How Do Negative Emotions Influence on the Conceptual Models Verification? A live study proposal

Angela Mayhua-Quispe
Universidad Nacional de San Agustín de Arequipa
Peru
amayhuaq@unsa.edu.pe

Franci Suni-Lopez
Universidad Nacional de San Agustín de Arequipa
Peru
fsunilo@unsa.edu.pe

Maria Fernanda Granda
Computer Science Department of University of Cuenca
Ecuador
fernanda.granda@ucuenca.edu.ec

Nelly Condori-Fernandez
Universidade da Coruña
Vrije Universiteit Amsterdam
Coruña, Spain, Amsterdam, The Netherlands
n.condori.fernandez@udc.es
n.condori-fernandez@vu.nl

ABSTRACT

The present live study is proposed with the objective of investigating the influence of negative emotions (*i.e.,* stress) in the efficiency for verifying conceptual models. To conduct this study, we use a Model-driven Testing tool, named CoSTest, and our own version of stress detector within a competition setting. The experiment design, overview of the empirical procedure, instrumentation and potential threats are presented in the proposal.

1 INTRODUCTION

1.1 Research problem

Emotions and moods are inherent to all human experiences and consequently have effects on our work performance [15]. For example, emotions might affect the correctness and duration of activities, or the perceived experience. In the last years, some researchers have been focused on analyzing what emotions are present and their influence in the software engineering field (e.g., [2, 4, 11, 13, 18]), being the software development process one of the main topics of interest.

The influence of the emotional state on the performance of programming tasks was analyzed by Wrobel [18], who applied a questionnaire and an interview for each participant to collect data. In this study, the Job Emotions Scale (JES) was applied to measure human emotions. Romano et al. [13] analyzed emotions when novice developers apply the TDD (Test-Driven Development) approach on change tasks; similarly, they collected data through self-reported emotions by the participants.

On the other hand, researchers have not focused only on analyzing self-reported emotional states, we have also found some attempts in using physiological data to determine automatically some emotions in a specific context. For instance, Müller and Fritz [11] collected bio-metric measures (*e.g.,* electro-dermal activity, electroencephalography, skin temperature, heart rate) using different devices (Empatica E3-wristband, Neurosky MindBand and Eye Tribe) to distinguish positive and negative emotions using machine learning techniques in the context of software change tasks; also, participants assessed their emotions answering periodically a short questionnaire. Girardi et al. [4] replicated the work of Müller and Fritz in the same context, using similar devices (Empatica E4-wristband, Emotiv Insight and Tobii 4C), including more subjects in their experiment. Other work was proposed by Suni Lopez et al. [10], they were focused on detecting physiological stress in real-time in a quiet office workplace environment, using some well-known emotional triggers from the Psychology community. Authors used the E4-wristband for gathering the electrodermal activity (EDA) and applying an arousal-based statistical approach for the stress detection.

Although the amount of research for understanding the influence of emotions in software development process is rising, the impact of emotions in model-driven development (MDD) has not been yet well investigated. We selected the MDD context, because its adoption in the industry is growing quite rapidly [12]. Additionally, we have access to CoSTest, a model-driven testing tool. A conceptual model (CM) is a key asset in MDD because represents abstract concepts of the relationships between objects in a specific problem. If a CM has defects, these are passed on to the following stages (*e.g.,* coding) and could be more expensive the correction. According to Granda et al. [5], defects in conceptual models (*e.g.,* missing, wrong and unnecessary elements) can be located in several ways through Validation & Verification techniques, which can be statically or dynamically supported by a tool and can have different scope and limitation depending on its purpose (*i.e.,* detect, prevent and resolve).

For these reasons, we propose to analyze the influence of negative emotions (focusing mainly on detect stress from physiological data) on the conceptual models verification process. This verification process includes the correction of defects found in the conceptual model.

