



Human-Technology Frontier: Measuring Student Performance-Related Responses to Authentic Engineering Education Activities via Physiological Sensing

Idalis Villanueva¹(✉), Edwin Marte Zorrilla¹, Jenefer Husman²,
and Matthew Graham²

¹ Engineering Education Department, University of Florida, Gainesville, FL 32611, USA
i.villanueva@ufl.edu

² Education Studies, University of Oregon, Eugene, OR, USA

Abstract. Multi-modal approaches have increasingly shown promise in exploring the human side of engineering via the assessment of authentic responses to learning or working environments. This study explores the utility of non-invasive physiological wrist sensors in measuring the reactive and regulatory responses of a group of 161 engineering students taking an authentic engineering practice exam. The practice exam was categorized into Conceptual problems (e.g., rote memorization) and Analytical problems (e.g., requiring application of learned concepts through equations and free-body diagrams). Responses were measured through electrodermal activity and indicators of performance. Findings identified that the type of practice exam problem, even if designed to be within the moderate range of difficulty, influenced how students reacted to and regulated their performance to the problem (as seen by stronger positive correlations in the Analytical problems) and that these may occur via multi-componential processes.

Keywords: Human-technology frontier · Human side of engineering · Engineering education · Physiological sensing · Student performance

1 Introduction

In science, technology, engineering, and math (STEM), workforce development initiatives necessitate that trainees learn about and perform as intended by their leads in their immediate learning environment. In turn, the actions and decisions of STEM trainees will determine how these individuals will manage future societal and complex problems post-graduation. In the classroom environment, learning and performance are traditionally studied through various techniques such as self-report instruments, artifact retrieval and analysis, sampling of experiences, interviews, or observational procedures [1–3]. These techniques, while valuable, are constrained to participant or observer subjectivity and restricts its understanding of individuals' authentic responses to a task in real-time.

The latter limitation poses a challenge for many scholars seeking to objectively measure such events in real-time.

The situational contexts individuals may find themselves in, for example in the classroom, are important to consider when examining students' performance. Educational psychology scholars posit that in the context of the classroom, learners experience test-related, class-related, or learning-related multi-component input processes such as affect, physiological responses, motivational changes, and cognitive fluctuations [2, 3]. Subsequently, these alter outcome processes such as decision-making and performance [2, 3]. Yet, there is a scarcity of empirical investigations exploring the relationships of these input and outcomes processes in real-time.

Recent and nascent literature has shown the promise of multi-modal approaches to explore the authentic responses of students to their learning environments and tasks [2–9]. For example, Pekrun and his colleagues examined stress responses during a final oral exam [2] using salivary cortisol; we have used this same method during engineering exams [5–7]. Azevedo and others have examined students' metacognitive judgments using physiological sensors when individuals interact with a meta-tutor environment [8]. Villanueva and colleagues have begun to tie engineering content (i.e., engineering statics and design) to physiological responses during in-classroom activities (active learning; [9]) and out-of-classroom exam-taking experiences (statics practice exams; [7]). Recently, Strohmeier and colleagues [10] utilized electrodermal activity sensors with self-reports to explore state (*momentary*) anxiety to mathematical tests. Together, these studies are pointing to an increasing desire from scholars to develop studies and methods that are more authentic to the situations and contexts that different learners are tasked to perform in.

For this work, the utility of electrodermal activity as a non-invasive electrodermal activity sensing tool to explore students' input physiological processes (i.e., reactive or regulatory emotions, motivations, etc.) related to its output processes (i.e., effort, performance, etc.) was explored. In addition, this study also tabulated the instructors' expected difficulty levels for the exam questions and students' actual performances to better contextualize the practice exam used for this study.

Electrodermal activity (EDA) is an umbrella term used to define changes in the electrical properties of the skin. The most widely studied electrical property of the skin is known as *skin conductance*. Skin conductance is quantified by applying an electrical potential between two contact points in the skin and measuring its electrical current flow. The resultant flow, stemming from sympathetic neuronal activity, is captured as both a background tonic (skin conductance level or SCL) and fast-varying phasic (skin conductance response or SCR) signal [11–13]. EDA is useful in identifying changes in sympathetic arousal, either through emotional or cognitive states, and it is the only autonomic psychophysiological variable that is not connected to parasympathetic activity [11–13]. In other words, when a person becomes nervous or anxious about a task, their body stimulates metabolic outputs to deal with external challenges (e.g., their palms become sweaty). Since emotional regulation and cognitive processes, among other brain functions, can influence the control of sweating, a fluctuation in the electrical conductivity of the skin can be measured. Also, these measurements can be collected rather

cheaply and non-invasively [14, 15] (e.g., EDA wrist sensors) making physiological EDA sensing tools attractive for scholars.

