PREPRINT - Accepted at the 45th IEEE/ACM International Conference on Software Engineering (ICSE 2023) - NIER Track

Towards Supporting Emotion Awareness in Retrospective Meetings

Daniela Grassi, Filippo Lanubile, Nicole Novielli
University of Bari
{daniela.grassi, filippo.lanubile, nicole.novielli}@uniba.it

Alexander Serebrenik
Eindhoven University of Technology
a.serebrenik@tue.nl

***Abstract*—Emotion awareness is a key antecedent to team effectiveness and the use of biometrics can help software developers in gaining awareness of emotions at the individual and team level. In this paper, we propose an approach to include emotional feedback in agile retrospective meetings as a proxy to identify developers' feelings in association with the activity performed by the team. As a proof of concept, we developed an emotion visualization tool that provides an integrated visualization of self-reported emotions, activities, and biometrics. We run a pilot study to evaluate our approach with the agile retrospective meetings of a software engineering capstone project. The preliminary findings suggest that integrated emotion visualization can be useful to inform discussion and reflection around the potential causes of unhappiness, thus triggering actionable insights that could enhance team productivity and improve collaboration.**

***Index Terms*—Emotion awareness, agile teams, retrospective meetings, biometric sensors, visualization.**

I. Introduction

Findings of psychological research show how emotion self-awareness is an antecedent of team effectiveness [1]. Andriyani et al. [2] performed a case study involving developers and their finding suggest that software developers openly discussed their feelings during agile retrospective meetings [3]. Fountaine and Sharif [4] suggest that emotion awareness might help increasing developers' progress through mitigation of negative emotions. Recent work by Girardi et al. [5] suggests the use of biometrics to help developers in improving awareness of emotions at both individual and team levels.

In line with recent findings, *we envision the emergence of tools to support emotion awareness in Agile software development and trigger actionable insights to enhance team productivity and improve collaboration.*

Emotion visualization can be integrated into an agile process to understand when they experience negative emotions, such as stress, thus helping them identifying the causes for such emotional reactions [5] and enabling corrective actions. Working towards this vision, we developed EmoVizPhy, an emotion visualization tool that provides the developer with an integrated overview of own self-reported emotions, activities, and physiological signals collected through biometric sensors. The tool leverages data coming from different information sources, including self-reported emotions and biofeedback acquired through a sensor-capturing electrodermal activity (EDA). The goal is to build views of the aligned data to present the user with clear and intuitive (self-)reflection material. We evaluate the usefulness of tool-supported emotion visualization through a pilot study with Computer Science students involved in a software engineering capstone project. As a main contribution, we show how emotion visualization can be used in retrospective meetings to inform discussion and reflection around the potential causes of unhappiness during software development.

This paper represents a first step towards assessing the usefulness and feasibility of supporting the emotion awareness and well-being of software developers by leveraging non-invasive biometric sensors in combination with self-report.

II. Background and Related Work

In line with previous work on emotions of software developers [5]–[7], we ground our study on dimensional models of emotions. In particular, we include consideration of three dimensions for emotions. Two of them are originally defined by Russel [8], who operationalizes emotions in terms of *valence*, i.e. the pleasantness vs. unpleasantness of the emotion stimuli, and *arousal*, i.e. its level of activation vs. deactivation. Pleasant emotional states are associated with *positive* valence, while unpleasant ones are associated with *negative* valence. Conversely, arousal refers to the level of activation of the emotional state ranging from inactive or *low* to *high*. Furthermore, in line with previous studies [9], [10], we also measure emotions according to a third dimension known as *dominance* (or control), that is the extent to which an individual feels in control of the situation.

The link between emotions and biometrics has been investigated for a long by affective computing research. Specifically, changes in biometrics associated with the electrical activity of the brain (EEG), the electrical activity of the skin (EDA), and heart-related metrics, such as blood volume pressure (BVP), heart rate (HR) and its variability (HRV) have been successfully used for emotion detection [11]–[13]. In recent years, software engineering researchers recently investigated the emotion detection using lightweight biometric sensors that can be comfortably worn while coding [6], [7], [14]. Recognition of negative emotions received special attention [5], as these might be detrimental to developers' well-being and productivity [15]. Fritz et al. [16] rely on a combination of EEG, BVP, and eye tracker to assess difficulty in code comprehension and prevent developers from introducing bugs. In a follow-up study, they employ the same set of sensors to distinguish between positive and negative emotions while programming [7]. Along the same line, Girardi and colleagues use biometrics to classify developers' emotions during a lab study [6] as well as at the work place [5]. valence and arousal.

