[redacted]

Wearable Sensing and Quantified-self to explain Learning Experience

Kshitij Sharma[1], Ilias Pappas[1,2], Sofia Papavlasopoulou[1], Michail Giannakos[1]
1. Norwegian University of Science and Technology; 2. University of Agder
kshitij.sharma@ntnu.no, ilias.pappas@uia.no, spapav@ntnu.no, michailg@ntnu.no

***Abstract*—The confluence of wearable technologies for sensing learners and the quantified-self provides a unique opportunity to understand learners' experience in diverse learning contexts. We use data from learners using Empatica Wristbands and self-reported questionnaire. We compute stress, arousal, engagement and emotional regulation from physiological data; and perceived performance from the self-reported data. We use Fuzzy Set Qualitative Comparative Analysis (fsQCA) to find relations between the physiological measurements and the perceived learning performance. The results show how the presence or absence of arousal, engagement, emotional regulation, and stress, as well as their combinations, can be sufficient to explain high perceived learning performance**

***Index Terms*—fsQCA, wearable sensors, learner performance, multimodal learning analytics, collaborative learning**

I. Introduction and Background

Positive learning experience provides the learners an opportunity to achieve their learning goals [1]. Assessing learner experience in different contexts and in real-time might help us in designing tools to improve the landscape of teaching and learning [2]. Using questionnaires in real-time during learning sessions is rather challenging and would disrupt learners' experience [2]. One way to conduct real-time sensing, while the learners are engaged with their tasks during a learning session, is to use wearable sensors [3]. Several studies have explored the relationship between the physiological data, captured from the wearable devices, and emotions [4], collaboration quality [5] and learning experience [2]. Most of the results are based on either multiple sources data in addition to the wristband data. While multimodal data may offer rich insight, they hinder the pervasiveness of the data collection and raise certain invasive issues (eye-tracking glasses or EEG) or privacy issues (audio or video recordings). On the other hand, a few studies are also based on tedious and error-prone process of human labelling of the collaborative episodes [3], [5]. To overcome the aforementioned issues, we propose to use only the wristband data (to incorporate the sensing in a pervasive manner) and employ completely automatic yet explainable methods to explain the relationships between the physiological data (wristband data in our case) and the learning experiences. In this way, we can derive implications that can directly guide the design of scaffolding agents to support the learning and/or the teaching processes [6]. Physiological sensing have been proven to be generalizable across contexts [7], acceptable by students [8], and not disruptive of task performance [9]. Several efforts have shown the potential of wearable devices, in a multitude of learning contexts, to explain predict a multitude of learning-related constructs, such as collaboration quality [5]; and engagement, workload, motivational level and emotional state [10]. Despite promising findings from recent research studies, there is currently limited understanding of the ways in which quantified-self technologies can offer new insights into students' learning qualities [11].

In this contribution, we propose two shifts. First, we move from the invasive devices such as eye-tracking, EKG, EEG to solely pervasive wristbands. Second, we use a complementary analytical approach, namely fuzzy-sets Qualitative Comparative Analysis (fsQCA), to find the necessary and sufficient conditions from the metrics derived from sensor that capture learning experiences, and explain students' perceived performance. It has received increased attention in the field of learning technologies [12], [13]. We draw from complexity theory and configuration theory and employ fsQCA to identify multiple solutions of conditions that explain high student performance [14]. The two theories build on the notion of equifinality, which means that a specific outcome can be explained by different, equally effective set (combination) of conditions. Also, configuration theory builds on the principle of causal asymmetry, based on which the presence or the absence of a condition from the output, depends on how this condition combines with the other conditions that are examined, in order to explain the desired outcome. To this end, in this paper, we address the following research question: ***How do measures derived from wearable sensors combine to explain students' learning performance?***.

II. Methodology

We captured data from 31 university students. The students were divided in groups of 5–6 each. The context, setup and the research design of the present study is from the our previous contribution [2]. In this section, we will present the measures and the analysis. We used five measurements in this contribution. **Stress** is computed as heart rate's increasing slope. Large positive slope of the heart rate indicate higher stress [15]. **Emotional regulation (ER)** is directly computed from HRV. ER is computed as the rate of arrival of HRV peaks as suggested by [16]. Lower arrival rate of HRV peaks shows higher ER. **Arousal** is related to phasic component of the EDA signal. Phasic component is with rapid changes and is found to be related to physiological arousal [4]. **Engagement** is computed as a linear combination of EDA's increasing slope

[redacted]

[redacted]

TABLE I

CALIBRATION TABLE.

