Exploring relationships between electrodermal activity, skin temperature, and performance during engineering exams

Tarique Hasan Khan, Ph.D.
Engineering Education
Utah State University
Logan, UT, USA
tariquekhan@aggiemail.usu.edu

Idalis Villanueva, Ph.D.*
*corresponding author
Engineering Education
Utah State University
Logan, UT, USA
idalis.villanueva@usu.edu

Paul Vicioso, Ph.D. Candidate
Electrical & Computer Engineering
Utah State University
Logan, UT, USA
paulvicioso@gmail.com

Jenefer Husman, Ph.D.
Educational Psychology
University of Oregon
Eugene, OR, USA
jhusman@uoregon.edu

***Abstract*Students academic learning, performance, and motivation are ongoing topics in engineering education. Those studies that have attempted to understand the mechanisms of motivation in authentic classroom settings and scenarios are few and limited to the methods used (e.g., self-reports, observations). This Work-in-Progress study explores the utility of electrodermal activity (EDA) and temperature sensors in accurately informing scholars about student performance during an exam in real-time. Correlations between each factor were analyzed. Initial results suggest that peripheral skin temperature has a weak, positive but significant correlation to exam question difficulty r=0.08; p<0.001). Also, electrodermal activity and temperature showed a weak, positive but significant correlation (r=0.13; p<0.05). Electrodermal activity showed a weak, positive but significant correlation to exam question difficulty (r=0.16; p<0.01). Also, skin temperature correlations with difficulty index (did not) changed across semesters (r=0.18; p<0.001). We also developed a multiple regression model and found moderately significant relationships between EDA, difficulty index, and skin temperature (r=0.45; p<0.05). The findings suggest that performance is tied to physiological responses among students during exam taking, indicating a possible connection between emotions and cognition via physiology.**

Keywords*electrodermal activity, temperature, performance, difficulty, exams, engineering

I. Introduction

It is well known among educational researchers that emotion is inextricably linked to cognition [1]. However, the study of such connections is limited to its available methodologies [2,3] and tools to explore student performance in authentic and ecologically valid classroom settings [2,3]. For example, self-reports, interviews, and observations are used largely in studies of academic emotions [2,3]. Instruments like self-reports are limited due to high subjectivity, interpretability issues, and variability in replicability, among others [2,3]. To overcome these limitations, researchers have begun to use multi-modal approaches based on biology [4,5] and physiology [6-8].

In particular, for physiology, researchers are interested in the activation of the sympathetic nervous system (SNS), and in particular changes in skin conductivity through electrodermal activity (EDA) whose increased activation is connected to the cognitive load and/or strong emotional responses [6-8]. Biological and physiological researchers have suggested that EDA data is influenced by physical mobility [6,7], speaking [9], and bodily and ambient temperatures [10-12]. However, the role of these influencing factors in classroom environments, and more specifically in for educational research, is less known [13-15]. The purpose of this work-in-progress study is to explore how influencing factors in physiological data collection (i.e., electrodermal activity, skin temperature) correlate to student performance. For this work, we selected engineering exams as an ecologically valid scenario due to the difficulty of its exams and real-world applications [16]. Findings from this work will help educational and engineering educational researchers understand how to analyze and interpret physiological data in the context of examination experiences.

II. Literature review

A. EDA use in educational research

Some of the primary reasons why electrodermal activity (EDA) sensors are attractive in educational research is its ease of use, low cost, and ability to generate a near-real-time quantitative measure of testing events [15]. Some early studies have found its utility in detecting changes among students in different studies. For example, Harley and colleagues [17] conducted a study of a biology course offered to students of different disciplines (e.g., math, engineering, social science, business, art). Participants were randomly assigned to a multi-agent intelligent tutoring hypermedia system in a controlled laboratory environment. In this study, facial expressions, electrodermal activity, and self-reports were collected and showed that there was over a 60% agreement between facial expression readings and EDA and over 40% agreement between EDA and the self-reports. This suggests that EDA may have different correlations to other instruments, potentially due to varying contexts and the nature of the experimental design. Villanueva and colleagues [15], conducted a study on 88 first-year engineering students where self-reports and EDA were collected during five design courses throughout a semester, as part of a quasi-experimental study. Findings suggested that EDA increased for individual and collaborative active learning activities compared to lecture whereas no significant changes were found between individual and collaborative active learning activities.

