Relatable and Humorous Videos Reduce Hyperarousal in Math Exams

Fettah Kiran, Amanveer Wesley
Dept, of Computer Science
Univ. of Houston, Texas
{fkiran, awesley4}@uh.edu

Tammy Tolar, Paul Cirino
TIMES & Dept. of Psych.
Univ. of Houston, Texas
{ttolar, pcirino}@uh.edu

Panagiotis Tsiamyrtzis
Dept. of Mech. Eng. & Stat.
POLIMI, Italy & AUEB, Greece
pt@aueb.gr

Ioannis Pavlidis
Dept. of Computer Science
Univ. of Houston, Texas
ipavlidis@uh.edu

***Abstract*—We report on a study investigating the sympathetic and performance effects of relatable humorous videos in undergraduate math exams. We recruited 20 lower division students to test this novel form of questioning. The students took a foundational math exam that included 12 items from each of the following three categories: Abstract A, Word W, and Video V, where A featured formula-based questions and W analytic questions expressed in plain descriptive form. The V category had questions similar to the W category, but expressed in relatable humorous video form. Sympathetic arousal was measured through facial electrodermal activity (EDA_f) and heart rate (HR), where the former was extracted via thermal imaging, and the latter through smartwatches. Results from both the EDA_f and HR channels indicate that questions expressed in relatable humorous video form significantly curtail hyperarousal with respect to similar questions expressed in plain descriptive form. Furthermore, the study's results suggest that exam performance is negatively affected by pre-exam anxiety, while is positively affected by generous time allotment. The said findings highlight the potential of V questions in making the math experience less stressful and more endearing to undergraduate students. Due to the importance of foundational math courses, such a change stands to bring downstream benefits to STEM education.**

***Index Terms*—math exams, math anxiety, sympathetic arousal, math performance, affective computing, multimodal data, thermal imaging, perinasal perspiration, EDA, heart rate**

I. Introduction

A healthy supply of scientists, engineers, and technologists is increasingly critical to the world economy, fueling the need for attracting and training students in STEM fields. [1] Critical to the sustenance of such a trend is the acquisition of basic math skills. Prior studies have found that students with strong elementary math skills perform significantly better in applied contexts during their college education. [2] A strong body of evidence also suggests that strong math skills positively affect labor-market outcomes in a modern economy. [3]

Math anxiety, which is defined as a negative reaction to all math related matters, is a key impediment to the acquisition of mathematical skills. It is a major factor in 'global avoidance', that is, the tendency of math-anxious individuals to avoid elective math-oriented coursework in secondary and post-secondary education. As a consequence, math-anxious individuals avoid educational tracks and career avenues that depend on math, with downstream implications for the size and quality of the STEM workforce. [4]

Math anxiety is especially acute in exams. Roos et al. found that in 21 out of 31 recently conducted studies there was a significant positive relationship between self-reported exam anxiety and physiological arousal [5]. Hence, exam anxiety bears psychophysiological effects. One such effect is electrodermal activity (EDA), which Khan et al. found that has a positive correlation to exam question difficulty. [6] Furthermore, Qu et al. reported positive correlation between heart rate and math anxiety during a 90-minute real math exam. [7]

Math anxiety is not only hard to the psyche and the body but also has negative implications to educational performance. A meta-study conducted by Zhang et al. found that 84 studies from 2000 to 2019, powered by $N = 8,680$ subjects in total, indicated a negative correlation between math-anxiety and performance. [8] The only factor capable of counterbalancing the negative performance effect of math anxiety is the existence of strong math ability, [9] pointing to the following conundrum: How can people acquire mathematical skills if they are afflicted by math anxiety, which is a barrier to math competency in the first place?