Variable [min-max]	0.95–FM	0.50–CP	0.05–FNM
Stress [0-1]	0.83	0.34	0.02
Arousal [0-1]	0.76	0.33	0.03
ER [0-1]	0.82	0.41	0.05
Engagement [0-1]	0.90	0.39	0.03
PP [0-7]	6.07	4.20	3.35

and the arrival rate of EDA peaks. Large positive slope of EDA and high rate of arrival of peaks show high engagement [17]. **Perceived performance (PP)** is the learners' rating of their performance using a self-repoorted questionnaire [18].

To address its objective this study employs fsQCA [19] following recommendations from extant research [12]–[14]. FsQCA enables capturing conditions that are (1) sufficient or necessary to explain the outcome and (2) insufficient on their own but are necessary parts of solutions that can explain the result. Furthermore, fsQCA is a configurational approach, which means that the findings include multiple configurations, or combinations, of conditions that explain the same outcome. Also, these combinations include conditions that are not identified by variance-based analyses because they may represent only a relatively small number of cases, since they compute the single-best solution or model that fits the data in the best manner. FsQCA findings offer 3 types of conditions; present, absent, and on a "do not care" situation. The "do not care" situation indicates that the variable may either be present or absent and it does not play a role on a specific configuration.

a) Calibration: In fsQCA all values need to be transformed into fuzzy sets ranging from 0 to 1 [19]. By performing the calibration, the researcher defines the fuzzy-sets and creates the conditions that will be used for the analysis. This is the most important part in fsQCA.To perform the calibration three thresholds need to be chosen, which define the **full membership (FM)** in the set, **the full non-membership (FNM)** in the set, and **the cross over point (CP)**. The following values are recommended 0.95, 0.50, 0.05 as the three thresholds (or breakpoints), which will transform the data into the log-odds metric with all values being between 0 and 1 [19]. To find which values in our dataset correspond to the 0.95, 0.50, and 0.05 we use percentiles. Thus, the thresholds are set as shown in Table I.

b) Analysis of Necessity: First, we test if any of individual conditions, both their presence and absence is necessary to explain high perceived performance. In detail, for high perceived performance the consistency values range between 0.47–0.82, for both the presence and negation of the causal conditions. Because none of causal conditions exceeds the value of 0.9 they are not considered as necessary for high perceived performance. Next, we proceed with fuzzy set analysis to identify sufficient combinations of causal conditions that explain high perceived performance.

c) Analysis of Sufficiency:: FsQCA produces a truth table of 2^k rows, on which k represents the number of outcome predictors and each row represents every possible combination. The process is explained by [14]. The truth table is sorted based on frequency (i.e. number of observations for each possible combination) and consistency (i.e., the degree to which cases correspond to the set-theoretic relationships expressed in a solution). A frequency threshold is set to ensure that a minimum number of empirical observations is acquired for the analysis. For samples larger than 150 cases the threshold should be set at 3. Next, the threshold for raw consistency is set at .85, higher than the recommended threshold of 0.75. Also, the threshold for PRI consistency is set at 0.75, over the minimum threshold of 0.5. Observations above the consistency threshold are the ones that fully explain the outcome.

TABLE II

CONFIGURATIONS FOR ACHIEVING HIGH PERCEIVED PERFORMANCE.

	S1	S2	S3	S4
Stress		⊗	⊗	
Arousal	●			
Emotional Regulation		●		●
Engagement			●	●
Raw Coverage	0.70	0.49	0.45	0.47
Unique Coverage	0.14	0.04	0.03	0.05
Consistency	0.91	0.90	0.89	0.94
Overall Solution Coverage	0.87			
Overall Solution Consistency	0.86			

Note: Black circles indicate the presence of a condition, and circles with "x" indicate its absence. All conditions are core conditions. Blank spaces indicate "don't care" conditions.

