Stretched Too Much?

A Case Study of Engineering Exam-Related Predicted Performance, Electrodermal Activity, and Heart Rate

D. Christensen
Utah State University – Engineering Education Department
Logan, Utah

M. T. H. Khan
Utah State University – Engineering Education Department
Logan, Utah

I. Villanueva[1]
Utah State University – Engineering Education Department
Logan, Utah

J. Husman
University of Oregon – Education Studies Department
Logan, Utah

Conference Key Areas: New notions of interdisciplinarity in engineering education

Keywords: Electrodermal activity, performance, engineering, exam, heart rate

ABSTRACT

Test writing is one of the essential activities that university faculty must do. Evidence-based instructional practice indicates that the exam content and difficulty should match the content taught in the course. Many faculty, however, hold the belief that tests should "stretch" students to tease out the best students or to extend content beyond what is covered in a course. In this case study, we explored if exam items, which are in the scope of the course but are "a stretch," affected engineering students' ability to self-monitor and reflect on performance. We compared and contrasted two examination experiences from the same engineering statics course. In scenario one, students recently learned a concept, and their practice exam reflected that content. In scenario two, students had yet to learn the concepts contained in the practice exams, but the concepts were related to the course. We explored this from a pre- and post-dicted expected performance, actual performance, and physiological response (electrodermal activity and heart rate) perspective for 26 engineering students.

This research examines the relationship between expected performance, actual performance, time per question or exam, and arousal response. Findings suggest the pre- and post-dicted expected performances may influence physiological responses (e.g., electrodermal activity and heart rate), which may not necessarily support students' actual performances on the exam.

---

[1]Corresponding Author
I. Villanueva
idalis.villanueva@usu.edu

Introduction

The Centre for Teaching Excellence at the University of Waterloo [1] states that "A good exam gives all students an equal opportunity to fully demonstrate their learning." The complexity lies in determining what a good exam looks like. What content does it contain? How difficult is too difficult? The question about the difficulty of an exam becomes more prominent in fields like science, technology, engineering, or math (STEM); for example, in engineering, the typical weed-out culture in the field leads to high expectations, many times evidenced in the form of very challenging exams [2].

Vygotsky defined the zone of proximal development as the zone where students cannot necessarily do something entirely on their own but can achieve mastery with encouragement [3]. Transferring this to exams, exams can be an experience that is knowledge building or disabling depending on how much the students are expected to "stretch." In cases such as engineering, we suspect the latter may be present. One main motivation for this study was to determine the effect that exam experiences have on the performance of engineering students in near-real-time. Through physiological biometric tools, we hope to understand better how these engineering students respond to and react to different exam approaches and situations. The results from this study can better inform instructor formation of assessments and help educators understand what may "stretch" engineering students too much.

THEORETICAL FRAMEWORK

Pre- and Post-dicted Performance

A limited number of scholars have explored the relationships that student performance predictions (both 'pre-dictions' before the exam and 'post-diction' after the exam) had with the actual performance on exams. For example, in a study by Hacker et al. [4], researchers explored student pre- and post-dictions in an introductory educational psychology course at a mid-southern university (N=99 undergraduates). Students who answered more than 70% of test questions correctly had closer pre- and post-dictions of their actual exam scores than those answering fewer exam questions correctly. Students who answered less than 50% of test questions correctly were overly confident in their pre- and post-dictions [4]. It is important to note that the exam given to the students was within the scope of content that was familiar to them because it had been introduced in class previously [4]. To our knowledge, there is a lack of studies exploring student pre- and post-dictions on exams whose content is within the scope of the course but are outside the content that has been covered in class (referred to as a "stretch"). This study will explore pre- and post-dictions for the student in a "stretched" situation. Also, we will explore how students react to the "stretch" scenario near-real-time during the examination experience through physiology based techniques (e.g., electrodermal activity, heart rate).

Electrodermal Activity and Performance

Due to the development of wearable sensors, continuous skin conductance levels can be recorded in near-real-time scenarios [5], which is attractive to educational researchers. A commonly used signal that measures physiological arousal from the electrical conductivity of the skin is the electrodermal activity (EDA) [6]. EDA is a measure of the activity within the sympathetic (fight/flight system) autonomic nervous system [6]. EDA data can be divided into phasic (stimulus-specific and immediate) and tonic (baseline) forms [6]. Phasic EDA evidences physiological arousal, indicating a cognitive activation or emotional strain [6].