Among other affective states, stress was recently investigated by Westerink et al. [17]. In their study, they demonstrated a correlation between peaks of cortisol, which is the hormone associated with the experience of psychological stress, and peaks of skin conductance, thus suggesting that peaks in the EDA signal can be successfully used to identify stress episodes. In our study we consider EDA as it can be collected using low-cost non-invasive sensors [5], [7], [14], [18] that can be comfortably used by developers during programming tasks (see Section III-B). This choice is in line with current research investigating the use of lightweight biometric sensors, including EDA, for emotion recognition in software development [19], [20].

III. Supporting Emotion Awareness in Retrospective Meetings

A. The idea

This study fits in the vein of ongoing research investigating the link between developers' emotions and productivity [5], [7], [15]. We propose an approach to support self-emotion awareness of developers to help them gaining insights on the causes for the negative and positive emotions experienced during an Agile development iteration. Such insights can be shared, on a voluntary basis, with the other team members to inform and guide the discussion during the retrospective meeting usually organized at the end of the iteration.

The closest study aiming at supporting emotion awareness of agile teams is the case study by El-Migid and colleagues [21]. Grounding on previous work by Madampe et al. [22], they developed Emotimonitor, a tool to capture emotions of Agile team members with respect to technical tasks using emoji reactions on Trello cards. Their findings provide evidence that this information can be used to summarize the team emotional reaction, thus enabling emotion identification as a central part of retrospective meetings.

However, self-reported emotions might be influenced by cognitive processing as well as by emotion regulation tendencies. It is the case, for example, of emotional labor of software developers, that is the "process by which workers are expected to manage their feelings in accordance with organizationally defined rules and guidelines" [23], which might reduce the intention to disclose negative emotions considered not acceptable in collaborative software development [24]. Indeed, emotions can be seen as a coherent response among different components [25], including cognitive assessment of a situation (i.e., worrying about something threatening my goals) and the way the emotions reflect in biometrics changes (e.g, EDA changes due to sweating and heart rate rising in presence of anxiety). Findings from affective computing research suggest that multiple emotion assessment methods (e.g., self-report vs. recognition of emotions based on facial expressions) might not necessarily align at a particular moment in time [26]. In particular, previous studies report the correlation between biometrics collected using non-invasive sensors and the emergence of emotional states while programming [5]–[7], [14].

Hence, we argue that to fully support emotion awareness during software development, a combination of multiple approaches for emotion assessment is needed. Specifically, we advocate in favor of tools and practices including both self-reporting through experience sampling and visualization of biofeedback as a proxy to identify relevant emotional episodes, as they might provide complementary information on the emotional status of an individual.

B. Data collection

In line with previous work [5]–[7], we use the Empatica E4 wristband[1] to collect the EDA signal, which is recorded with a sample frequency of 4Hz. As for self-report of emotions, we replicate the approach based on experience sampling [27] used by previous studies to collect the participants' emotions and activities while programming [5]–[7]. Specifically, we use the pop-up application developed by Girardi et al. [6], which the authors made publicily available, to collect the self-reported scores for valence, arousal and dominance as well as the activity the participant is performing at the moment of the interruption. The pop-up application relies on the Self-Assessment Manikin (SAM) for self-report [28], using a 5-point pictorial scale used for assessing and collecting scores for each emotion dimension. Furthermore, we ask the participants to report the activity they are performing at the moment of the interruption by selecting it from a drop-down list, based on previous work [5], [29], i.e., *coding, bug fixing, testing, design, meeting, email, helping, networking, learning, administrative task, documentation, just arrived, other*. Finally, we ask the participants to motivate the ratings provided, i.e. to explain the causes for the reported emotions. These choices are in line with the experimental protocol adopted by Girardi et al. in their field study on developers' self-reported emotions and perceived productivity at the workplace [5].