III. RESULTS AND DISCUSSION

The findings present combinations of causal conditions that are sufficient in explaining high perceived performance (Table II). The solution presents the core conditions, as the intermediate and parsimonious solutions are the same Table II presents consistency values for each combination and for the overall solutions, with all values being above the recommended threshold (>0.75). The overall solution coverage shows the extent that high perceived performance can be determined based on the identified configurations and is comparable to the R-square value. The overall solution coverage of 0.87 suggests that the solutions account for a substantial proportion of high perceived performance. Findings show that 4 configurations can explain high perceived performance. In detail:

S1: Students that experience high arousal will also have high perceived performance, regardless of the other conditions.

S2: Students that show high emotional regulation will have high perceived performance, when their stress is at low levels, regardless of their engagement and arousal.

S3: Students that show high engagement will have high perceived performance, when their stress is at low levels, regardless of their emotional regulation and arousal.

S4: Students that combine high emotional regulation and high engagement will have high perceived performance, regardless of their stress and arousal levels.

Overall, the findings highlight the importance of arousal, which is identified as a sufficient condition to achieve high perceived performance. Among the other conditions, none is sufficient on its own, but they are necessary for their respective configurations (S2-S4), which in turn are sufficient to achieve the same outcome. Thus, when the students are not stressed, high emotional regulation or engagement are sufficient to achieve high performance. Finally, in the case that students show high levels on both emotional regulation and engagement, then stress and arousal do not influence their performance. These observations, indicate a complementarity between emotional regulation and engagement.

The first configuration is solely based on high arousal (**S1**), which is also the only configuration with a single measurement. The underlying reason for arousal being positively associated with the performance could be the learners' familiarity with the situations and the learning context [20], which in turn increases the performance [21]. Therefore, to support students in conditions with low levels of arousal, the scaffolds can attempt to make students comfortable with learning contexts by providing more information about the design learning context, and roles and responsibilities of different actors.

The next three configurations complement each other with high emotional regulation, high engagement and low stress explaining the high perceived performance in a pairwise manner in the three configurations respectively. High engagement often indicates high quality interaction within the learning context and there fore often leads to high performance [3]. Adapting to the learners capability is one of the most sought for research outcome in the related fields such as, Learning Analytics, User Modelling, and Artificial Intelligence in Education. Most of the occasions when learners experience lack of engagement is due to the learning material being a lot easier than their capabilities [22]. Delivering the most appropriate content to the learners can retain their engagement [13]. Considering the emotional regulation being positively associated with performance, one can explain this relationship based on learners having control over challenging situations [23], [24]. To scaffold for emotional regulation it is important for the learners to be able to understand their own emotions and the also understand the value of the feedback [24].

Finally, we observed in two configurations that students with low levels of stress will be able to achieve high performance, indicating a negative association between the two. The relationship between the stress and learners' performance have been found in related studies with wearable sensor-based data [3], [6]. High stressful situations can be avoided with prompting and supporting the learners to understand the challenges that they are facing and hinting towards proper/correct solutions [3]. Another way to scaffold stressful situations is to provide more time for solving the problem at hand or to resolve the conflict during collaborative learning [6].

REFERENCES

[1] M. Schmidt, A. A. Tawfik, I. Jahnke, and Y. Earnshaw, "Learner and user experience research," 2020.

[2] M. N. Giannakos, K. Sharma, S. Papavlasopoulou, I. O. Pappas, and V. Kostakos, "Fitbit for learning: Towards capturing the learning experience using wearable sensing," *Intl. Jour. of Human-Computer Studies*, vol. 136, 2020.

[3] S. Lee-Cultura, K. Sharma, G. Cosentino, S. Papavlasopoulou, and M. Giannakos, "Children's play and problem solving in motion-based educational games: synergies between human annotations and multimodal data," in *Interaction Design and Children*, 2021.

[4] E. Di Lascio, S. Gashi, and S. Santini, "Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, 2018.

[5] G. Chanel, M. Bétrancourt, T. Pun, D. Cereghetti, and G. Molinari, "Assessment of computer-supported collaborative processes using interpersonal physiological and eye-movement coupling," in *2013 Humaine Association Conf. on Affective Computing and Intelligent Interaction*. IEEE, 2013.