EDA has been reported to be a function of task difficulty, with lower levels of arousal measured when the problem is more difficult [7]. It is assumed that this limited rise in arousal may have to do with inhibitory control [7], which is the ability one has to resist distractions and give full attention to the relevant stimuli due to increased cognitive load [8].

Changes in EDA can provide insight into understanding student performance [9]. High skin conductance indicates an increase in arousal, which is significantly positively correlated with peak performance [10]. Both the studies previously stated dealt with sports and video games, but significant correlations have also been found between academic performance and EDA responses [5]. There is a lack of studies of academic situations when students are not familiar with exam content. This study aims to explore further this phenomenon through a closer examination of the relationships between EDA, expected performance (as reported by the student), and actual performance in both a context where students are familiar with the exam content and a context where the students have not yet learned the exam content.

Heart Rate and Performance

Heart rate has been investigated as an indicator of student stress during various academic activities [11]. It has been found that anxiety is a strong influencing factor for heart rate increases with exams [12]. Heart rate can give insight into the magnitude of stress experienced by students [11] while participating in an exam experience. It has been found that student heart rates can be about 35% higher during an exam than during lecture and a 26% increase in heart rate was observed while students were receiving the results of an exam [11]. During a French oral exam (N=23 first-year college students), a significant positive correlation was found between average heart rate and performance [13], but during a driving test (N=13), a significant negative correlation was found between average heart rate and performance [14]. This suggests that attentional and cognitive shifts may affect or even reduce changes in an individual's heart rate [15]. Since there is a gap for written exam performance and heart rate, this study also sought to determine the relationship between heart rate and performance during a written exam with differing testing scenarios.

Perceived Stress and Performance

As mentioned in the Heart Rate and Performance section, anxiety and stress can be influencers of physiological responses and performance. In a study of 202 dental students, who are known to report high levels of stress, it was found that there was limited support for an association between academic performance and perceived dental environment stress [16]. At the beginning and middle of a semester in a pre-degree science university program, there were no significant correlations between perceived stress and academic performance, but at the end of the semester, there was a significant negative correlation between perceived stress and academic performance [17]. There appears to be a gap in the understanding of perceived student stress and exam performance in a different exam situation. This study will further investigate this gap by exploring perceived stress in students who are not familiar with exam content.

METHODS

All procedures were approved by the Utah State University and University of Oregon Institutional Review Board (IRB) offices for studies on human subjects.

Research Design

This study involves a customized experimental setup that includes a laptop computer interfaced with two web-cameras, a customized timestamping program created in MATLAB, and a customized exam protocol, which recorded student responses as well as the time it took the students to respond to the questions. A representative experimental timeline can be found in *Fig. 1*.

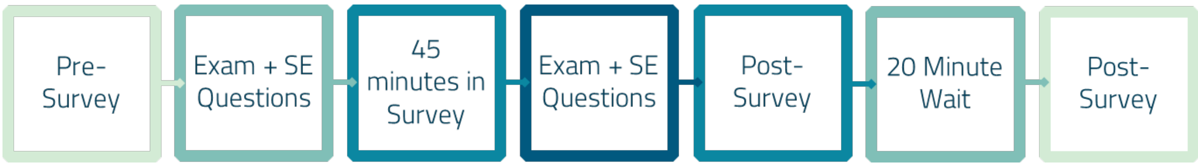

Fig. 1. Experimental timeline for the research study. Survey questions refer to a set of self-efficacy (SE) questions relating to student confidence on individual questions. These questions are not covered in this study.

Our research study focuses on the comparison of one set of students that took an exam with content that had been covered in the course (control group) against one set of students that took the same exam, but the students had not yet covered the content in class (experimental group).

Hypotheses

In this experiment, we hypothesized the following:

H1: Predicted performance by students across the exam will decrease with the magnitude of change being greater for the experimental group compared to the control group.

H2: Time spent on the exam for the experimental group will be significantly less than the control group.