Emotion regulation (ER) methods may reduce the negative relationship between math anxiety and math performance. For instance, researchers found that cognitive reappraisal weakened the correlation between physiological arousal and math accuracy, such that even high physiological arousal levels no longer had a negative association with math accuracy. [10]

In addition to psychological interventions, such as cognitive reappraisal, math anxiety can also be countered with educational methods. In this direction, Gaidi et al. introduced contextualized teaching of mathematics through everyday examples, to which students could relate from their life experiences. This approach was antithetical to the conventional abstract teaching of mathematics in Sweden, which the investigators thought was the source of poor educational results. Their experiment supported the validity of their hypothesis, demonstrating the potential of math instruction centered on everyday problems. [11]

Another novel and under-explored method to counter math anxiety is humor. Humor can function as an adaptive coping mechanism. Specifically, using humor in a testing situation produces a cognitive shift in perspective that allows students to distance themselves from the immediate threat, that is, the

979-8-3503-2745-8/23/$31.00 ©2023 IEEE

exam. Humor produces an emotional response that decreases stress hormones such as serum cortisol and epinephrine. To use humor effectively in exams, it is suggested to allow extra time and employ culturally relevant themes and styles. [12] In fact, the dearth of studies about the effectiveness of humor in exams, leaves plenty of room for imaginative interventions.

A. Research Questions

We aim to ameliorate stress responses in mathematical testing by introducing questions in video form, where actors express the problem in a relatable daily situation with a dose of humor. We focus on math exams in lower division courses, which are critical to student retention and career orientation. Should the manipulation be successful, it stands to curtail 'global avoidance', increasing the flows towards STEM disciplines. We also wonder if reduced stress responses in math exams are accompanied by increases in performance. Accordingly, we ask the following research questions:

- **RQ1:** Do students tackling math questions in relatable humorous video form exhibit lower sympathetic arousal with respect to similar math questions in conventional form?
- **RQ2:** Do students tackling math questions in relatable humorous video form perform better with respect to similar math questions in conventional form?

B. Contributions

To address research questions RQ1 and RQ2 we conducted an experiment, where we monitored and analyzed the performance, physiological, and psychometric state of $n = 20$ lower division students as they took a foundational math exam with questions in both relatable humorous video and conventional forms. Our research makes the following contributions:

1) It brings to the fore the sympathetically moderating role of math questions in relatable humorous video form.
2) It demonstrates the agreement between facial electrodermal activity (EDA_f) and heart rate (HR) in determining arousal caused by mathematical stressors.
3) It shows the negative effect anxiety can have on exam performance.
4) It makes public a unique dataset that would feed further research on the subject [https://osf.io/qx6et/].

II. Study Design

We recruited lower division students ($n = 20$) per an institutionally approved protocol. Advertisement was carried out through class announcements. After consenting, the participants went through the experiment, which featured three sessions: 1) baselining; 2) psychometric questionnaire; 3) foundational math exam. In the baseline and math exam sessions there was continuous recording of facial electrodermal activity (EDA_f) and heart rate (HR) - Fig. 1.

The purpose of the baseline session was to establish intra-individual physiological references for the EDA_f and HR measurements [13]. The tonic level of sympathetic activation is characterized by significant inter-individual differences. This implies that in the absence of stimuli, some people have higher background arousal than others. As a result, analysts may misinterpret experimental measurements of participants with naturally high baseline as indicators of strong sympathetic activation. For this reason, during stimulation, which in our case is the math exam, absolute physiological measurements are not as informative as normalized physiological measurements (i.e., differences from the participants' baseline).

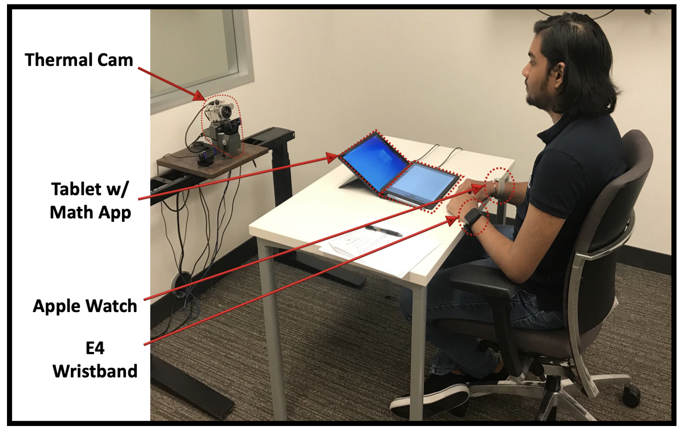

Fig. 1. **Experimental setup.** Shown is a participant about to take the foundational math exam.