1.2 Motivation to conduct the study

In this live study proposal, we start from the hypothesis that negative emotions of workers influences on their productivity when they develop model-driven tasks. As the emotional state could be influenced by the user profile (*e.g.,* background, personality, experience in the task) or the way how the emotions are experiment for each one; we focus on investigating the emotions that could be generated along the tasks of defect detection in CMs, which are supported by a testing tool [6].

Therefore, in order to investigate if negative emotions are experienced by the subjects during the verification and correction of conceptual models, we present the design and plan of a live study to be conducted with the SEmotion's attendees.

2 SCENARIO

The model-driven development includes different tasks, where testing and correction are important to ensure the quality of the conceptual models. The scenario of this live study is based on the context of correcting defects, where requirements engineers, analysts, and testers have an important role, using a UML modeling editor for applying changes on the class diagrams. CoSTest [6] is used as support tool to detect defects and verify the correctness and completeness of these changes. Doing these tasks, subjects could experience negative emotions (*e.g.,* frustration, stress, anxiety), which might be caused by different factors such as the lack of familiarity with the tools (i.e., CoSTest, UML editor), difficulties to correct some type of defects. For this reason, subjects will use the E4-Wristband[1] for capturing physiological data (*e.g.,* EDA, heart rate, skin temperature), and a software app in a smartphone for detecting real-time physiological stress. Moreover, the experimenter will be able to monitor the emotional state from all participants (See Figure 1).

3 STUDY DESIGN

3.1 Goal and research questions

The live study proposal aims to *analyze* the influence of negative emotions on the efficiency for verifying conceptual models.
From this goal, the following research question is derived:
RQ$_1$: *How do negative emotions influence on efficiency for performing CM verification tasks?*

Type of study: We propose to conduct a *quasi-experiment*, where all participants have the same set of defects to be corrected and will use CoSTest[2] as a support tool.
Variables and metrics: the following variables were identified: **independent variables:** *CoSTest tool* that is used to automatically detect defects in conceptual models. The selected *conceptual models* and the *defects injected* into the CMs also can impact the results. As **dependent variables:** We identified the following variables: 1) *user emotional state* that is determined by (i) the stress detector proposed by Suni Lopez et al. [10] and (ii) the self-reported emotions through some questionnaires like PANAS and VAS for stress and anxiety (see Section 3.3.2 for more details); 2) *efficiency for correcting defects* that represents the relationship between the total corrected defects and the total time assigned for correcting all defects.

In this live study we focus on negative emotions due to that we are interested in: 1) validating our stress detector, and 2) creating datasets for detecting anxiety. It is important to remark that technology adoption is influenced strongly by negative emotions [14]. In the present live study, our technology is the CoSTest tool that we will use as support for verifying conceptual models.

3.2 Population of interest

3.2.1 Profile of the intended subjects. We plan for 16 subjects; students, researchers, and practitioners are welcome to this live study. Prior knowledge and experience on modeling UML class diagrams using tools or editors (*e.g.,* UML2Tools editor[3]) is required. We choose SEmotion 2020 to run this experiment thanks to the knowledge in software engineering of SEmotion's attendees, particularly UML modeling and testing.

3.2.2 Benefits to the subjects of participating in the study. The testing tasks are the key leverage point for practitioners (*e.g.,* project managers, analysts, testers) who want to develop software systems with high quality level. In this context, we think that this study might benefit to participants by getting:

- Training on a tool to support the verification of conceptual models through test cases, which could be applied on their workplaces or development tasks.
- Experience in running studies to analyze emotions based on physiological data.
- Access to instruments for evaluating and measuring emotions, such as the Positive and Negative Affect Schedule (PANAS) and Visual Analogue Scale (VAS), which could be used for the participants in similar studies based on human emotions.