H3: EDA arousal intensity will show lower arousal responses for the experimental group compared to the control group.

H4: Heart rate will be higher for the experimental group and will increase across the exam in comparison with the control group, which will have lower heart rates and will decrease over the exam

The rationale for these hypotheses stems from the findings stated in the theoretical framework. It is expected that students who are in the "stretch" or experimental scenario will face lower arousal because of the increased difficulty since the problems are not familiar, leading to inhibitory control [16]. The experimental group is expected to have a higher heart rate due to higher stress and anxiety resulting from a lack of familiarity with content [11-12]. More significant decreases in predicted performance will be evident in the experimental group since the actual performance is expected to be lower because of the lack of familiarity with the content. Less time will be spent on exam questions for the experimental group because they are not familiar with the content of the exam, though it is within the scope of the class.

Exam Context and Participants

For this study, an engineering statics course was chosen since it is historically the first engineering course that students across multiple engineering majors are required to take in colleges of engineering in the U.S. [18]. This study was conducted on a U.S. western institution of higher education with a higher than average non-traditional, first-generation, and rural student enrollment [19].

The statics course has three midterm exams and one final exam. This analysis focused on the content of the third midterm exam in Fall 2018 and Spring 2019, which was around week 12 of a 16-week semester and whose length and difficulty paralleled the actual exam. The practice exam contained 16 multiple-choice questions, which required some data analysis before selecting a response.

Students were recruited to take a practice exam one week before their actual exam for this research study from an engineering statics course in preparation for their actual exam. As an incentive, students were offered extra credit (decided by the instructor) and a $5 gift card (provided by the research team). Attending this study also fulfilled the class assignment requirement for completing the practice examination. This work-in-progress study shows a case of 13 participants from Fall 2018 who took Exam 3 after learning the content in class (control group) and 13 participants from Spring 2018 who took Exam 3 before learning the content in class (experimental group). Even though participant numbers are small, the amount of data collected and processed was very high, as explained in the sections below.

Experimental Setup

Students are situated at a station featured in *Fig. 2.* As shown in *Fig. 1*, the students take a series of surveys throughout their multiple-choice exam. Additional time is allotted on the exam to account for the time expended during survey taking. In the pre- and post-survey, students self-assess their performance as their predicted percentage of correct answers in the exam. Correct responses to each exam question were collected from the course instructor (who designed the practice and actual exams in the course), and performance data were collected from each student via our custom-created program. Students' perceived life stress was self-reported using the validated Perceived Stress Scale from Cohen, Kamarck, and Mermelstein [20].

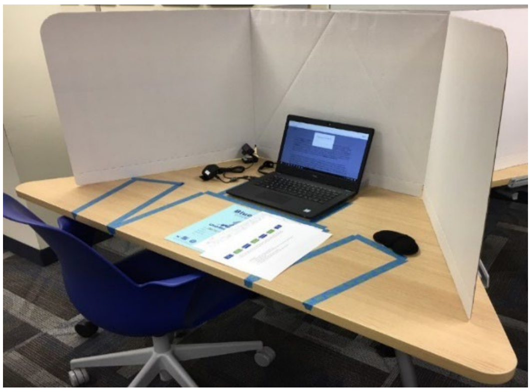

Fig. 2. Individual participant experimental setup for the research study.

Electrodermal Activity and Heart Rate Collection

During the entire duration of the exam period, EDA signals were collected from each participant at a rate of 4 Hz for ~2-hour exam (~28,800 data points per student) via an Empatica E4 wrist sensor (Empatica, Boston, MA). After collection, the EDA data must be cleaned, which was done according to a customized protocol [21-22] that filters out sources of noise in the data due to hand or body movement. To compare across the semesters, the EDA data (in the form of arousal intensity or the number of peaks) was normalized by time spent on each question of the exam. Heart rate was collected by the same sensor at 1Hz, which is slower compared to EDA data collection.

Ecological Validity

The ecological validity in this experiment was maintained by closely resembling the students' actual exam conditions. The students were provided with the same equation sheet, which was developed by the instructor, given in the actual exam. The workbooks were similar to what is offered in an actual exam. The test content was an electronic subset of practice test questions developed by the instructor, which paralleled the content and structure of the actual exam. The same amount of time was given for exam time (extra time was allotted for surveys). The actual exam that students take is also administered on a computer.