McNeal and others [18] compared students engagement on different pedagogical approaches (traditional lectures, movie viewing, class dialogue, and journaling) and found increased levels of EDA during movie viewing. Dragon and colleagues [19] used EDA sensors to find a correlation between physical behavior, emotional state, and the learning process of 34 students of a Massachusetts public school where EDA was collected during a standard mathematics test. The results showed a strong positive correlation of EDA with physical movement, where authors coined these movement experiences as Joy, Yes, or Aha moments. Hardy and colleagues [20] measured EDA in 38 trainees in a computer programming (Java programming) training session to find a relationship between EDA and novice participants learning curve, as measured by a standardized test. A positive correlation was found between EDA and learning gains, a similar finding to other studies [21-23].

Collectively, the findings point to two factors: (a) the type of experimental design matters, as evidenced by work conducted in Harleys [17] and Villanuevas research designs [15]; (b) EDA signals are being equated to either emotions or learning, with no consideration or limited explanation of how the data was processed. The latter presents an area that warrants more exploration, as erroneous interpretations due to lack of proper data cleaning and processing can greatly influence the types of interventions and inferences conducted by researchers. To this end, our study provides details about how we cleaned our data by using movement as the parameter for filtration, as recommended previously by the authors [15].

B. Skin temperature as a thermoregulatory mechanism of cognitive function and performance

A limited amount of studies suggest that there is a potential link between skin temperature and performance. The phenomenon, thermoregulation, suggests that as skin temperature changes, so do an individuals ability to perform. Hartley and McCabe [24] performed a study on ten males and ten females where their core body temperatures was dropped to 35.5^{o}C, and a Performance Assessment Battery of tests was administered. The same test was conducted at room temperature. It was determined that subjects performed better at room temperatures compared to colder temperatures suggesting a temperature-dependent influence on cognitive information processing.

Lai and colleagues [25] conducted a study on 594 participants where the residential air temperature was changed between -5.8^{0}C to 25.7^{0}C as participants took a Mini-Mental State Examination (MMSE), a quantitative measure of cognitive function. The data showed that there is a higher risk of having a low cognitive function when extreme temperatures are present (high or low) but this cognitive function increased at ambient temperatures.

Collectively, the findings suggest a potential thermoregulatory influence of temperature on cognitive function and performance. To our knowledge, no research group has attempted to explore this in the context of classroom examination performance and more specifically, in engineering.

III. METHODOLOGY

A. Research Design

This study is part of a larger experimental design conducted at a western institution of the United States on engineering examination experiences, whose study was approved by the Institutional Review Boards of the home institutions of the authors. A subset of 76 participants between Fall 2017 and Spring 2018 was selected for this exploratory study. These two time points were selected as they represent different atmospheric seasonal temperatures, which may influence skin temperatures among participants differently. As such, a sub-analysis between the Fall term (44 participants) and the Spring term (32 participants) was conducted. The exam content between these two semesters was nearly the same in style and content and pre-approved by the course instructor.

To briefly discuss the overall experimental design, engineering students enrolled in a statics course were asked to take a practice exam, provided by the instructor, one week before the actual exam. For this particular course, exams are taken in a testing center in computer form. To mimic this same setup and ensure ecological validity, the experiment was conducted similarly to the testing center. The content of the exam also paralleled the actual exam in terms of topics and overall format. Similar to the actual exam, students worked on their exam problems on a testing booklet and entered their responses in multiple-choice on the computer. The instructor approved an exam cheat sheet, which was administered during the exam. As students completed the exam, they were asked to enter self-report information on several surveys while EDA data was collected.