To obtain baseline measurements, we asked participants to close their eyes and envision a relaxing nature scene for five minutes while sitting alone in a dimly lit room. The absence of external stimuli and the focus on a relaxing internal stimulus meant to bring participants' sympathetic state close to their tonic level or at the very least to an equitable reference level.

After completing the baseline session, participants filled out the State and Trait Anxiety Inventory (STAI) Form Y-1, to obtain their pre-exam anxiety score. Then, they took the math exam, which was delivered via a tablet app (Fig. 1). There was no strict time limit for the exam, which included 36 questions, grouped into three categories with 12 matching questions each: Abstract (A), Word (W), and Video (V). The category order was randomized to ameliorate practice effects. W questions involved a plain description of a basic analytic problem; for instance, '*If a student loan had a yearly interest rate of 2%, how much interest would be charged on $1000 at the end of one year?*' The matching V question for this sample W question was expressed through a video, where a student after admitting to his friend through self-deprecating humor that he lost money on a game bet, he was asking how to compute compound interest, now that he needs to resave for his planned dream trip. The matching A question, for the said W and V pair, involved an interest calculation formula the result of which had to be filled out by the student: $(\frac{25}{100}) \cdot 1000 = ?$.

A. Study Variables and Justification

Psychometric Score. Anxiety is intimately linked to sympathetic activation [14] and thus of interest to our investigation. We measured the pre-exam anxiety score of participants through the STAI Form Y-1. [15] STAI scores take values in the range $[20, 80]$, with higher scores indicating higher anxiety.

Heart Rate Signals. For sitting individuals, subjected to a well-known stressor, such as math testing, heart rate is largely associated with sympathetic activation. To ensure validity, we recorded participants' heart rate with two wearable sensors - E4 and Apple Watch worn on the non-dominant and dominant hand, respectively (Fig. 1). We paired the time registered heart rate measurements from the two sensors every 5 s, producing a regression plot with 16779 points. Using Cook's Distance, we identified 838 pairs that were not in high agreement and removed them from consideration as unreliable, leaving 15941 solid HR measurements.

Facial EDA Signals. A physiological variable that is exclusively associated with sympathetic activation is electrodermal activity (EDA). EDA is typically measured in the palms (EDA_p). However, the palm location entails poor usability, especially for participants taking exams, with their hands busy swiping, tapping, and writing. A more user friendly alternative is EDA measurements in the wrist through a wearable sensor. This solution has its own problems, as a portion of the population has weak or non-existent EDA responses in the wrist [16]. For this reason, we opted to measure EDA in the face (EDA_f) through thermal imaging. EDA_f quantifies perspiration in the perinasal region, which has been shown to commensurate with EDA_p - the gold standard. [17]

To ameliorate the effect of head motion in the EDA_f signal, we used the facial tissue tracker reported by Zhou et al. [18] Zhou's algorithm tracks the participant's perinasal region from frame to frame via spatiotemporal smoothing. In the said region, sympathetically activated sweat glands appear as 'cold' (dark) spots, amidst 'hot' surrounding tissue - a phenomenon quantified by a morphology-based algorithm reported by Shastri et al. [17], yielding the arousal signal.

Question Grade (QG). Depending on the participant's answer to each question, the math app records a grade, which is 1 for correct answer and 0 for incorrect answer. This binary question grade is the key measure of performance in the exam.

Question Time (QT). The math app presents the questions to the participants sequentially. Participants have to thoroughly negotiate each question before they move to the next. The math app records the time participants spent solving each question.