3.3 Instrumentation

3.3.1 Stress Detector. In order to detect the presence of stress in the participants during the CM verification contest, we are going to use our own detector which was implemented and evaluated in a controlled experimental context [10]. The detector uses the EDA signals collected with the E4-wristband as main input. Those signals are filtered, applying a median filter. Then, it is used two algorithms to aggregate and to apply a discretization to a normalized time series which values are between 1 and 5 [10]. These values are interpreted as levels of stress variation (1: completely relaxed to 5: maximum arousal). Lastly, the approach uses a change detection algorithm based on ADaptive WINdowing (ADWIN) method [1] to assign a stress/not stress label.

The participants will be asked to install our mobile application on their smartphones, which will be connected to the E4-wristband (placed in the non-dominant hand). The mobile application send the physiological data to our server to be processed and to determine whether the user$_x$ is stressed or not. Figure 1 shows a pipeline of the stress detection process.

3.3.2 Questionnaire. We implement a web-based survey using the Qualtrics tool[4], which is composed by three set of questions regarding:

- **Demographic data** (*e.g.,* sex, age, educational degree, domain expertise)
- **Emotion state**; where we use the PANAS questionnaire and the VAS scale. The PANAS is a list of 20 adjectives used to describe different emotional states: 10 states of Positive Affect (PA) and 10 states of Negative Affect (NA). The PA scale measures activity and pleasure, while the NA scale relates

[1]https://www.empatica.com/en-eu/research/e4/
[2]https://costestproject2017.wordpress.com/
[3]https://www.eclipse.org/modeling/mdt/?project=uml2tools
[4]https://www.qualtrics.com

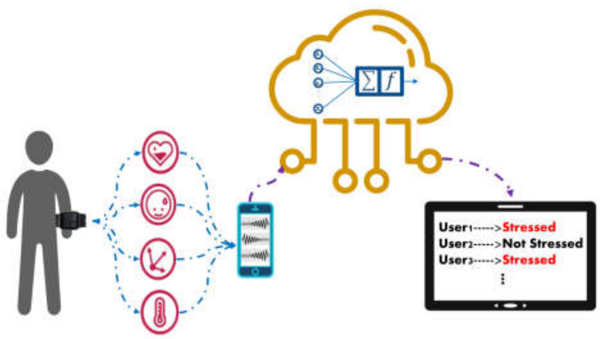

Figure 1: Pipeline of stress detection process.

to fear and stress [3]. Although we are mainly interested in analyzing the negative emotions, we are going also to collect positive affect values for further analysis. The Visual Analogue Scale (VAS) can be used as a measurement instrument for assessing anxiety level [16]; this instrument uses commonly a horizontal line to represent a range of values, from the minimum to the maximum value, so that subject marks a point on the range where he perceives his anxiety state has been located during the different situations presented on the contest.

- **Experiment feedback**, a post questionnaire that includes closed and open questions about the instrumentation, the timing allocated for each phase, and complexity of the verification task.

3.3.3 Verification tasks. In this live study, we ask the participants to carry out two verification tasks (*i.e.,* two conceptual models, CM1 and CM2) and run six test cases associated with each one by using the CoSTest tool. Based on the defect classification proposed by Granda et al. [5], we will use three defect types (*i.e.,* missing, wrong and unnecessary elements) to inject six defects per CM (two for each defect type).

Participants will have access to the six test cases corresponding to the current delivered CM; after analyzing the test cases at the same time using the CoSTest tool, it will show a list of defects which will have to be corrected in any order depending on the participant. Once one or several defects are corrected, participants will have to rerun the test cases to verify the correctness of the changes. The participants will be able to perform the next conceptual model when finishing the current one. The total time allocated to perform the two verification tasks is 40 minutes. For the verification of both CMs, the participants need to bring their laptops with VirtualBox[5] installed to execute CoSTest, because we will provide a virtual machine with all the software required in this study to avoid compatibility issues related to some programs (*e.g.,* Java).

