**Online Supplementary Materials for**

**Investigating Cortisol in a STEM** **Classroom:**

**The Association Between Cortisol and Academic Performance**

**S1. Information About**

**Introduction to Molecular and Cellular Biology Course**

Informal focus groups conducted before the study revealed that majority of prior students who had taken this class indicated that it was the most stressful class of their major, making it particularly suitable to investigate cortisol responses. This class was perceived as both difficult and important. The baseline survey (assessed prior to the class starting or in the first weeks of the semester) had two questions reported in the main text about participants’ rating of the difficulty and importance of the course. Participants were asked how difficult they expected the course to be, with five response options from very easy to very difficult. They were also asked their agreement with the statement, “It is important to me to do well in this course,” with five response options from 1 (strongly disagree) to 5 (strongly agree). Most participants (51.3%) reported expecting the class to be very difficult, and the vast majority (86.7%) strongly agreed that it was important for them to do well.

# S2. Information About Participating and Non-Participating Students

Among the 552 students who were enrolled in this class, 328 (59%) consented to participate. Consenting participants were asked to sign transcript release forms and most (95%) agreed. As we did not have any demographic information on the 224 non-consenting students, we could not identify differences between consenting and non-consenting students. Of the 328 students who consented, 53 did not submit any saliva samples and thus, are not included in the results presented here. Of the 53 students without saliva samples, 28 remained in the class but also did not complete any other study activities, 13 remained in the class and participated in other study activities, and 12 dropped the class (7 without taking any exams or participating in any study activities; 5 who took 1 – 2 of 4 exams and partially completed other study activities).

In the main text, we note that we excluded two participants because their average cortisol levels were more than three standard deviations above the mean. Including these two participants does not alter the patterns of 1) within- and between-class cortisol trajectories or 2) the association of cortisol and race/ethnicity on performance.

# S3. Information About Participant Characteristics

Of the 271 participants, most (79%) submitted at least 75% of the requested saliva samples (i.e., ≥ 6 out of 8 samples) and most (87.5%) completed all four exams. Consistent with typical demographic characteristics of biology-focused STEM classes (Sax et al., 2018), women outnumbered men by about 2:1. The racial composition of the course is consistent with past research suggesting the underrepresentation of Hispanic, Black and Native American students in college biology classes (Jordt et al., 2017).

# S4. Coding Race/Ethnicity for Students who Reported Their Race/Ethnicity as Other

Four students had ambiguous or missing information as to how they would describe their race/ethnicity and selected a race/ethnicity of “other.” We categorized one student, who wrote in a race/ethnicity of “Trinidadian,” as URM based on our understanding of the racial and cultural history of Trinidad. The other three participants did not provide any demographic information and were identified through phenotypical traits as non-URM students.

**S5. Participant Recruitment**

We emailed all enrolled students a study description and a link to begin participation before classes started and recruited in-person at the start of the semester. Upon opening the link, students were presented with a consent form and a set of baseline measures (see https://bit.ly/3EB2v5i for all measures). During the first week of the semester, researchers distributed consent forms in class for students interested in participating who had not yet done so, who were then sent another email with a link to the baseline measures.

**S6. Accounting for Time Since Awakening on Cortisol Levels**

We reasoned that time-since-awakening would be a more accurate covariate of cortisol than class time because it tracks how long participants had been awake independent of class time. However, we also tested class time as a covariate. Adding class time as a non-interacting covariate to models predicting cortisol did not change the pattern or statistical significance of any reported results.

Adding (1) the class time x time within-class interaction and (2) the class time x time between-classes interactions to the models predicting cortisol revealed a significant interaction for (1) but not (2). Simple effects revealed that cortisol declined more steeply within morning classes (b = -.20, *p* < .001) than within afternoon classes (b = -.15, *p* < .001), but the pattern of significant within-class decline remained for both. Specifying class time as a nesting factor in longitudinal analyses also did not change the pattern of results. Furthermore, adding class time as a covariate to models predicting performance or as a moderator of cortisol or race/ethnicity on performance did not change the pattern of results.