III. Results

To address research question RQ1, we construct a family of multiple linear regression models that feature as response variable the mean normalized sympathetic arousal of participant $\mathcal{P}_i$ in question $\mathcal{Q}_j$. Sympathetic arousal can be quantified either through facial EDA or heart rate measurements, with the corresponding sympathetic models shown in Eq. (1) and Eq. (2). In Eq. (1), the response variable $\overline{\Delta \ln EDA_f}(i,j)$ is the logarithmically corrected mean normalized perinasal perspiration of participant $\mathcal{P}_i$ in question $\mathcal{Q}_j$. In Eq. (2), the response variable $\overline{\Delta HR}(i,j)$ is the mean normalized heart rate of participant $\mathcal{P}_i$ in question $\mathcal{Q}_j$. Normalization is achieved by subtracting from each momentary physiological value during the exam, the mean baseline value of the participant.

$$\begin{aligned}\overline{\Delta \ln EDA_f}(i,j) \sim \beta_0 + &\underline{\beta_1 SA(i) + \beta_2 QC(j)} + \\ &\underline{\beta_3 QG(i,j) + \beta_4 QT(i,j)} + (1|\mathcal{P}_i)\end{aligned} \tag{1}$$

$$\begin{aligned}\overline{\Delta HR}(i,j) \sim b_0 + &\underline{b_1 SA(i) + b_2 QC(i,j)} + \\ &\underline{b_3 QG(i,j) + b_4 QT(i,j)} + (1|\mathcal{P}_i)\end{aligned} \tag{2}$$

The first line in the sympathetic arousal models (1) and (2) – underlined in green – holds information about the participants' state of anxiety and the category of exam questions. In more detail, $SA(i)$ stands for the STAI Form Y-1 score of participant $\mathcal{P}_i$ just before the start of the exam. $QC(i,j)$ stands for the category of question $\mathcal{Q}_j$ negotiated by participant $\mathcal{P}_i$, with levels A, W, and V. The second line in the sympathetic arousal models – underlined in magenta – holds the participants' exam performance information. Specifically, $QG(i,j)$ is the grade participant $\mathcal{P}_i$ received in question $\mathcal{Q}_j$ - 0 for incorrect and 1 for correct answer. $QT(i,j)$ is the time in seconds participant $\mathcal{P}_i$ devoted to question $\mathcal{Q}_j$. The term $(1|\mathcal{P}_i)$ indicates that we take into account participant-centered random effects.

To address research question RQ2, we construct a logistic regression model in Eq. (3) that features as response variable the odds $P(QG(i,j))$ of participant $\mathcal{P}_i$ successfully answering question $\mathcal{Q}_j$.

$$\begin{aligned}\text{logit}(P(QG(i,j))) \sim \; &\gamma_0 + \underline{\gamma_1 SA(i) + \gamma_2 QC(i,j)} + \\ &\underline{\gamma_3 \overline{\Delta \ln EDA_f}(i,j) + \gamma_4 \overline{\Delta HR}(i,j)} + \\ &\gamma_5 QT(i,j) + (1|\mathcal{Q}_j).\end{aligned} \tag{3}$$

The first line of the exam performance model – underlined in green – holds information about the participants' state of anxiety $SA(i)$ and the category of exam questions $QC(i,j)$. The second line of the exam performance model – underlined in magenta – holds the participants' sympathetic activation information. Specifically, $\overline{\Delta \ln EDA_f}(i,j)$ and $\overline{\Delta HR}(i,j)$ express normalized arousal manifested via facial EDA and heart rate, respectively. The last line of the model features the time $QT(i,j)$ it takes participant $\mathcal{P}_i$ to address question $\mathcal{Q}_j$. The term $(1|\mathcal{Q}_j)$ indicates that we take into account question-centered random effects.

Table I provides the parameter estimates for the significant predictors of the facial EDA model in Eq. (1), heart rate model in Eq. (2), and exam performance model in Eq. (3). Figure 2 shows the plots of the models' results.