The experimental objects consist of the specifications of two conceptual models: CM1 is a Super Stationery (SS) system, where CM defines the information system of a company that provides stationery and office material to its clients; and CM2 is a Photography Agency (PA) system where a CM defines the information system that manages photographers and their photographic reports for distribution to newspaper publishers. Regardless of the experimental object, we provide the participants with the following experimental material: (i) a brief description for each information system modeled in CM1 and CM2; and (ii) an example test suite (each one with six test cases), developed by the authors to verify each CM using CoSTest. We opted for SS and PA as experimental objects because they are often adopted to learn/practice CoSTest and were used in past empirical studies on Mutation Testing [8].

3.3.4 Other material.

- Consent form[6] that outlines the informed consent of an individual for the live study, where the privacy and confidentiality terms are detailed;
- CoSTest training material, where we provide the required material to use the tool (demo-video, examples, instructions);
- A relaxing video to be used before starting the correction contest.

3.4 Procedure

The study is composed of three phases, as shown in Figure 2.

- *Preparation*: first we explain details about the study and request to read and sign the informed consent form. Then, we provide some specific instructions to use the E4 wristband and the mobile phone devices.
  Moreover, we give instructions to configure the virtual machine, which contains the CoSTest, in the participants' laptops. Furthermore, as the CoSTest could be a new tool for the participants, we will give a training for about 30 minutes.
  As next step, we need to uniform the emotional state of all participants (*e.g.,* someone could come to the experiment already stressed) to avoid the influence of previous emotions in our experiment. To do this, participants are asked to stay quiet and watch a video during five minutes to get relaxed.
- *Correction contest*: This phase takes 40 min, and it is organized as follows: (i) Participants start identifying defects in the CM without the Tool support. (ii) We provide different test cases for using the CoSTest tool (see Subsection 3.3.3 for more details regarding the test cases used in the experiment). CMs are delivered in random order. (iii) Participants identify defects with the Tool support. (iv) Participants start correcting the list of defects reported by the tool. 5 minutes before the finalization of this phase, participants are warned to upload their corresponding solutions on time. It is important to remark that participants will be able to submit their solutions as long as they consider.
  After the contest, participants will be asked to complete a brief demographic questionnaire, and self-response emotional questionnaires to report their perceived emotions during the contest (see Section 3.3.2 for more details about these questionnaires).
- *Post-experiment*: With the purpose of getting feedback for improving the experiment, participants are requested to complete a post questionnaire.

[5] https://www.virtualbox.org/

[6] This consent form can be found in https://www.dropbox.com/s/mpz1px9h18d7taw/SEmotion-ConsentForm.pdf and its web-based version will be implemented using the Qualtrics tool.