Performance data were collected from students and was analyzed through its *difficulty index* (DI). DI accounts for the number of correct responses from a group of students divided by the total number of collective student responses [26]. DI also indicates whether a problem created by an instructor was difficult for the students or not. DI values of under 0.3 are considered too hard, and values over 0.8 are considered too easy. DI values in between 0.3 and 0.8 are deliberated as moderately difficult.

B. Hypothesis

In this experiment, we hypothesized the following:

- H1: There will be a strong negative correlation between electrodermal activity and skin temperature
- H2: There will be a strong positive correlation between electrodermal activity and difficulty index

This work is funded by the National Science Foundation Grant No. 1661100 and 1661117. Any opinions, findings, and conclusions or recommendations expressed in this materia
the author(s) and do not necessarily reflect the views of NSF.

- H3: There will be a strong negative correlation between skin temperature and difficulty index
- H4: There will be inverse correlations between varying seasonal temperatures and difficulty index

The rationale for these hypotheses is that as difficulty index increases, there will be a higher use of cognitive resources used by students that may lead to higher emotional responses and activation of electrodermal activity, which is reflected by an increase in sweat secretion and lowered skin temperature.

C. *EDA Data Cleaning and Processing*

To clean EDA data, we developed a customized protocol to filter out sources of noise in the data due to movement. The Empatica E4 sensors (Empatica, Inc.) includes four sensors, one of which is an accelerometer. Accelerometer data is collected from three directions (x, y, z) alongside time and EDA data, which is collected in micro Siemens. Together, they were used to remove any potential sources of noise due to motion artifacts (e.g., excessive hand or body movements) from the students as they solved their exam questions.

Noise reduction due to motion artifact in EDA was considered to include two primary steps: (1) rank-order participants EDA data; and (2) use accelerometer data to synchronize with EDA for outlier determination. A more detailed procedure is provided below:

1)*Data collection:* Data collection occurs as students work throughout the exam. For this study, Empatica E4 sensors were used for each student according to manufacturer recommendations [27, 28]. The E4 sensor collects data at 4Hz (1/4 second) [27, 28].

2)*Data repository:* Using the E4 manufacturer data management software, download the EDA data alongside its other four measures (ACC or accelerometer data: movement of the hands), blood volume pressure and body temperature. All these data are transferred to a local computer.

3)*L2-Norm calculation:* Pairing time, EDA, and ACC on a single data sheet, rank-order participant data using ACC. The aim is to identify those students who moved minimally during the exam since less movement helps provide a cleaner EDA response. ACC data is composed of 3 signals as movement can occur in the x-, y-, or z-direction. Thus, we combined the three ACC directions to calculate the total movement using the L2-norm equation [29] (Eqn. 1).

$$\text{L2-norm} = sqrt(x^2+y^2+z^2) \quad \textit{(Eqn. 1)}$$

4)*Standard deviation calculation:* To account for variations and skewness of the movement data, we calculated the standard deviation of the L2-norm across participants.

5)*Ranking:* We re-ranked the orders of the standard deviation calculated from the L2-norm equation. Smaller deviations correspond to minimum movement of hands. The mean of this standard deviation as well as two standard deviations to find the top and bottom 5% of the L2-norm data set (assuming a Gaussian distribution) was removed.

6)*Noise removal:* We removed the top and bottom 5% of the standard deviation of the total movement. With the time synchs between ACC and EDA, we paralleled the process with the EDA data . The removed EDA data was then replaced with the mean EDA data from the rest of the data. Using Ledalab, we extracted the corrected EDA data into tonic and phasic signals. Phasic EDA is used for studies as they represent a more immediate, reactive response compared to a tonic (baseline signal) [15].

For this study, we used the filtered phasic EDA for analysis as it represents the most immediate responses by the participants [15]. Skin temperature was paired with the EDA data and use to determine the correlations described below.