In addition, we also tested whether students’ within-class cortisol patterns changed across the semester. Indeed, there was a significant interaction between within-class cortisol patterns and timing of the class session in the semester, b = 0.03, *p* = 0.01. Follow-up tests revealed that the within-class decline in cortisol was more pronounced in later classes (Saliva Class 2, b = -0.14; Saliva Class 3, b = -.20; Saliva Class 4, b = -0.14) relative to Saliva Class 1 (b = -0.11). Notably, however, the decline within class remained significant in every class. Students may have become more comfortable in and accustomed to the classroom over the semester, leading to a greater reduction in cortisol from the beginning to end of class as the semester progressed. It is also possible that students’ sleeping cycles, which can impact students’ diurnal patterns, may vary across the semester (e.g., sleeping later in the beginning of the semester). However, students’ time since awakening at the saliva collection (which may be closely associated with students’ diurnal pattern and wake-up time) did not change across the semester. Overall, cortisol declined within classes in all classes we assessed.

# S7. Information About Saliva Sample Collection Procedure

On the day before each saliva sample collection, the research team emailed participants and asked them to refrain from activities that can influence cortisol in the hours before samples were collected.

On the day of the collection, once participants received plastic bags that contained saliva collection materials, they were asked to write their school ID on a label attached to the bag with the saliva collection materials. At the beginning of the first class in which saliva samples were collected, the researchers made an announcement to explain the saliva collection procedures. When finished, participants were asked to return the tubes and questionnaire to the plastic bag. After class, the researchers collected the bags. Saliva samples were then labeled with participants’ de-identified study ID number and stored at or below -25°C.

# S8. Longitudinal Cortisol Pattern Across Classes Moderated by Intervention Condition

The larger project included a randomly-assigned affirmation intervention or control exercise delivered in Week 3 (Cohen et al., 2006). Because of this, many of our pre-registered predictions (see <https://bit.ly/3EB2v5i>) were centered around the effect of the intervention and those in the current manuscript are based on the typical situation in the absence of intervention.

We did not find the pre-registered effect of the intervention on cortisol levels or trajectories. Overall, the intervention had a small, difficult-to-interpret effect on longitudinal cortisol patterns. We followed the model building approach described in the main text to test whether students’ cortisol patterns across classes were moderated by the intervention condition (see https://bit.ly/3EB2v5i for intervention materials). Among our sample of 271 students, 27 were excluded because they did not complete the intervention activities. Thus, these analyses are based on 244 students, who submitted 1,601 saliva samples. In addition to the final model noted in the main text, intervention condition was included as a predictor of the intercept and time slopes. Also, to further test whether any effect of intervention condition varied by race/ethnicity, we included at Level 3 the interaction between race/ethnicity and intervention condition as a predictor of the intercept and time slopes. As in the main text, we started with a linear model at Level 2 to represent the four measurement occasions across classes, including all upper-level predictors. Then, we tested whether adding quadratic and cubic terms (with all upper-level predictors) improved model fit. As in the main text, results revealed that a model with linear, quadratic, and cubic terms fit better than either a linear-only model, χ²(12, *N* = 244) = 36.07, *p* < .001, or a linear-plus-quadratic model, χ²(6, *N* = 244) = 25.27, *p* < .001. Consequently, we retained the linear, quadratic, and cubic predictors. However, we note that in the linear-only and linear-plus-quadratic models, neither the time slope(s) nor the intercept was moderated by race/ethnicity, intervention condition, or their interaction.

Results revealed little evidence that cortisol levels were affected by the intervention. In fact, the only evidence of any intervention effect on cortisol patterns was a marginal interaction between race/ethnicity and intervention condition on the cubic cortisol trajectory across classes, = -0.02, *t*(541) = -1.93, *p* = .054, 95% CI [-0.04, 0.0001]. Simple effects tests revealed that among URM students, there was a downward cubic trajectory of cortisol, = -0.04, *t*(541) = -2.22, *p* = .03, 95% CI [-0.07, -0.004], that was not moderated by intervention condition, = -0.01, *t*(541) = -0.39, *p* = .70, 95% CI [-0.04, 0.02]. Among non-URM students, there was also a downward cubic trajectory of cortisol, = -0.03, *t*(541) = -3.16, *p* = .002, 95% CI [-0.05, -0.01], but this trajectory was moderated by intervention condition, = 0.03, *t*(541) = 2.93, *p* = .004, 95% CI [0.01, 0.05].