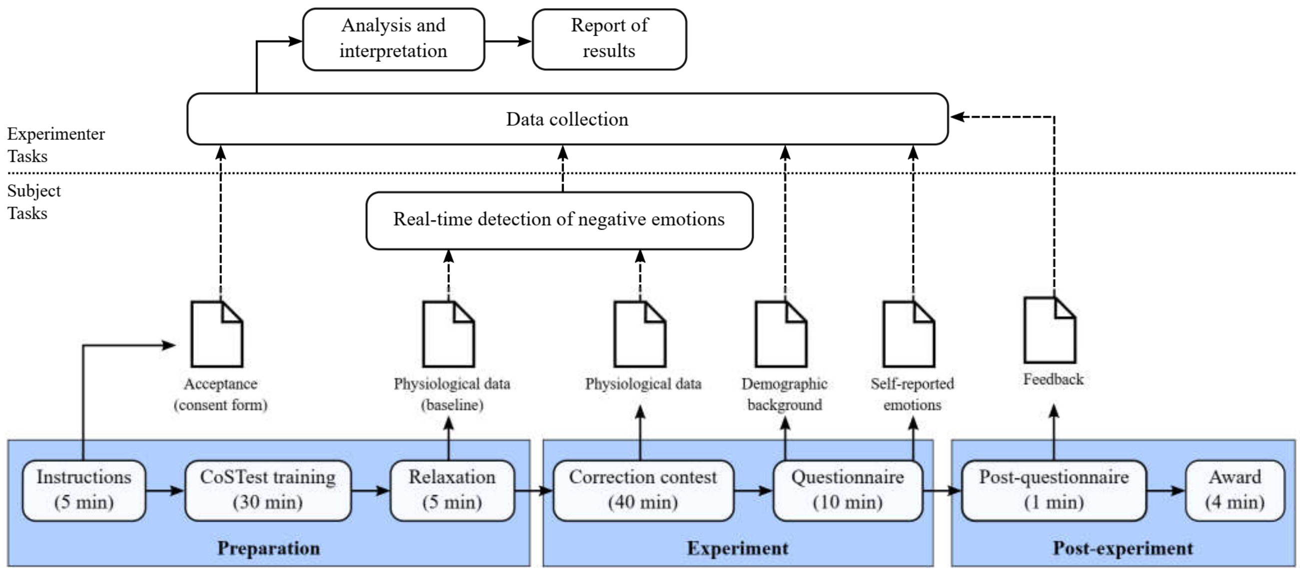

Figure 2: Live study's procedure

After processing all submissions of the participants, we will reward three participants who detect and correct more defects in the less time.

4 THREATS TO VALIDITY AND ETHICAL ISSUES

Internal validity. The different factors triggered by live study (*e.g.,* place, settings) might affect the observed variables. We mitigate this threat by performing the study in similar conditions for each participant (*e.g.,* material, verification tasks, rules of contest). The tools (*i.e.,* CoSTest and UML2Tools Editor) used in this experiment will be tested to adjust the settings on the virtual machine used for software installation. For this purpose, we will use five reference points, taken from other CMs, which are not part of this study. Additionally, our study could be negatively affected if both artifacts (CMs and injected defects) were not properly selected. The CMs have been taken from other experiments that have been carried out with the tool [7] and the defects are those generated by the mutation tool [8]. Another threat is the emotions of the participants before starting this live study (due to some previous activities or experiences on the day) might affect the perception of their emotional state during this study; to mitigate this threat, we have prepared a relaxing activity to uniform the emotions of all participants before starting this experiment.

External validity. This validity is regarding the generalization of our results outside the experiment setting; in this context, a possible threat could be the selection of participants. Nonetheless, it is mitigated by the SEmotion's attendees, because they have different personalities, experiences and educational backgrounds, such as master/PhD students, senior researchers, and practitioners from the Software Engineering community. Another possible threat is the low number of participants, which is limited by the number of available devices. However, we plan to replicate this experiment in further studies. Additionally, the environment where this study will be carried out might not be a typical context for doing software work; to mitigate this threat, we will conduct the study in a quiet environment, trying to give comfortable work-spaces for the participants as well as putting them to work under time pressure.

Construct validity. As part of this study, our instruments are based on questionnaires with self-reported responses and as consequence, participants could be afraid of giving information about their emotional states or personal information; however, this threat is mitigated through our privacy and confidentiality terms that specify their information and responses are going to be anonymous. Furthermore, the selected instruments are well known and have been used in other works to measure emotions [9, 17]. Some other possible threat is determining the correctness of a solution for a defect because it could affect the measure of efficiency; this threat is mitigated by the CoSTest tool, which verifies automatically if the defect was solved successfully or not.

Ethical issues. The live study will be implemented in a survey platform (*i.e.,* Qualtrics), where at the beginning of the study, we present to participants information about the experiment and privacy statement, and they will give their consent to participate in the study. The participation in the study will be anonymous and volunteer.

5 PUBLICITY AND DISSEMINATION PLAN

To make publicity of our study for attracting potential participants we plan to use the social networks of SEmotion 2020 (*e.g.,* Twitter, Facebook). A summary of our preliminary results with attendees will be disseminated in a short presentation on the last day of the workshop. The final results, discussion and conclusions will be published as a research paper and submitted to one of the appropriate venues either a conference (*e.g.,* ER, ESEM, MODELS, REFSQ, CAISE, CHI) or journal (*e.g.,* Journal of Systems and Software).