C. Data Processing and Cleaning

Physiological signals obtained with wearable sensors are noisy and may contain erroneous data due to, e.g., electrode contact loss, or movement artifacts. As a result, the raw signals recorded during the experimental sessions need to be cleaned in order to allow for meaningful data interpretation. We remove the signal recorded before the first pop-up in order to remove anomalies due to wearing the Empatica E4. Then, we use the tool by Taylor et al. [30] to identify and remove artifacts in the signal, i.e. peaks due to noise rather than genuine changes in skin conductance values. We compute EDA peaks using the same tool, which implements a method capitalizing on the 1st derivative of the signal curve. The choice to include identification of signal peaks is informed by findings by Westerink et al. [17], demonstrating that stressful events are associated to EDA peaks.

[1] https://www.empatica.com/en-eu/research/e4/

D. Emotion Visualization: EmoVizPhy

EmoVizPhy provides a visualization of the EDA signal to identify peaks that can be used as a proxy for stress. This information is combined with self-reported valence, arousal, and dominance scores, as well as the comments provided by the user through self-report. The emotion information is aligned with time and date to inform the developer of work-related emotion triggers during the workday, so as to facilitate the recalling of such episodes during the retrospective meeting.

As an example of the visualization output, Figure 1 shows in blue the EDA signal with peaks represented as vertical lines, which are coloured differently based on the most recently self-reported arousal. If the self-reported arousal is 1 or 2, the vertical line is green; if the arousal is 3, the line is orange; and if the arousal is 4 or 5, the line is red. The tool shows the timestamp every hour (in the format "MM-DD-HH"), which is the axis on which the self-reports and the signal are aligned. To facilitate comprehension, the values of Arousal and Valence are colored green, orange, and red. In the example in figure, there are a significant number of red peaks in the centre area of the plotted signal. In correspondence to these peaks, the participant reports arousal of 4 and provides an explanation in the notes stating that "*we have reached 89% of coverage but we must reach 90%*", which overall indicates the participant was experiencing pressure to complete the task.

IV. Pilot Study

We aim at addressing the following research question: *To what extent the visualization of biofeedback and self-reported emotions enhance the effectiveness of Agile retrospective meetings?*

A. Participants

We recruited 17 undergraduate CS students involved in the capstone project of a Software Engineering course in which they have to complete 3 Scrum Sprints, working in teams of up 7 members. The experimenters explained the study protocol to all students and then used a web form to collect the volunteers' names. As a result, 4 teams out of 30 participated in the study. For each team, one participant operates in the *experimental* condition, i.e. she wears the Empatica E4 sensor for collecting the EDA signal and uses the pop-up self-report application while working at the capstone project during the final Sprint. The other team members operate in the *control* condition in which information about emotions and biofeedback is not collected. Overall, 4 students in different teams are in the experimental condition while 13 are in the control group (4 in Team 1; 2 in Team 2; 3 in Team 3; and 4 in Team 4).

B. Study protocol

Pre-experimental briefing. Before starting the study, the experimenter meets the participants and demonstrates how to correctly wear the wristband, install the pop-up application, and use the E4 manager for downloading the biometric data from the device. The raw data is then shared with the experimenter using a private channel. Next, the experimenter explains how to use the SAM scales for self-report of valence, arousal, and dominance. Then, the participant signs the informed consent form[2].

Data collection. For each participant in the experimental condition, we collect data for two weeks, which is the duration of the final Sprint (see Figure 2). The participants wear the Empatica E4 and use the pop-up application while performing all activities related to the capstone project final Sprint. By default, the popup appears on the participant's monitor once every 30 minutes. We believe that this interval is a fair trade-off between an adequate number of self-reports and the required interruptions of the participant. When the participants do not want to be interrupted, they can postpone answering the pop-up. To reduce the intrusiveness of the pop-up we allow the participants to dismiss the pop-up for the entire day. Conversely, the participants can invoke the pop-up manually, when experiencing strong emotions that they believe are important to be reported. At the end of each working session, they turn off the device, and share data with the experimenter who reviews them to check for consistency and completeness.