Statistical Analysis

To compare differences across the semesters, a paired one-tail t-test analysis was conducted for time, EDA, heart rate, and self-reported perceived stress across the semesters. To identify correlations between each variable, a Delta score was calculated for each parameter across the two semesters. This Delta score was then used to conduct a Pearson correlation analysis for each variable.

RESULTS

The time spent on the exam was compared between the fall (control group) and spring semester (experimental group). We found a significant decrease in time spent on the exam between the two-time points (t=4.36; $p<0.001$). For EDA, the number of peaks were normalized by time and then compared. There were no differences found between the experimental and control group (t=-0.38; $p=0.35$). For heart rate, no changes were found either (t=1.05; $p=0.15$). Interestingly, when comparing expected performance across the two groups, we found that between pre- and post- predictions, there was a ~14% (control group) and a ~29% (experimental group) self-reported reduction in their expected performances to the exam when comparing the pre-dicted and post-dicted conditions (t=12.49; p<0.001 and t=36.86; p<0.001, respectively). Finally, actual performances on the exams were compared between the control and experimental group. The actual performance for the experimental group was found to be nearly 30% lower than the control group (t=2.72; p<0.01).

Correlation analysis was conducted for the Delta scores between the control and experimental group in terms of actual performance and: (a) normalized EDA arousal intensity (number of phasic EDA peaks/time); (b) heart rate; and (c) perceived stress. No statistical significance was found between actual performance and EDA arousal intensity (r=0.13; p=0.068). We also did not find a statistical significance between heart rate and actual performance (r=0.05; p=0.49). Perceived stress and actual performance also did not show significant correlations (r=0.04; p=0.52).

In a similar fashion, pre- and post-dicted performance was examined for correlations to: (a) normalized EDA arousal intensity; (b) heart rate; and (c) perceived stress. For predicted performance, we found a moderate positive correlation to normalized EDA arousal intensity (r=0.36; p<0.001), heart rate (r=0.36; p<0.001), and perceived stress (r=0.35; p<0.001). For post-dicted performance, we found a moderate positive correlation to normalized EDA arousal intensity (r=0.36; p<0.001) and to perceived stress (r=0.30; p<0.001) but not to heart rate (r=0.02; p=0.77).

Upon closer examination of the variables, we also found a weak positive correlation between normalized EDA arousal intensity and heart rate (r=0.14; p=0.03) and a moderate positive correlation between normalized EDA arousal intensity and perceived stress (r=0.29; p<0.001). Actual performance when compared with both predicted performance (r=0.00; p=0.93) and post-dicted performance (r=0.04; p=0.55) did not have correlations with each other.

DISCUSSION

We found that there were overall decreases in both pre- and post-dicted performance and actual performance in both experimental and control groups with a larger magnitude of change in the experimental group suggesting that exposure and familiarity with the exam content do matter. This was also confirmed with decreased time spent on the exam by the experimental group due to unfamiliarity with the content. Literature suggests that if the cognitive load exceeds our processing capacity, individuals will struggle to complete the activity successfully [23]. It was interesting to find that while no correlations were found to actual performance, we did find correlations with pre- and post-dicted expected performances for EDA, heart rate, and perceived stress. This suggests potential anticipatory influences on physiological and psychological constructs of emotions and stress [24]. Also, we found that for post-dicted performances, there was a reduction in heart rate. This finding may suggest that reflective processes were at play, which may have influenced or reshifted cognitive attention on the task, leading to reduced heart rate, as suggested in other studies [15]. Moderate positive correlations were found between EDA and perceived stress. This points to a potential feedback role with reflective emotional processes among individuals, as suggested by the circumplex model of affect [25]. However, it is also possible that this may support the feedback mechanisms of the control-value theory [26] where individuals who may not have felt in full control of their exam experience may have some residual physiological and psychological responses lingering after the exam.

Collectively, the data suggest the importance of anticipatory and reflective processes during exam-taking, particularly as they relate to a students' well-being and behavior during these types of experiences. Overall, exam content that seemed to be a 'stretch' for these students did ignite these physiological responses, which may have been reflective of their sympathetic nervous system responses (e.g., fight or flight) [6].