2 Methods

Research Design. This study is composed of a subset of data derived from a larger study exploring the motivational and performance experiences of engineering students performing on an authentic practice exam in the topic of statics [7]. The study was quasi-experimental and integrated EDA sensing with salivary biomarkers of stress along with self-reports in the practice engineering exam for triangulation. Per Institutional Review Board (IRB) guidelines and in alignment with human subjects' research, a practice exam equivalent in content to the actual exam was provided to prevent unnecessary harm to students' grades in the course.

The total number of participants reported in this study was 161 engineering students. These students took an engineering statics course for the first time in a Mountain West institution in the United States between Fall 2018, Spring 2019, and Fall 2019. The instructor has been teaching engineering statics for several years and the exam used for this study was designed by this faculty member. The 15-question exam (6 conceptual and 9 analytical) used was the same throughout the semesters in this study. For this study, students were given ample time to complete the questions, as approved by the instructor.

Data Collection and Measures. The setup of the experiment and its considerations is described elsewhere [7] but in summary, involved a time-stamped research design that measured the time students took to perform per practice exam question, their performance to each item, the raw EDA data from each practice exam question and event (before and after the exam), salivary data (not reported in this work) and self-reports (not reported in this work). For this study, we wanted to explore how reactive input processes, in this case, sympathetic emotional arousal (measured through skin conductance responses), are related to output processes (measured through difficulty index scores of performance) in authentic engineering examination settings.

There were several measurements collected in our study. The first measurement was the *event-specific skin conductance response* (es.SCR), which allowed us to attribute a specific eliciting stimulus [14] to its fast-varying reactive responses. The second measurement was the *difficulty index* (DI), which is the measure of the number of questions that students answered correctly on the practice exam divided by the total number of responses. DI is also an indication of whether a problem created by the instructor was difficult ($DI < 0.3$), moderate ($0.3 < DI < 0.8$) or easy ($DI \geq 0.8$) [16]. The third measurement was the instructor's self-reported difficulty level for the exam, attained through a three-point Likert scale where '1' was reported as 'easy', '2' was reported as 'moderate' and '3' was reported as 'difficult'. The fourth measure was the average time spent in Conceptual and Analytical exam problems; this measured time is an indication of effort executed by the student. Finally, we tracked the percentage of students who received fully correct responses to Conceptual and Analytical exam problems to understand students' effectiveness in their performance.

Data Processing and Analysis. For the collected and continuous raw EDA data, we first visually identified and removed any data tied to potential manufacturer sources of outliers, such as sensor malfunctions, incomplete collection, or poor contact of the electrodes in the skin. These were evidenced by negative values or constant near-zero continual data segments [7]. Then, additional EDA outliers due to user-generated sources such as erratic movements (e.g., hand hitting desk or nervous tapping), survey or salivary biomarker collection periods, or large changes in body temperatures or blood volume pressure readings [7] were removed. This required that the temperature and blood volume pressure sensors found in the wrist sensor (E4, Empatica, Inc.) was used in parallel to the collected EDA data. Next, the authors used a custom-developed process to remove any potential contributions to the raw EDA signal via fine motor body movements from the participant (e.g., typing) [7]. This process required that the 3-axis accelerometer found in the EDA sensor was used in parallel to the collected galvanic skin response data.

To summarize, this processing method used an L2-norm accelerometer transformation of the 3-axis of movement (in the x, y, z directions) to calculate the Euclidean distance of movement (total movement). From the collected signal, the standard deviation and coefficient of variation (CV) of the accelerometer data were calculated, ranked-ordered to identify its upper and lower limits, and a threshold equivalent to 95% of the standard deviation of the accelerometer data was established. The corresponding raw EDA values to this transformed accelerometer data were further used for filtration. Filtration of this signal was conducted through continuous deconvolution methods developed in the open-source MATLAB software, Ledalab [17]. The resultant continuous data was then separated into baseline tonic (skin conductance level-SCL) and fast-varying phasic (skin conductance responses-SCR) EDA signals. For this study, we were interested in the deconvoluted continuous phasic EDA signal, although we acknowledge that this represents “a small proportion of the overall EDA complex” [17, p. 4]. The resultant processed and filtered EDA phasic signal labeled as a non-specific skin conductance response ($ns.SCR_i$) was aligned to the timestamps of a particular event (i.e., duration of time for a practice exam question) to extract the *event-specific SCRs* ($es.SCR_i$), which measures the specific eliciting stimulus [17] of a given event to its fast-varying reactive responses. The number of peaks present in these $es.SCRs$ ($es.SCR_{peak}$) was used to measure the reactivity of the response while the mean of these $es.SCRs$ ($es.SCR_{mean}$) was used to measure the regulatory potential of these responses.