Among non-URM students, those in the intervention condition did not show evidence of a cubic pattern in cortisol levels, = -0.004, *t*(541) = -0.29, *p* = .77, 95% CI [-0.03, 0.02], whereas those in the control condition evidenced a downward cubic slope, = -0.06, *t*(541) = -4.12, *p* < .001, 95% CI [-0.09, -0.03]. Thus, there was a downward cubic trajectory in cortisol, on average, for everyone except for non-URM students in the intervention condition. Among non-URM students, although they did not have a significant cubic slope, they still had a downward cortisol trajectory across time with less fluctuation. In part, these results suggest that those in the intervention condition may have had less fluctuation in cortisol across classes compared to the control group or to URM students. However, it is unclear what this result may represent, and as this pattern was small in magnitude and not hypothesized, we do not interpret it further.

**S9. Information about Smartphone Studies**

Although a full description is beyond the scope of this article, the smartphone study procedures involved a set of questions each morning and evening, as well as after each class for 21 days that partially overlapped with the timing of the saliva sample collection. The smartphone survey included a daily question about stress, which we report in Figure S1. Detailed procedures and full measures are available on OSF (https://bit.ly/3EB2v5i). Participants who indicated in the baseline measures that they were interested in the smartphone part of the study were invited to participate (approximately 90% were interested). We oversampled URM students in this sub-sample.

**S10. Cortisol Assay Procedure**

Ninety percent of samples were analyzed in singlets and 10% (selected randomly) were analyzed in duplicate (one sample tested twice) for assessing intra-assay coefficients. Duplicate cortisol results were averaged, and averaged values were used in all analyses.

**S11.** **Information About Final Class Score and Exam 2**

**Final Score**

In addition to retaining students who did not get a final grade, we use final score (instead of final letter grade) because, as a continuous variable, final scores are a better outcome from a statistical perspective (Altman & Royston, 2006). Because students who dropped the class would have a lower final score because of missing exams, we also calculated final scores using an average of available exams. The pattern of results using the average, presented in Table S1, did not change. We retained students who ultimately dropped the class because they were likely to have been struggling with the material, which might reveal itself in the association of cortisol with performance. However, the pattern of results did not substantively change if these students were excluded from the analyses.

**Exam 2**

Similar to Exam 1, Exam 2 had 4 problems, each of which was comprised of several short-answer questions which had a maximum point-value of 102. Scores ranged from 22 to 99. On average, students scored 70.26 (69%) (*SD* = 16.44).

**S12. Information About Socioeconomic Status (SES) Measures**

**Family Income**

Participants reported their parents’ (or guardians’) average household income on a scale from 1 ($*0-$30,000*) to 9 (*$250,000+*), with an option for not knowing. For the 10 participants who did not know, this item was coded as missing.

**Parental Education**

Participants reported the highest level of education of their parents/guardians on a scale from 1 (*less than high school*) to 6 (*Ph.D. or professional degree (MD, MBA, JD)*). We used the highest level of either parent in the SES composite.

**Perceived Social Status**

To assess perceived SES, participants completed the *MacArthur Scale of Subjective Social Status*, which shows a picture of a ladder with 10 rungs representing SES in the United States, from lowest to highest (Adler et al., 2000). Participants were asked to locate their rung on the ladder compared to others in the United States based on their income, education, and occupation.

**S13. Addressing Missing Data and Controlling for Prior Performance**

Among our final sample, a total of 33 participants had no prior performance data on their transcripts. Nine students had not previously taken chemistry or biology classes, so we used their overall prior GPA (which, among other students, was correlated with chemistry/biology GPA at *r* = .78). Fourteen students did not previously attend the university, so we used their GPA in other classes from the same semester as the current class. The other 10 students did not consent to the release of their transcripts. Because this was not a critical variable for the analyses and data were missing from a small number of participants, we opted to use a strategy of mean replacement with an indicator variable. We deemed this strategy as acceptable because of the relatively small number of cases (< 4%) and the fact that prior performance was a covariate rather than a primary predictor of interest (Cheema, 2014). Specifically, we used the mean of other students’ prior biology and chemistry classes, along with an indicator variable, where 0 corresponded to students with intact prior performance data and 1 corresponded to those whose prior performance was mean imputed.