CONCLUSIONS & IMPLICATIONS

The data suggest that when students are placed in situations where their knowledge is 'stretched' beyond domains that are attainable to them, especially in high-stakes situations like exams, reactive physiological and psychological responses surface. This, in turn, influences the magnitude of change over time suggesting that their reactive physiological responses are not supporting their actual performances, as evidenced by the 30% decrease in performance score for the experimental group.

Our findings point to the importance of understanding beforehand the implications of "stretching" students during an exam. Rather than responding to the "stretch" by trying harder, students put less effort into the exam. Over-stretching them may result in decreased performance and maladaptive psychological/physiological processes. If not handled appropriately, these decreases can lead to dire consequences (e.g., failure, drop-out). At the same time, identifying the proper zone by which to stretch students may potentially have a contrary effect such as positive performance outcomes and warrants more in-depth exploration. One intended future aim of this work will be exploring advanced modeling (such as those based on machine learning) to develop a student prediction model based on various physiological response data (e.g., EDA, Heart Rate) on a higher population of participants.

LIMITATIONS

This study is limited in its sample size, which limits statistical power. Also, this study was conducted in a laboratory environment that, while it was ecologically valid, is still not representative of the full, high-stakes environment that students would experience in a real exam experience. Also, EDA is influenced by other factors other than movement such as temperature and humidity [6]; additional work is needed to identify if these factors may have influenced some of the findings in this data. We did not explore additional outcomes (e.g., self-efficacy) that could have informed us more about how students coped with the two exam scenarios. Another limitation is that since students were not introduced to the exam concepts while in the classroom context, they may have been demotivated when continuing the exam. To fully understand the phenomenon we explored in this study, we would have to address different levels of content familiarity to understand this in more detail.

ACKNOWLEDGMENTS

This material is based upon work supported by National Science Foundation Grants No. 1661100 and No. 1661117 and the National Science Foundation Graduate Research Fellowship Program under Grant No. 120214. 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 the National Science Foundation.

Dr. Shawn Lampkins, Matthew Graham, and Cynthia Rigby are acknowledged for their contribution to data collection.

REFERENCES

[1] University of Waterloo (n.d.), Preparing Tests and Exams, *Centre for Teaching Excellence,* https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/developing-assignments/exams/exam-preparation

[2] Meyer, M. (2015), Persistence of Engineering Undergraduates at a Public Research University, *Utah State University Dissertation.*

[3] McLeod, S. (2019), What is the Zone of Proximal Development?, *Simply Psychology,*https://www.simplypsychology.org/Zone-of-Proximal-Development.html

[4] Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000), Test Prediction and Performance in a Classroom Context, *Journal of Educational Psychology,* Vol. 92, No. 1, pp. 160-170. doi: 10.1037//0022-0663.92.1.160

[5] Zhang, Y., Qin, F., Liu, B., Qi, X., Zhao, Y., & Zhang, D. (2018), Wearable Neurophysiological Recordings in Middle-School Classroom Correlate With Students' Academic Performance, *Frontiers in Human Neuroscience,* Vol. 12, No. 457, pp. 1-8. doi: 10.3389/fnhum.2018.00457

[6] Boucsein, W. (2012), *Electrodermal Activity* (2nd ed.), New York: Springer Science+Business Media, LLC.

[7] Hammond, G. R. & Jordan, P. M. (1984), Bilateral Electrodermal Activity During Performance of Cognitive Tasks of Varying Difficulty Levels, *Physiological Psychology,* Vol. 11, No. 4, pp. 256-260. doi: 10.3758/BF03326804

[8] Administration for Children and Families (n.d.), *National Survey of Child and Adolescent Well-Being* (Methods Brief No. 1). Retrieved from https://www.acf.hhs.gov/sites/default/files/opre/inhibitory_control.pdf

[9] Badami, R., VaezMousavi, M., Wulf, G., & Namazizadeh, M. (2012), Feedback about more accurate versus less accurate trials, *Research Quarterly for Exercise and Sport,* Vol. 83, No. 2, pp. 196-203. doi: 10.1080/02701367.2012.10599850

[10] Kramer, D. (2007), Predictions of Performance by EEG and Skin Conductance, *Indiana Undergraduate Journal of Cognitive Science,* Vol. 2, pp. 3-13.

