2021. Multimodal engagement analysis from facial videos in the classroom. *IEEE Transactions on Affective Computing* (2021). https://doi.org/10.1109/TAFFC.2021.3127692

[50] T. Susnjak, G. S. Ramaswami, and A. Mathrani. 2022. Learning analytics dashboard: a tool for providing actionable insights to learners. *Int. J. Educ. Technol. High. Educ.* 19, 1 (2022), 1–23. https://doi.org/10.1186/s41239-021-00313-7

[51] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard. 2015. Automatic Identification of Artifacts in Electrodermal Activity Data. In *2015 37th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 1934–1937. https://doi.org/10.1109/embc.2015.7318762

[52] J. Tervonen, S. Puttonen, M. J. Sillanpää, L. Hopsu, Z. Homorodi, J. Keränen, J. Pajukanta, A. Tolonen, A. Lämsä, and J. Mäntyjärvi. 2020. Personalized mental stress detection with self-organizing map: From laboratory to the field. *Comput. Biol. Med.* 124 (Sept. 2020).

[53] K. Williamson and R. Kizilcec. 2022. A review of learning analytics dashboard research in higher education: Implications for justice, equity, diversity, and inclusion. In *LAK22: 12th Int. Learning Analytics and Knowledge Conf.* 260–270. https://doi.org/10.1145/3506860.3506900

[54] J. C. Y. Wong, J. Wang, E. Y. Fu, H. V. Leong, and G. Ngai. 2019. Activity Recognition and Stress Detection via Wristband. In *Proc. of the 17th Int. Conf. on Advances in Mobile Computing & Multimedia*. 102–106. https://doi.org/10.1145/3365921.3365950

[55] L. Yan, L. Zhao, D. Gasevic, and R. Martinez-Maldonado. 2022. Scalability, Sustainability, and Ethicality of Multimodal Learning Analytics. In *LAK22: 12th Int. Learning Analytics and Knowledge Conf.* 13–23. https://doi.org/10.1145/3506860.3506862

[56] H. Yaribeygi, Y. Panahi, H. Sahraei, T. P. Johnston, and A. Sahebkar. 2017. The impact of stress on body function: A review. *EXCLI Journal* 16 (2017), 1057. https://doi.org/10.17179/excli2017-480