D. *Skin Temperature Collection*

Within the Empatica E4 sensor, there are many sensors, including an infrared thermophile sensor that collects data at 4Hz [28]. Thermopiles are designed to measure temperature from a distance by detecting an object's infrared (IR) energy. The higher the temperature, the more IR energy is emitted. The thermopile sensing element, composed of small thermocouples on a silicon chip, absorbs the energy and produces an output signal. In parallel to the EDA data, skin temperature was collected at the same time.

E. *Statistical Analysis*

Pearson correlation analysis was conducted between the filtered phasic EDA and skin temperature and difficulty index. Also, to compare if different seasonal temperatures in examination experiences influence performance, a t-test analysis was done between Fall 2017 and Spring 2018 data sets. Finally, to explore potential relationships between EDA, temperature, and difficulty index, a multiple regression model was conducted. For all purposes, significance was considered if p-values were under 0.05.

IV. results

To answer the first hypothesis, scatter plots were conducted between the mean of the filtered phasic EDA data and skin temperature of participants across the two semesters (Fig. 1). A weak, positive but significant correlation was found (r=0.13; p<0.05).

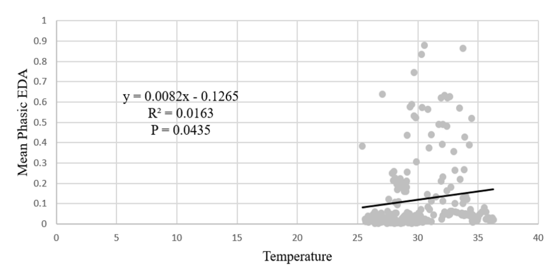

Fig. 1. Scatter plot of the relationship between mean filtered phasic EDA and skin temperature; p<0.05 indicates statistical significance

To explore the second hypothesis, a correlation analysis was done between the mean of the filtered phasic EDA and the mean of the difficulty indexes of the exams. We found a weak and significant positive correlation (r=0.16; p<0.01). The scatter plot is shown in Fig. 2.

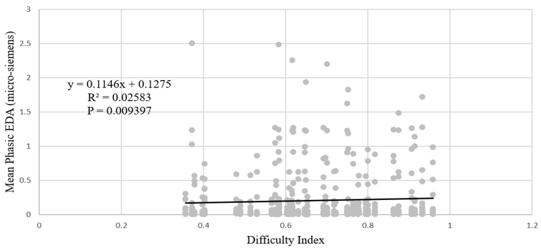

Fig. 2. Scatter plot of the relationship between mean filtered phasic EDA and difficulty index; p<0.05 indicates statistical significance

To test the third hypothesis, a correlation between skin temperature and the difficulty index was conducted (Fig. 3).

We found a weak, positive but a significant correlation (r=0.18; p<0.001).

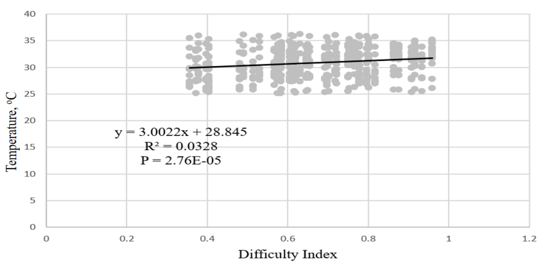

Fig. 3. Scatter plot of the relationship between skin temperature and difficulty index; p<0.05 indicates statistical significance

To test the fourth hypothesis, we compared the correlations between the two semesters and difficulty index. We found a weak positive correlation in the fall term (r=0.14; p<0.05) and a very weak positive correlation in the spring term (r=0.08; p<0.001).