For the processing the DI data, we calculated the number of exam questions answered correctly on the exam divided by the total number of responses. This DI value was divided by time for each event. For the instructor assessment, we calculated the average scores from the Likert Scale responses for the three semesters. The practice exam questions were categorized as Conceptual (those requiring rote memorization or interpretation of terminology) or Analytical (those that required application of the concepts learned in the form of equations, identification of parameters provided, and drawing of free body diagrams to attain an answer). The conceptual and analytical questions were timestamped and aligned to its corresponding $es.SCR_i$ and DI data. Time and percentages of performance for the exam problems were tabulated and a t-test analysis was conducted.

For analysis of the $es.SCR_i$ data, we followed recommended practices to transform and standardize EDA data [7, 11, 14, 17]. For a normalized transformation, the $\text{Log}(es.SCR_i + 1)$ was calculated to remove any problems with skewness, kurtosis, and

heterogeneity of variance in the EDA data [17]. For standardization of the EDA data, the $\text{Log}(\text{es.SCR}_i + 1)$ was divided by the time per practice exam question to make group comparisons by event. Descriptive statistics (a t-test for equality of means) was determined for the transformed and standardized mean and peak es.SCRs, DI, the instructor averaged estimated scores, time, and percentage of students who answered the practice exam questions correctly; data are presented as mean $\pm$ standard error of the mean. Additionally, a Pearson correlation analysis was conducted for the transformed and standardized mean and peak es.SCRs with DI scores per type of practice exam question (Conceptual and Analytical).

3 Results and Discussion

T-test analysis was conducted for the Conceptual and Analytical problems and significance was determined by a p-value under 0.05. The time spent on the exam problems was tabulated as a measure of effort from students for the Conceptual and Analytical questions. There was a 3.5x-fold increase in the Analytical problems compared to the Conceptual problems (372.98 ± 8.47 s versus 80.58 ± 1.22 s; $p < 0.001$). Also, 65% of the students answered the Conceptual questions correctly while 58% of the students answered the Analytical questions correctly; no statistical significance was found between these groups ($p = 0.69$). The DI scores were 0.64 ± 0.01 for the Conceptual questions and 0.68 ± 0.00 for the Analytical questions and these differences were significant ($p < 0.001$).

The instructor estimated the difficulty level of practice exam questions as being twice as higher for the Analytical problems (2.04 ± 0.01 ; moderate) compared to the Conceptual problems (1.21 ± 0.02 ; easy). These estimates were significantly different from each other ($p < 0.001$). A summary of these t-tests is provided in Table 1.

Table 1. Descriptive statistics for conceptual and analytical problems in an engineering statics practice exam; N = 161; data presented as Mean $\pm$ Standard Error of Mean; statistical significance as calculated through T-test is * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ when comparing analytical to conceptual problems.

Measurement	Exam problem type	Mean $\pm$ SEM
Time spent	Conceptual	$80.58 \text{ s} \pm 1.22 \text{ s}$
	Analytical	$372.98 \text{ s} \pm 8.47 \text{ s}^{***}$
Students who answered correctly	Conceptual	$65\% \pm 0.00\%$
	Analytical	$58\% \pm 0.00\%$
Instructor guess	Conceptual	1.21 ± 0.02
	Analytical	$2.04 \pm 0.02^{**}$
Difficulty index	Conceptual	0.64 ± 0.01
	Analytical	$0.68 \pm 0.00^{***}$

Pearson correlation analysis (Table 2) was conducted for the transformed and standardized es.SCR_{mean} and es.SCR_{peak} data and DI scores (both divided by time) for the Conceptual and Analytical problems. For reactive responses (through es.SCR_{peak} values), both Conceptual and Analytical exam problems resulted in strongly positively correlated reactive responses when compared to DI/time ($r = 0.64$, $p < 0.01$ and $r = 0.85$, $p < 0.01$). For regulatory potential responses (through es.SCR_{mean} values), a strong positive correlation for the Conceptual problems ($r = 0.59$, $p < 0.01$) and a weak positive correlation for the Analytical problems ($r = 0.21$, $p < 0.01$) compared to DI/time was found.