6 ABOUT THE RESEARCHERS

Angela Mayhua Quispe is a PhD candidate at National University of San Agustin of Arequipa. Her main research focuses on data visualization and pattern recognition from images and signals. She has a particular interest in analyzing physiological signals to recognize human emotions, with special emphasis on negative emotions.

Franci Suni Lopez is a PhD candidate at National University of San Agustin of Arequipa. His main research focuses on using human emotions to empower the self-adaptation capability of software services. His research interest also includes software engineering for mobile development and self-adaptive software. He executed one empirical studies in the live study track of REFSQ'19 and an experiment at MeGSuS 2018.

Maria Fernanda Granda is a professor and researcher of the Computer Science Department at University of Cuenca (Ecuador). She obtained her Ph.D. in Computing in 2017 at the Universitat Politècnica de València (Spain). Her main research focuses on Requirements Engineering, Software Testing & Quality and Model-driven Development. She has executed empirical studies using mutation testing and the results have been published in several conferences (i.e., EASE, CAISE, ISD)

Nelly Condori-Fernandez is an assistant professor at the Universidade da Coruña (Spain) and research associate of the Vrije Universiteit Amsterdam (the Netherlands). Her main empirically-driven research focuses on topics related to quality requirements prioritization and requirements validation. She has a particular interest in applying Human Computer Interaction technologies to support software engineering activities. Her research interests also include sustainability design and assessment with special emphasis on social and technical aspects of context-aware systems. She has executed various type of empirical studies and published in conferences like REFSQ, ESEM, EASE and journals as JSS and IST.

ACKNOWLEDGMENTS

This work has been supported by the projects: KUSISQA (014-2019-FONDECYT-BM-INC.INV) from the National Fund for Scientific and Technological Development (FONDECYT-PERU) and World Bank, Datos 4.0 (TIN2016-78011-C4-1-R) supported by MINECO-AEI/FEDER-UE, and the Dirección de Investigación de la Universidad de Cuenca (DIUC) - Ecuador.

REFERENCES

[1] Albert Bifet and Ricard Gavaldà. 2007. Learning from Time-Changing Data with Adaptive Windowing. In *Proceedings of the 2007 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 443–448.

[2] Broderick Crawford, Ricardo Soto, Claudio León de la Barra, Kathleen Crawford, and Eduardo Olguín. 2014. The Influence of Emotions on Productivity in Software Engineering. In *HCI International 2014 - Posters' Extended Abstracts*, Constantine Stephanidis (Ed.). Springer International Publishing, Cham, 307–310.

[3] Ute Engelen, Steven De Peuter, An Victoir, Ilse Van Diest, and Omer Van den Bergh. 2006. Verdere validering van de Positive and Negative Affect Schedule (PANAS) en vergelijking van twee Nederlandstalige versies. *gedrag en gezondheid* 34, 2 (01 Apr 2006), 61–70. https://doi.org/10.1007/BF03087979

[4] Daniela Girardi, Filippo Lanubile, Nicole Novielli, and Davide Fucci. 2018. Sensing Developers' Emotions: The Design of a Replicated Experiment. In *Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering* (Gothenburg, Sweden) *(SEmotion '18)*. Association for Computing Machinery, New York, NY, USA, 51–54. https://doi.org/10.1145/3194932.3194940

[5] Maria F. Granda, Nelly Condori-Fernández, Tanja E. J. Vos, and Oscar Pastor. 2015. What do we know about the defect types detected in conceptual models?. In *2015 IEEE 9th IEEE International Conference on Research Challenges in Information Science (RCIS)*. 88–99. https://doi.org/10.1109/RCIS.2015.7128867