Retrospective meeting. Once the Sprint is over, the experimenter performs the data processing and cleaning step (see Section III-C), runs the processing on the clean EDA signal, and uses EmoVizPhy to create a comprehensive view of the emotions and activities (see Figure 1). The experimenter shares this visualization with the participant in the experimental condition only. During the retrospective meeting, the experimenter plays the role of the agile coach. We follow the *Mad Sad Glad* template [3] for the retrospective meeting as it helps to release emotions and connect them to events that happened during the Sprint. The retrospective meeting starts with the participants individually writing down personal cards, using a different frame color (see Figure 3). Afterwards, the cards are shared on a digital board and the discussion starts. For each team, only the participant in the experimental condition can use the EmoVizPhy visualization to recall significant events and thus write down their personal cards. The experimenter reads aloud the cards and team members state their agreement by up-voting the cards. Cards reporting the same issues are merged. Figure 3 shows an example of the final whiteboard produced during one of the retrospective meetings. After a retrospective meeting is over, the experimenter conducts a follow-up semi-structured interview with the experimental subjects.

V. Results and Discussion

We summarize the insights obtained with the semi-structured interviews and discuss the study limitations.

Readability: All participants rated the tool output as clear and understandable. One participant only reported minor difficulties and suggested using a more expressive label or symbol instead of the Arousal and Valence scores: *"Except for the numbers that represented valence and arousal, the final plot was easy to interpret. In fact, while entering the data in the pop-up, I focused on the SAM icons rather than the numerical*

[2]Ethical review board: ERB2022MCS11, Eindhoven Univ. of Technology

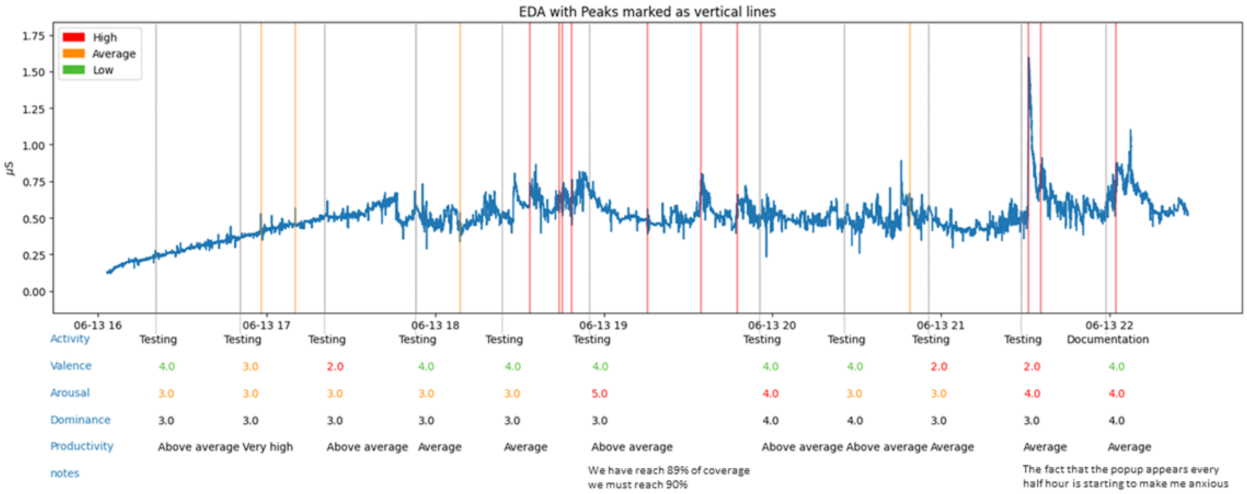

Fig. 1. Visualization of self-reported emotions, biometrics, and activities.

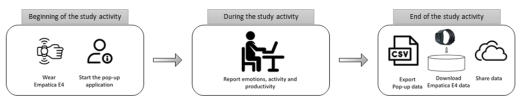

Fig. 2. Timeline of the data collection for each participant during the Sprint

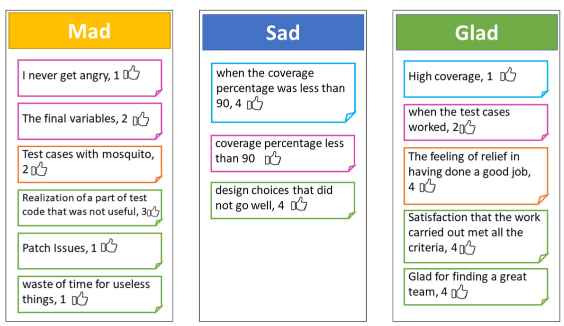

Fig. 3. An example of the Whiteboard resulting from a retrospective meeting.

values. I would alter the numerical numbers in an icon or a statement that provides an additional explanation."