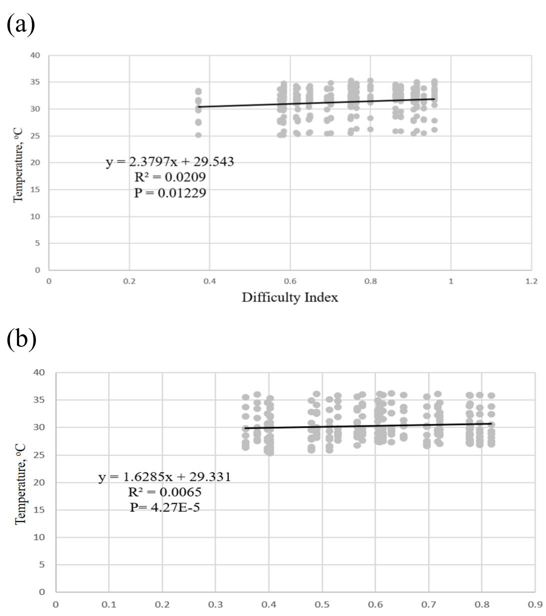

Fig. 4. The representative plots include correlations between skin temperature and difficulty index for Fall 2017 (panel a) and Spring 2018 (panel b).

Finally, a multiple regression model was conducted between skin temperature, difficulty index, and EDA. The equation is summarized in Eqn. 2.

$$\text{Mean Phasic EDA} = -0.02183 * \text{Temperature} - 0.11168 * \text{DI} + 1.03125 \qquad (\textit{Eqn. 2})$$

We found a moderate statistical significance among the three factors (r= 0.45; p< 0.05).

V. discussion

We found overall that in general, there was a very weak to weak positive correlation between skin temperature and DI and electrodermal activity and DI, respectively. Also, between electrodermal activity and skin temperature, a weak positive significant correlation was found. Furthermore, a multiple regression correlation plot for the mean filtered phasic EDA, mean skin temperature, and mean DI showed a moderate statistically significant model. Collectively, the data suggest a potential connection between cognition and physiological response, which corresponds to the circumplex model of affect [7] and what other researchers suggest [2,5]. It also shows a link between electrodermal activity, performance, and body temperature, which suggests a thermo-regulatory role to cognitive function [24, 25], a unique finding in this work.

VI. limitation

One of the limitations of this work is participant numbers. It is possible that these relationships may be more pronounced upon larger participant numbers or in conjunction with other constructs (e.g., self-report data), which our group is discovering (unpublished findings). We did not explore the entire array of data (e.g., blood volume pressure) that the Empatica E4 sensor has. As such, there may be additional information that may explain the trends found here. Future work will explore the role that these other factors (e.g., blood volume pressure) had in these observations. Besides, we did not include self-reported data based on academic emotions and self-efficacy [5], which we collected in this study. This may help us to more definitely delineate an emotional or cognitive role of these relationships [14, 30-35] to the physiological findings we found and help us identify associations with student performance across different types of exams. It is worth mentioning that while we reported r-values and p-values in this work, it is hard to infer any meaningful effect without gauging the strength of those effects in more detail [36].

VII. Conclusion

This research effort aims to understand if there are any physiological underpinnings to student performance on engineering exams. We found significant weak positive correlations between student performance and mean phasic EDA and skin temperature, respectively. The early findings suggest a physiological connection to exam performance. We also found a potential thermo-regulatory role in students cognitive performance, a unique finding in this work.

VIII.Implications

This work-in-progress can inform scholars and engineering educators of students near-real-time responses to examination experiences and can help shed light to exam design and conditions that may improve performance in these students. The physiological links to performance highlight the importance of students bodily temperature regulation in cognitive function suggesting the importance of learning environment ergonomics [24] in examination experiences.

Acknowledgment

This material is based upon work supported by the National Science Foundation (NSF) Grant No. 1661100 and No. 1661117. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF. We also would like to thank Darcie Christensen, Matthew Graham, Kate Youmans, and Dr. Shawn Lampkins for their assistance in data collection.

References

[1] P. Cartney and A. Rouse, The emotional impact of learning in small groups: highlighting the impact on student progression and retention, Teaching in Higher Education, vol. 11, no. 1, pp. 7991, 2006.

[2] R. Pekrun and M. B hner, Self-report measures of academic emotions, In R. Pekrun, & L. Linnenbrink Garcia, (Eds) *International Handbook of Emotions in Education* (pp. 561-566). London: Routledge Press, 2014.