[6] Maria. F. Granda, Nelly Condori-Fernández, Tanja E. J. Vos, and Oscar Pastor. 2017. CoSTest: A Tool for Validation of Requirements at Model Level. In *2017 IEEE 25th International Requirements Engineering Conference (RE)*. 464–467. https://doi.org/10.1109/RE.2017.69

[7] Maria F. Granda, Nelly Condori-Fernández, Tanja E. J. Vos, and Oscar Pastor. 2017. Effectiveness Assessment of an Early Testing Technique using Model-Level Mutants. In *2017 21st International Conference on Evaluation and Assessment in Software Engineering (EASE)*. 98–107. https://doi.org/10.1145/3084226.3084257

[8] Maria F. Granda, Nelly Condori-Fernández, Tanja E. J. Vos, and Oscar Pastor. 2017. A model-level mutation tool to support the assessment of the test case quality. In *2017 Lecture Notes in Information Systems and Organisation*, Vol. 11. 17–37. https://doi.org/10.1007/978-3-319-52593-8_2

[9] Barbara A. Karanian, Andrew Parlier, Ville Taajamaa, and Gloria Monaghan. 2018. Engineering Emotion : Students tell Stories about the Costs of Being Innovative. In *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE. https://doi.org/10.1109/fie.2018.8659349

[10] Franci Suni Lopez, Nelly Condori-Fernandez, and Alejandro Catala. 2019. Towards Real-Time Automatic Stress Detection for Office Workplaces. In *Information Management and Big Data*. Springer International Publishing, 273–288. https://doi.org/10.1007/978-3-030-11680-4_27

[11] Sebastian C. Müller and Thomas Fritz. 2015. Stuck and Frustrated or in Flow and Happy: Sensing Developers' Emotions and Progress. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, Vol. 1. IEEE, Florence, Italy, 688–699. https://doi.org/10.1109/ICSE.2015.334

[12] Önder Babur, Loek Cleophas, Mark van den Brand, Bedir Tekinerdogan, and Mehmet Aksit. 2018. Models, More Models, and Then a Lot More. In *Software Technologies: Applications and Foundations (STAF 2017). Lecture Notes in Computer Science*, Vol. 10748. https://doi.org/10.1007/978-3-319-74730-9_10

[13] Simone Romano, Davide Fucci, Maria Teresa Baldassarre, Danilo Caivano, and Giuseppe Scanniello. 2019. An Empirical Assessment on Affective Reactions of Novice Developers when Applying Test-Driven Development. https://arxiv.org/abs/1907.12290. arXiv:1907.12290 [cs.SE]

[14] Evan T. Straub. 2009. Understanding Technology Adoption: Theory and Future Directions for Informal Learning. *Review of Educational Research* 79, 2 (June 2009), 625–649. https://doi.org/10.3102/0034654308325896

[15] Howard Weiss and Russell Cropanzano. 1996. Affective Events Theory: A Theoretical Discussion of The Structure, Cause and Consequences of Affective Experiences at Work. *Research in Organizational Behavior* (01 1996).

[16] Valerie SL Williams, Robert J Morlock, and Douglas Feltner. 2010. Psychometric evaluation of a visual analog scale for the assessment of anxiety. *Health and Quality of Life Outcomes* 8, 1 (2010), 57. https://doi.org/10.1186/1477-7525-8-57

[17] Monika Wróbel, Maria Finogenow, Paulina Szymańska, and Jeff Laurent. 2019. Measuring Positive and Negative Affect in a School-Based Sample: A Polish Version of the PANAS-C. *Journal of Psychopathology and Behavioral Assessment* 41, 4 (Feb. 2019), 598–611. https://doi.org/10.1007/s10862-019-09720-7

[18] Michal R. Wrobel. 2013. Emotions in the software development process. In *2013 6th International Conference on Human System Interactions (HSI)*. 518–523. https://doi.org/10.1109/HSI.2013.6577875