Usefulness in recalling emotion episodes: Three participants found the tool useful for recalling both the negative and positive events of the Sprint. Two of them pointed out that the notes were particularly useful and one of them suggested that they should be mandatory to better support recalling the episodes corresponding to self-reported emotions and associating stress-triggering events with EDA peaks. One of the participants specifically refers to the usefulness of the tool in writing the card: *"During the previous retrospective meetings with my colleagues I always had difficulty in writing the cards. Thanks to this tool, I was able to remember emotion-triggering events more easily"*

Usefulness in supporting self-awareness: Two participants indicated that the tool gave them new insights on the causes of their emotions (e.g.: *"I did not expect that creating UML diagrams can be so stressful"*). Another participant reported that instantly expressing her emotional state through the self-report enabled her to timely recognize her feelings.

Further usage scenarios: Two of the participants claimed that they would use the tool to keep track of their stress levels during their daily studying activities in order to monitor the stress associated with exam preparation.

Reflection on the process: All the four participants agreed that wearing the wristband was not uncomfortable. As for answering the pop-up, three participants reported that it was not perceived as intrusive and only one reported feeling nervous when the pop-up appeared, but never skipped it.

Limitations: We acknowledge the preliminary nature of the findings as well as some limitations due to the reduced number of participants. Other concerns might be due to the participants being students. On the one hand, they might be not representative of professional software developers. On the other hand, the participants might feel under pressure and not disclose negative feedback. To mitigate this threat and let students freely express their feedback, the lecturer did not take part in the retrospective meetings and could not access the emotion data. We share the replication material to encourage follow-up studies [3].

VI. Conclusions and Future Plans

While preliminary, our findings show that the emotion visualization provided by the tool is perceived as useful to identify developers' feelings in association with the activity performed, thus informing agile teams during retrospective meetings. This information can also inspire developers to consider changing their behaviour to improve their reactions to negative emotions and stress. To avoid the potential misuse of technology to monitor people's behaviour, we advocate in favour of using the tool on a voluntary basis to share emotional feedback with colleagues, e.g. during retrospective meetings. In the next future, we plan to perform an extended replication of our pilot study with a broader, more diverse pool of participants, also including professional IT developers.

[3]https://figshare.com/s/9f1de1570530f654fc20

References

[1] P. Jordan and N. Ashkanasy, *Emotional Intelligence, Emotional Self-Awareness, and Team Effectiveness*. Lawrence Erlbaum Associates Publishers, 2006, pp. 145–163.

[2] Y. Andriyani, R. Hoda, and R. Amor, "Reflection in agile retrospectives," in *Agile Processes in Software Engineering and Extreme Programming*, H. Baumeister, H. Lichter, and M. Riebisch, Eds. Cham: Springer International Publishing, 2017, pp. 3–19.

[3] E. Derby, D. Larsen, and K. Schwaber, *Agile retrospectives: Making good teams great*. Pragmatic Bookshelf, 2006.

[4] A. Fountaine and B. Sharif, "Emotional awareness in software development: Theory and measurement," in *2017 IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering (SEmotion)*, 2017, pp. 28–31.

[5] D. Girardi, F. Lanubile, N. Novielli, and A. Serebrenik, "Emotions and perceived productivity of software developers at the workplace," *IEEE Transactions on Software Engineering*, pp. 1–1, 2021.

[6] D. Girardi, N. Novielli, D. Fucci, and F. Lanubile, "Recognizing Developers' Emotions while Programming," in *42nd Int. Conf. on Software Engineering (ICSE '20), July, 2020, virtual conference*, 2020, pp. 666–677.

[7] S. C. Müller and T. Fritz, "Stuck and frustrated or in flow and happy: Sensing developers' emotions and progress," in *ICSE*, 2015, pp. 688–699.

[8] J. Russell, "Culture and the categorization of emotions," *Psychological Bulletin*, vol. 110 (3), pp. 426–450, 1991.