[3] D. Hopkins, A teacher's guide to classroom research, *Philadelphia: Open University Press*, 1985.

[4] G. Spangler, R. Pekrun, K. Kramer, and H. Hofman, Students emotions, physiological reactions, and coping in academic exams, *Anxiety, Stress, & Coping*, vol. 15, no. 4, pp. 413-432, 2002.

[5] J. Husman, K.C. Cheng, K. Puruhito, and E.J. Fishman, Understanding engineering students stress and emotions during an introductory engineering course, *American Society of Engineering Education*, Paper ID 13148, 2015.

[6] M. Benedek and C. Kaernbach, A continuous measure of phasic electrodermal activity, *Journal of Neuroscience Methods*, vol. 190, no. 1, pp. 8091, 2010.

[7] W. Boucsein and R.W. Backs, Engineering psychophysiology as a discipline: Historical and theoretical aspects. In R.W. Backs & W. Boucsein (Eds.), *Engineering psychophysiology*. Issues and applications (pp. 330). Mahwah, NJ: Lawrence Erlbaum, 2000.

[8] W. Boucsein and R.W. Backs, The psychophysiology of emotion, arousal, and personality: Methods and models, In V. G. Duffy (Ed.), *Handbook of digital human modelin*g (pp. 3538). CRC: Boca Raton, 2009.

[9] A. Schwerdtfeger, Predicting autonomic reactivity to public speaking: don't get fixed on self-report data! *International Journal of Psychophysiology*, vol. 52, no. 3, pp. 217-224, 2004.

[10] G. Turpin, P. Shine, and M.H. Lader, Ambulatory electrodermal monitoring: effects of ambient temperature, general activity, electrolyte media, and length of recording, *Psychophysiology*, vol. 20, pp. 219224, 1983.

[11] T. Lobstein and J. Cort, The relationship between skin temperature and skin conductance activity: Indications of genetic and fitness determinants, *Biological Psychology*, vol. 7, pp. 139143, 1978.

[12] T. Scholander, Some measures of electrodermal activity and their relationships as affected by varied temperatures, *Journal of Psychosomatic Research*, vol. 7, pp. 151158, 1963.

[13] I. Villanueva, M. Valladares, and W. Goodridge, Use of galvanic skin responses, salivary biomarkers, and self-reports to assess undergraduate student performance during a laboratory exam activity, *Journal of Visualized Experiments*, vol. 108, no. e53255, pp. 1-8, 2016.

[14] I. Villanueva, J. Husman, K. Youmans*, D. Christensen*, P. Vicioso*, M.T.H. Khan, S. Lampkins, & M. Graham. (2019). Considerations for studying near-real-time authentic examination experiences: A cross-disciplinary & multi-modal experimental approach. *Journal of Visualized Experiments,* 2019, under review.

[15] I. Villanueva, B. Campbell, A. Raikes, S. Jones, and L. Putney, A multi-modal exploration of engineering students emotions and electrodermal activity in design activities, *Journal of Engineering Education*, vol. 107, no. 3, pp. 414-441, 2018.

[16] I. Singh and A. Jha, Anxiety, Optimism and Academic Achievement among Students of Private Medical and Engineering Colleges: A Comparative Study, Journal of Educational and Developmental Psychology, vol. 3, no. 1, 2013.

[17] J.M. Harley, F. Bouchet, S. Hussain, R. Azevedo, & 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.

[18] K. S. Mcneal, J. M. Spry, R. Mitra, and J. L. Tipton, Measuring Student Engagement, Knowledge, and Perceptions of Climate Change in an Introductory Environmental Geology Course, Journal of Geoscience Education, vol. 62, no. 4, pp. 655667, 2014.

[19] T. Dragon, I. Arroyo, B. P. Woolf, W. Burleson, R. E. Kaliouby, and H. Eydgahi, Viewing Student Affect and Learning through Classroom Observation and Physical Sensors, Intelligent Tutoring Systems Lecture Notes in Computer Science, pp. 2939, 2008.