IV. Conclusion

The results suggest that: 1) W questions are associated with significantly higher arousal than V questions. 2) Pre-exam anxiety and time devoted to solving a question are significant predictors of exam performance, with lower anxiety and longer solving times being associated with higher odds of successful answers. Hence, casting math exams in relatable humorous video form (i.e., giving V instead of W questions) would

TABLE I
RESULTS FOR THE SIGNIFICANT PREDICTORS FEATURED IN THE FACIAL EDA, HEART RATE, AND EXAM PERFORMANCE MODELS. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

FACIAL EDA	Estim.	Std Error	DOF	t-value	Pr(>|t|)
Intercept	0.162	0.039	22.225	4.195	< 0.001***
QC_A	0.013	0.015	689.036	0.874	0.382
QC_W	0.087	0.015	689.045	5.710	< 0.001***
HEART RATE	**Estim.**	**Std Error**	**DOF**	**t-value**	**Pr(>|t|)**
Intercept	-3.255	1.360	21.181	-2.394	0.026*
QC_A	0.498	0.400	651.032	1.243	0.214
QC_W	1.590	0.410	651.121	3.879	< 0.001***
EXAM PERFORM.	**Estim.**	**Std Error**		**z-value**	**Pr(>|t|)**
Intercept	0.955	0.286		3.334	< 0.001***
SA	-0.416	0.096		-4.357	< 0.001***
QT	0.620	0.123		5.036	< 0.001***

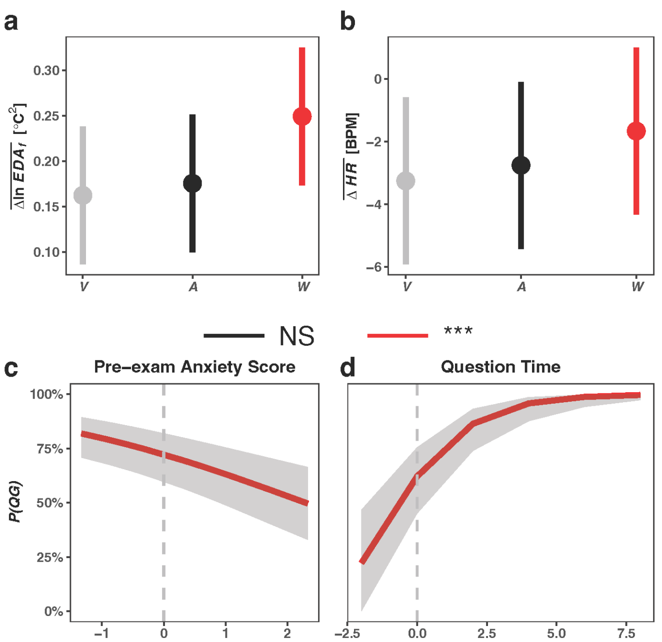

Fig. 2. **Association of question categories with: a.** facial EDA; **b.** heart rate. V questions were selected as base levels and colored in gray. Black color indicates No Significance (NS: $p \geq 0.05$) while red indicates High Significance (***: $p < 0.001$). **Association of the odds $P(QG)$ of correctly answering a question with: c.** participant's pre-exam anxiety; **d.** the time spent on the question. The variables are scaled.

render them less stressful without affecting performance. In this context, it is important to note that humor was not a central part of the initial video design. The focus was on the production of culturally relatable content. It was this focus, however, that naturally led to humorous overtones. It is also not clear what would have happened if the video scripts were delivered as write-ups instead of acted plays. Consequently, we measured the compound effect of relatedness, humor, and the artistic medium. We are currently working on a larger study to see if these promising pilot results would scale up.

ETHICAL IMPACT STATEMENT

We caution about the use of methodological shortcuts for expediency. They may lead to arousal misestimation with downstream implications for exam fairness.

Acknowledgment: NSF Award # 1760760.

REFERENCES

[1] D. I. Spang and K. Spang, "Real-world applications of mathematical and scientific principles in the curriculum for college and career success," in *2012 ASEE Annual Conference & Exposition*, 2012, pp. 25.1100.1–25.1100.13.