[9] D. Graziotin, X. Wang, and P. Abrahamsson, "Are happy developers more productive? - the correlation of affective states of software developers and their self-assessed productivity," in *PROFES*, 2013, pp. 50–64.

[10] M. Mäntylä, B. Adams, G. Destefanis, D. Graziotin, and M. Ortu, "Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity?" in *Proceedings of the 13th International Conference on Mining Software Repositories, MSR 2016, Austin, TX, USA, May 14-22, 2016*, 2016, pp. 247–258.

[11] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.

[12] S. Koelstra, C. Mühl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. on Affective Comp.*, vol. 3, no. 1, pp. 18–31, 2012.

[13] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Trans. Aff. Comp.*, vol. 7, no. 1, pp. 17–28, 2016.

[14] H. Vrzakova, A. Begel, L. Mehtätalo, and R. Bednarik, "Affect recognition in code review: An in-situ biometric study of reviewer's affect," *Journal of Systems and Software*, vol. 159, p. 110434, 2020.

[15] D. Graziotin, X. Wang, and P. Abrahamsson, "Happy software developers solve problems better: psychological measurements in empirical software engineering," *PeerJ*, 2014.

[16] T. Fritz, A. Begel, S. C. Müller, S. Yigit-Elliott, and M. Züger, "Using psycho-physiological measures to assess task difficulty in software development," in *36th Int'l Conf. on Software Engineering, ICSE '14*, 2014, pp. 402–413.

[17] J. H. Westerink, R. J. Rajae-Joordens, M. Ouwerkerk, M. van Dooren, S. Jelfs, A. J. Denissen, E. Penning de Vries, and R. van Ee, "Deriving a cortisol-related stress indicator from wearable skin conductance measurements: Quantitative model & experimental validation," *Frontiers in Computer Science*, vol. 2, p. 39, 2020.

[18] D. Girardi, F. Lanubile, and N. Novielli, "Emotion detection using noninvasive low cost sensors," in *ACII 2017*, 2017, pp. 125–130.

[19] S. C. Müller and T. Fritz, "Stuck and frustrated or in flow and happy: Sensing developers' emotions and progress," in *37th IEEE/ACM International Conference on Software Engineering, ICSE 2015, Florence, Italy, May 16-24, 2015, Volume 1*. IEEE Computer Society, 2015, pp. 688–699.

[20] D. Girardi, F. Lanubile, and N. Novielli, "Emotion detection using noninvasive low cost sensors," in *Seventh International Conference on Affective Computing and Intelligent Interaction, ACII 2017, San Antonio, TX, USA, October 23-26, 2017*, 2017, pp. 125–130.

[21] M.-A. A. El-Migid, D. Cai, T. Niven, J. Vo, K. Madampe, J. Grundy, and R. Hoda, "Emotimonitor: A trello power-up to capture and monitor emotions of agile teams," *Journal of Systems and Software*, vol. 186, p. 111206, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016412122100279X

[22] K. Madampe, R. Hoda, and P. Singh, "Towards understanding emotional response to requirements changes in agile teams," in *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: New Ideas and Emerging Results*, ser. ICSE-NIER '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 37–40. [Online]. Available: https://doi.org/10.1145/3377816.3381722

[23] R. Hochschild, *The managed heart: Commercialization of human feeling*. The University of California Press, 1983.

[24] A. Serebrenik, "Emotional labor of software engineers," in *Proc.of the 16th Ed. of the BElgian-NEtherlands software eVOLution symposium*, 2017, pp. 1–6.

[25] R. Pekrun, *Emotions as Drivers of Learning and Cognitive Development*. New York, NY: Springer New York, 2011, pp. 23–39.

[26] J. M. Harley, F. Bouchet, M. S. Hussain, R. Azevedo, and R. Calvo, "A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system," *Computers in Human Behavior*, vol. 48, pp. 615–625, 2015.

[27] R. Larson and M. Csikszentmihalyi, *The Experience Sampling*. Springer Netherlands, 2014, pp. 21–34.

[28] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy & Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[29] A. Meyer, E. T. Barr, C. Bird, and T. Zimmermann, "Today was a good day: The daily life of software developers," *IEEE Trans. on Software Eng*, pp. 1–1, 2019.

[30] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, "Automatic identification of artifacts in electrodermal activity data," vol. 2015, 08 2015, pp. 1934–1937.