[20] M. Hardy, E. N. Wiebe, J. F. Grafsgaard, K. E. Boyer, and J. C. Lester, Physiological Responses to Events during Training, Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 57, no. 1, pp. 21012105, 2013.

[21] B. Woolf, W. Burleson, I. Arroyo, T. Dragon, D. Cooper, and R. Picard, Affect-aware tutors: recognising and responding to student affect, International Journal of Learning Technology, vol. 4, no. 3/4, p. 129, 2009.

[22] L. Shen, W. Minjuan, and S. Ruimin, "Affective e-learning: Using emotional data to improve learning in pervasive learning environment." Journal of Educational Technology & Society, vol. 12, no. 2, pp. 176-189, 2009.

[23] I. Arroyo, D. G. Cooper, B. Winslow, W. B. Park, M. Kasia, and C. Robert. "Emotion sensors go to school." In AIED, vol. 200, pp. 17-24. 2009.

[24] M.D. Hartley and J. McCabe, The effects of cold on human cognitive performance: Implications for design." In *IEE conference publication*, pp. 310-315. Institution of Electrical Engineers, 2001.

[25] L. Dai, I. Kloog, A.B. Coull, D. Sparrow, A. Spiro, P.S.Vokonas, and J.D. Schwarz, Cognitive function and short-term exposure to residential air temperature: A repeated measures study based on spatiotemporal estimates of temperature, Environmental Research, vol. 150, pp. 446-451, 2016.

[26] J. Johari, J. Sahari, D.A. Wahab, S. Abdullah, S. Abdullah, M.Z. Omar, and N. Muhamad, Difficulty index of examinations and their relation to the achievement of programme outcomes, Procedua-Social and Behavioral Sciences, vol. 17, pp. 71-80, 2011.

[27] S. C. Mukhopadhyay, Wearable Sensors for Human Activity Monitoring: A Review, IEEE Sensors Journal, vol. 15, no. 3, pp. 13211330, 2015.

[28] Empatica. E4 wristband from Empatica: Users manual. Empatica, 1-32, 2018

[29] R. Abbasi-Kesbi and A. Nikfarjam, Denoising MEMS accelerometer sensors based on L2-norm total variation algorithm, Electronics Letters, vol. 53, no. 5, pp. 322324, 2017.

[30] L. Kulke, D. Feyerabend, and A. Schacht, A comparison of the Affectiva iMotions Facial Expression Analysis Software with EMG for identifying facial expressions of emotion, 2018

[31] R. W. Picard, S. Fedor, and Y. Ayzenberg, Multiple Arousal Theory and Daily-Life Electrodermal Activity Asymmetry, Emotion Review, vol. 8, no. 1, pp. 6275, 2015.

[32] T. Deltombe, P. Hanson, J. Jamart, and M. ClRin, The influence of skin temperature on latency and amplitude of the sympathetic skin response in normal subjects, Muscle & Nerve, vol. 21, no. 1, pp. 3439, 1998.

[33] R. Vetrugno, R. Liguori, P. Cortelli, and P. Montagna, Sympathetic skin response, Clinical Autonomic Research, vol. 13, no. 4, pp. 256270, 2003.

[34] E. Austin, D. Saklofske, and S. Mastoras, Emotional intelligence, coping and exam-related stress in Canadian undergraduate students, Australian Journal of Psychology, vol. 62, no. 1, pp. 4250, 2010.

[35] MD. T. H. Khan, K. Raoufi, K. Park, T. Reza, C. E. Psenka, K. S. Jackson, K. R. Haapala, G. E. O. Kremer, and K.-Y. Kim, Development of learning modules for sustainable life cycle product design: a constructionist

approach, ASEE, 2017. [Online]. Available: https://peer.asee.org/28174.pdf.

[36] R. Wasserstein, American Statistical Association releases statement on statistical significance and p-values: provides principles to improve the conduct and interpretation of quantitative science, pp. 1-3, 2016.