Table 2. Pearson correlation analysis between $\log(\text{es.SCR}_{\text{peak}} + 1)$ or $\log(\text{es.SCR}_{\text{mean}} + 1)$ versus DI/time to measure students' reactive or regulatory responses, respectively; $N = 161$; data presented as Mean $\pm$ Standard Error of Mean; statistical significance is * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Response type	Exam problem type	Correlation
Reactive	Conceptual	0.59
	Analytical	0.21**
Regulatory	Conceptual	0.64
	Analytical	0.85**

Collectively, the data suggest that: (1) instructors' assessment of the difficulty level of the practice exam problems does not necessarily match students' performance to the same practice exam problems; (2) more effort (as seen by increases in time) by the student to practice exam problem types may slightly improve performance but not sufficiently to cause an improvement in the number of students excelling; (3) the type of practice exam problem, even if perceived to be within the moderate range of difficulty (DI scores of 0.64 and 0.68, respectively) influences how students react to and regulate their execution of the problem (stronger correlations in Analytical problems).

These findings suggest the need to better align exam difficulty levels with students' zone of proximal development [18, 19], a learning theory suggesting that students best perform and learn when instructors understand the zone by which students best learn and can be 'pushed' to gain further motivation to learn. The zone of proximal development states that when faculty pushes students through unattainable perceived boundaries, a students' motivation decreases, and their performance may be negatively impacted. Thus, finding a right balance between perceived difficulty and actual performance of students is important in examination contexts.

Furthermore, the findings in this study demonstrate that responses to an exam experience can be measured, quantified, and connected to its context (e.g., exam problem types, instructor creation of 'just-right' levels of difficulty for students). This suggests the presence of multi-componential processes tying motivational with cognitive constructs. This parallels the educational psychology framework of the Control-Value Theory of emotions [2, 20] that posits that emotions are inextricably linked to multi-componential processes such as motivation, affect, cognition, and psychophysiology. Thus, it is important to not disconnect individuals' output processes (i.e., performance) from these input

processes (e.g., emotions) in their learning or working environments. Collectively, the data points to the need to guide students to cope with unexpected challenges and to perform in ways originally anticipated by the instructor.

4 Implications to the Human-Technology Frontier

Taleb suggests [21] that 21st century skills (e.g., complex problem solving) are both applicable and antifragile (involving individual growth when exposed to “volatility, randomness, disorder, and stressors... risk, and uncertainty.” [21, p. 3]. Even the National Academy of Engineering has recognized that 21st century skills for engineering can’t function without equipping its workforce to manage emotional experiences such as social distress, mental and emotional health, which influence decision-making and performance [22]. As suggested by these publications and the Accreditation Board for Engineering and Technology (criterion 7; [23]), there is and will be a need to holistically integrate learning experiences to include appropriate teaching/training strategies that help students gain an ability to multi-modally acquire and apply new knowledge effectively.

There seems to be a growing recognition of the importance of multi-componential processes in learning and performance, not just in the classroom but also in the way society views the future of engineering. Thomas Insights published an article delineating the top 20-in demand skills for employees in unexpected environments and scenarios such as those expected to be experienced in a post-COVID-19 world [24]. Some of these skills include critical thinking, complex problem solving, and stress tolerance. In a similar vein, Qadar and Al-Fuqaha [25] conducted an in-depth literature review of the seven steps needed to succeed in engineering education post-COVID-19. Step one of their model suggests the need for students to start with the end in mind by aligning efforts with product outcomes and motivations with effective performance.

The findings demonstrated in this work demonstrate the utility that physiological sensing can have in helping scholars/educators to cross the human-technology frontiers in workforce training and development. Applying multi-modal approaches such as the ones presented here provide a potential pathway to study more authentically the human side of learning and performance in engineering.

Acknowledgment. This material is based upon work supported in part by the National Science Foundation (NSF) No. EED-1661100/1661177. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect those of NSF.

Author contributions in this paper: Villanueva Alarcón (research design, data collection and analysis, writing, editing); Marte Zorrilla (data collection, analysis, writing, editing); Husman (research design, data collection, editing); Graham (writing and editing).