[11] Elwess, N. L. and Vogt, D. (2005), Heart Rate and Stress in a College Setting, *Bioscene: Journal of College Biology Teaching,* Vol. 31, No. 4, pp. 20-23.

[12] Zhang, Z., Hai, S., Qiang, P., Qing, Y., and Cheng, X. (2011), Exam Anxiety Induces Significant Blood Pressure and Heart Rate Increase in College Students, *Clinical & Experimental Hypertension,* Vol. 33, No. 5, pp. 281-286.

[13] Daly, A. L., Chamberlain, S., and Spalding, V. (2011), Test Anxiety, Heart Rate and Performance in A-level French Speaking Mock Exams: an Exploratory Study, *Educational Research,* Vol. 53, No. 3, pp. 321-330. doi: 10.1080/00131881.2011.598660

[14] Fairclough, S. H., Tattersall, A. J., and Houston, K. (2006), Anxiety and Performance in the British Driving Test, *Trasportation Research Part F,* Vol. 9, pp. 43-52. doi: 10.1016/j.trf.2005.08.004

[15] Phillips, A. C. (2011), Blunted cardiovascular reactivity relates to depression, obesity, and self-reported health. *Biological psychology*, Vol. 86, No. 2, pp. 106–113.

[16] Sanders, A. E. and Lushington, K. (2001), Effect of Perceived Stress on Student Performance in Dental School, *Journal of Dental Education,* Vol. 66, No. 1, pp. 75-81.

[17] Rafidah, K., Azizah, A., Norzaidi, M. D., Chong, S. C., Salwani, M. I., & Noraini, I. (2009), The Impact of Perceived Stress and Stress Factors on Academic Performance of Pre-Diploma Science Students: A Malaysian Study, *International Journal of Scientific Research in Education,* Vol. 2, No. 1, pp. 13-26.

[18] STEM Course Prep (n.d.), Engineering Mechanics – Statics, https://stemcourseprep.com/p/engineering-mechanics-statics

[19] Office of Analysis, Assessment, and Accreditation (2018), Utah State University, Accessed 23 July 2018.

[20] Cohen, S., Kamarck, T., and Mermelstein, R. (1983), A Global Measure of Perceived Stress, *Journal of Health and Social Behavior,* Vol. 24, pp. 385-396.

[21] Khan, T. H., Villanueva, I., Vicioso, P., and Husman, J. (2019), Exploring Relationships Between Electrodermal Activity, Skin Temperature, and Performance During Engineering Exams, *2019 IEEE Frontiers in Education Conference (FIE),* October 16-19, 2019, in review.

[22] Villanueva, I., Husman, J., Youmans, K., Christensen, D., Vicioso, P., Khan, M.T.H., Lampkins, S., and Graham, M. (2019). A Cross-Disciplinary and Multi-Modal Experimental Design for Studying Near-Real-Time Authentic Examination Experiences. *Journal of Visualized Experiments*, 2019, in review.

[23] Chandler, P. and Sweller, J. (1991), Cognitive Load Theory and the Format of Instruction, *Cognition and Instruction,* Vol. 8, No. 4, pp. 293–332.

[24] Husman, J., Cheng, K. C., Puruhito, K., and Fishman, E. J. (2015), Understanding Engineering Students' Stress and Emotions During an Introductory Engineering Course, Paper presented at the 122nd American Society of Engineering Education Annual Conference & Exposition, Seattle, Washington.

[25] Boucsein, W. and Backs, R. W. (2009), The Psychophysiology of Emotion, Arousal, and Personality: Methods and Models, in Duffy, V. G. (Ed.), *Handbook of Digital Human Modeling: Research for Applied Ergonomics and Human Factors Engineering* (pp. 35-1-35-18). Boca Raton, FL: CRC Press.

[26] Pekrun, R., and Linnenbrink-Garcia, L. (2014), Emotions in education:Conclusions and future directions. In *International Handbook of Emotions in Education.* First Edition. Taylor Francis: New York, NY, USA.