[2] A. A. Gallo and C. K. Johnson, "Mathematics skills and everyday problem solving," *Journal of Economics and Finance Education*, vol. 7, no. 1, pp. 7–20, 2008.

[3] C. Koedel and E. Tyhurst, "Math skills and labor-market outcomes: Evidence from a resume-based field experiment," *Economics of Education Review*, vol. 31, no. 1, pp. 131–140, 2012.

[4] M. H. Ashcraft and K. S. Ridley, "Math anxiety and its cognitive consequences," in *Handbook of Mathematical Cognition*, J. Campbell, Ed. Psychology Press, 2005, ch. 18, pp. 315–327.

[5] A.-L. Roos, T. Goetz, M. Voracek, M. Krannich, M. Bieg, A. Jarrell, and R. Pekrun, "Test anxiety and physiological arousal: A systematic review and meta-analysis," *Educational Psychology Review*, vol. 33, pp. 579–618, 2021.

[6] T. H. Khan, I. Villanueva, P. Vicioso, and J. Husman, "Exploring relationships between electrodermal activity, skin temperature, and performance during," in *2019 IEEE Frontiers in Education Conference (FIE)*, 2019, pp. 1–5.

[7] Z. Qu, J. Chen, B. Li, J. Tan, D. Zhang, and Y. Zhang, "Measurement of high-school students' trait math anxiety using neurophysiological recordings during math exam," *IEEE Access*, vol. 8, pp. 57 460–57 471, 2020.

[8] J. Zhang, N. Zhao, and Q. P. Kong, "The relationship between math anxiety and math performance: A meta-analytic investigation," *Frontiers in Psychology*, vol. 10, p. 1613, 2019.

[9] K. H. Dew, J. P. Galassi, and M. D. Galassi, "Math anxiety: Relation with situational test anxiety, performance, physiological arousal, and math avoidance behavior," *Journal of Counseling Psychology*, vol. 31, no. 4, p. 580, 1984.

[10] R. G. Pizzie and D. J. Kraemer, "The association between emotion regulation, physiological arousal, and performance in math anxiety," *Frontiers in Psychology*, vol. 12, p. 639448, 2021.

[11] K. El Gaidi and T. Ekholm, "Contextualizing calculus with everyday examples to enhance conceptual learning," in *2015 ASEE Annual Conference & Exposition*, 2015, pp. 26–401.

[12] R. A. Berk, "Does humor in course tests reduce anxiety and improve performance?" *College Teaching*, vol. 48, no. 4, pp. 151–158, 2000.

[13] S. Zaman, A. Wesley, D. R. D. C. Silva, P. Buddharaju, F. Akbar, G. Gao, G. Mark, R. Gutierrez-Osuna, and I. Pavlidis, "Stress and productivity patterns of interrupted, synergistic, and antagonistic office activities," *Scientific ata*, vol. 6, no. 1, p. 264, 2019.

[14] M. M. Wenner, "Sympathetic activation in chronic anxiety: not just at the "height" of stress. Editorial Focus on "Relative burst amplitude of muscle sympathetic nerve activity is an indicator of altered sympathetic outflow in chronic anxiety"," *Journal of Neurophysiology*, vol. 120, no. 1, pp. 7–8, 2018.

[15] C. D. Spielberger, "State-trait anxiety inventory," in *The Corsini Encyclopedia of Psychology*. Hoboken, NJ: Wiley Online Library, 2010. [Online]. Available: https://doi.org/10.1002/9780470479216.corpsy0943

[16] P. Tsiamyrtzis, M. Dcosta, D. Shastri, E. Prasad, and I. Pavlidis, "Delineating the operational envelope of mobile and conventional eda sensing on key body locations," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 5665–5674.

[17] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. Pavlidis, "Perinasal imaging of physiological stress and its affective potential," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 366–378, 2012.

[18] Y. Zhou, P. Tsiamyrtzis, P. Lindner, I. Timofeyev, and I. Pavlidis, "Spatiotemporal smoothing as a basis for facial tissue tracking in thermal imaging," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 5, pp. 1280–1289, 2012.