References

1. Jarrell, A., Harley, J.M., Lajoie, S., Naismith, L.: Success, failure, and emotions: examining the relationship between performance feedback and emotions in diagnostic reasoning. *Educ. Tech. Res. Dev.* **65**(5), 1263–1284 (2017)
2. Pekrun, R., Bühner, M.: Self-report measures of academic emotions. In: Pekrun, R., Linnenbrink-Garcia, L. (eds.) *International Handbook of Emotions in Education*, pp. 561–566. Routledge Press, London (2014)

3. Roos, A.L., et al.: Test anxiety and physiological arousal: a systematic review and meta-analysis. *Educ. Psychol. Rev.* **33**, 1–40 (2020)
4. Spangler, G., Pekrun, R., Kramer, K., Hofmann, H.: Students' emotions, physiological reactions, and coping in academic exams. *Anxiety Stress Coping* **15**(4), 413–432 (2002)
5. Husman, J., Cheng, K.: When performance really matters: future time perspective and career aspirations. In: European Association of Learning and Instruction, Limassol, Cyprus (2015)
6. Husman, J., Cheng, K.C., Barnes, M.E., Espino, K.: Managing emotions when it counts: stress, academic achievement emotions, and perceptions of instrumentality. In: American Educational Research Association, Chicago, IL (2016)
7. Villanueva, I., et al.: A cross-disciplinary and multi-modal experimental design for studying near-real-time authentic examination experiences. *J. Vis. Exp.* **151**(e60037), 1–10 (2019)
8. Azevedo, R., Taub, M., Mudrick, N., Farnsworth, J., Martin, S.A.: Interdisciplinary research methods used to investigate emotions with advanced learning technologies. In: Zembylas, M., Schutz, P.A. (eds.) *Methodological Advances in Research on Emotion and Education*, pp. 231–243. Springer (2016)
9. Villanueva, I., Campbell, B., Raikes, A., Jones, S., Putney, L.: A multi-modal exploration of engineering students' emotions and electrodermal activity in design activities. *J. Eng. Educ.* **107**(3), 414–441 (2018)
10. Strohmaier, A.R., Schiepe-Tiska, A., Reiss, K.M.: A comparison of self-reports and electrodermal activity as indicators of mathematics state anxiety. An application of the control-value theory. *Frontline Learn. Res.* **8**(1), 16–32 (2020)
11. Boucsein, W., Backs, R.W.: Engineering psychophysiology as a discipline: historical and theoretical aspects. In: Backs, R.W., Boucsein, W. (eds.) *Engineering Psychophysiology. Issues and Applications*, pp. 3–30. Lawrence Erlbaum, Mahwah (2000)
12. Boucsein, W., Backs, R.W.: The psychophysiology of emotion, arousal, and personality: methods and models. In: Duffy, V.G. (ed.) *Handbook of Digital Human Modeling*, pp. 35–38. CRC, Boca Raton (2009)
13. Sano, A., Picard, R.W., Stickgold, R.: Quantitative analysis of wrist electrodermal activity during sleep. *Int. J. Psychophysiol.* **94**(3), 382–389 (2014)
14. Braithwaite, J.J., Watson, D.G., Jones, R., Rowe, M.: A guide for analyzing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments (Technical report, 2nd version). Behavioural Brain Sciences Centre, Birmingham (2015)
15. Empatica. Empatica E4 wristband (2016). <https://www.empatica.com/e4-wristband>
16. Johari, J., et al.: Difficulty index of examinations and their relation to the achievement of programme out comes. *Procedia Soc. Behav. Sci.* **18**, 71–80 (2011)
17. Benedek, M., Kaernbach, C.: A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **190**(1), 80–91 (2010)
18. Vygotsky, L.S.: *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge (1978)
19. Vygotsky, L.S.: *Thought and Language*. MIT Press, Cambridge (1986)
20. Pekrun, R.: The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. *Educ. Psychol. Rev.* **18**(4), 315–341 (2006)
21. Taleb, N.N.: *Antifragile: Things That Gain from Disorder*. Random House, New York (2012)
22. National Academy of Engineering. <https://www.nae.edu/230195/Call-for-Engineering-Response-to-the-COVID19-Crisis>
23. ABET Engineering Accreditation Commission. Criteria for accrediting engineering programs effective for reviews during the 2020–2021 accreditation cycle (2019)
24. Thomas Insights. <https://www.thomasnet.com/insights/top-10-in-demand-skills-in-a-post-covid-world-according-to-world-economic-forum/>
25. Qadir, J., Al-Fuqaha, A.: A student primer on how to thrive in engineering education during and beyond COVID-19. *Educ. Sci.* **10**(9), 1–22 (2020)