[6] K. Sharma, S. Lee-Cultura, and M. Giannakos, "Keep calm and do not carry-forward: Toward sensor-data driven ai agent to enhance human learning," *Frontiers in Artificial Intelligence*, vol. 4, 2022.

[7] K. Sharma, E. Niforatos, M. Giannakos, and V. Kostakos, "Assessing cognitive performance using physiological and facial features: Generalizing across contexts," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, 2020.

[8] M. Sung, J. Gips, N. Eagle, A. Madan, R. Caneel, R. DeVaul, J. Bonsen, and A. Pentland, "Mobile-it education (mit. edu): m-learning applications for classroom settings," *Jour. of Computer Assisted Learning*, vol. 21, no. 3, 2005.

[9] M. Feidakis, T. Daradoumis, and S. Caballé, "Emotion measurement in intelligent tutoring systems: what, when and how to measure," in *2011 Third Intl. Conf. on Intelligent Networking and Collaborative Systems*. IEEE, 2011.

[10] R. A. Sottilare, K. W. Brawner, B. S. Goldberg, and H. K. Holden, "The generalized intelligent framework for tutoring (gift)," *Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate (ARL-HRED)*, 2012.

[11] C. Wang and P. Cesar, "Physiological measurement on students' engagement in a distributed learning environment." *PhyCS*, vol. 10, 2015.

[12] Z. Papamitsiou, A. A. Economides, I. O. Pappas, and M. N. Giannakos, "Explaining learning performance using response-time, self-regulation and satisfaction from content: an fsqca approach," in *Proceedings of the 8th Intl. Conf. on learning analytics and knowledge*, 2018.

[13] Z. Papamitsiou, I. O. Pappas, K. Sharma, and M. N. Giannakos, "Utilizing multimodal data through fsqca to explain engagement in adaptive learning," *IEEE Transactions on Learning Technologies*, vol. 13, no. 4, 2020.

[14] I. O. Pappas and A. G. Woodside, "Fuzzy-set qualitative comparative analysis (fsqca): Guidelines for research practice in information systems and marketing," *Intl. Jour. of Information Management*, vol. 58, 2021.

[15] J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel, "Influence of mental stress on heart rate and heart rate variability," in *4th European conf. of the international federation for medical and biological engineering*. Springer, 2009.

[16] D. P. Williams, C. Cash, C. Rankin, A. Bernardi, J. Koenig, and J. F. Thayer, "Resting heart rate variability predicts self-reported difficulties in emotion regulation: a focus on different facets of emotion regulation," *Frontiers in psychology*, vol. 6, 2015.

[17] D. Leiner, A. Fahr, and H. Früh, "Eda positive change: A simple algorithm for electrodermal activity to measure general audience arousal during media exposure," *Communication Methods and Measures*, vol. 6, no. 4, 2012.

[18] B. Kuvaas, "Work performance, affective commitment, and work motivation: The roles of pay administration and pay level," *Jour. of Organizational Behavior*, vol. 27, no. 3, 2006.

[19] C. C. Ragin, "Redesigning social inquiry: Fuzzy sets and beyond," *Social Forces*, vol. 88, no. 4, 2009.

[20] J. P. Jamieson, W. B. Mendes, E. Blackstock, and T. Schmader, "Turning the knots in your stomach into bows: Reappraising arousal improves performance on the gre," *Jour. of experimental social psychology*, vol. 46, no. 1, 2010.

[21] L. F. Barrett, B. Mesquita, K. N. Ochsner, and J. J. Gross, "The experience of emotion," *Annu. Rev. Psychol.*, vol. 58, 2007.

[22] R. A. Sottilare, K. W. Brawner, B. S. Goldberg, and H. K. Holden, "The generalized intelligent framework for tutoring (gift)," in *Fundamental issues in defense training and simulation*.

[23] H. Järvenoja, P. Näykki, and T. Törmänen, "Emotional regulation in collaborative learning: when do higher education students activate group level regulation in the face of challenges?" *Studies in Higher Education*, vol. 44, no. 10, 2019.

[24] E. Molloy, C. Noble, and R. Ajjawi, "Attending to emotion in feedback," in *The impact of feedback in higher education*. Springer, 2019.