The number of biology and chemistry classes students had taken and their difficulty can affect students’ prior knowledge about biology and chemistry. Thus, we controlled for credits students earned from biology and chemistry courses and their mean levels of difficulty in our analyses predicting academic performance. The significance and direction of primary results are unchanged. As adding these two additional variables do not consistently improve the model fit across analyses, we did not include them in the final models.

**S14. Covariates Affecting Cortisol Concentrations**

**Time Since Awakening**

We used participants’ reports of their time of awakening to calculate the hours that participants had been awake at the time of each saliva sample collection.

**Medications**

Participants used a checklist to report if they had taken medications in the last 24-hours that could potentially alter cortisol levels, such as sedatives or barbiturates, antibiotics, anti-inflammatory medicines, or steroids (Granger et al., 2009). Responses were converted to a dichotomous variable (yes/no) indicating whether one or more of these medications had been taken.

**Tooth Brushing**

Participants were asked whether they had brushed their teeth in the hour before saliva collection (yes/no). Brushing teeth can cause gum bleeding that contaminates saliva and elevates salivary cortisol levels (Malamud & Rodriguez-Chavez, 2011).

**Additional Covariates**

Additional questions were assessed at saliva collection, but not included in the models because in preliminary analyses they were not associated with students’ cortisol levels. For a complete list, see https://bit.ly/3iaIgnp.

**S15. R Packages used in Current Study**

In this paper, we have used R packages, *ggplot2* (Wickham et al., 2016), *dplyr* (*Wickham* et al., 2014), *nlme* (Pinheiro et al., 2017), *psych* (Revelle, 2014), and *interactions* (Long, 2019).

# S16. Estimating Missing Saliva Questionnaire Data

**Covariates Related to Cortisol Concentration**

For participants who submitted at least one saliva questionnaire, we replaced missing saliva questionnaire data with their response from the next closest saliva questionnaire by class time and date. That is, we used data from each student’s completed questionnaire from the next closest class at the same time of day as the missing class. If a completed questionnaire was not available at the same class time, we used information from the next closest class by date.

Among the final sample of 271, 47 participants had missing data on at least some questions from the saliva questionnaire. Of these, 41 had submitted at least one other saliva questionnaire, so we replaced missing data for these participants with their responses following the rule noted above. For the six students who did not submit any saliva questionnaires, we calculated time since awakening based on these students’ typical school-day waking time reported in the baseline survey. For tooth brushing, we calculated the overall probability of other students having brushed their teeth before the morning and evening classes. We then used this probability to randomly generate an affirmative or negative response for these six students depending on the time they had attended class when they provided saliva. For medication use, we used the same strategy as tooth brushing (i.e., randomly assigning negative/affirmative responses based on overall sample probabilities) but without accounting for class time, since 24-hour medication would be unlikely to vary within a day.

# S17. Association Between Cortisol and Exam 2 Performance

In the following set of analyses, we controlled for cortisol from Saliva Class 1 and tested the association between averaged cortisol from Saliva Class 2 – 4 and performance on Exam 2, which took place a week after the last cortisol assessment (see Figure 1 in the main text).

We used the same hierarchical model-building approach as in the main text and above. The sample size for these analyses was reduced to the 225 students who submitted saliva samples for Saliva Class 1 and submitted at least one saliva sample across Saliva Classes 2, 3, and 4 and took the second exam.

In *Step 1*, controlling for Saliva Class 1 cortisol, students’ cortisol responses were not associated with their performance on the second exam, b = 2.30, *F*(1, 222) = 0.18, *p* = .67, 95% CI [-8.25, 12.84], = .001. In *Step 2*, adding the usual set of covariates associated with cortisol and/or academic performance, cortisol responses were marginally associated with better performance on the second exam, b = 8.91, *F*(1, 214) = 3.41, *p* = .07, 95% CI [-0.61, 18.43], = .016.

In *Step 3*, the interaction between students’ cortisol responses and their race/ethnicity predicting students’ performance on the second exam was not significant, b = 5.73, *F*(1, 213) = 2.40, *p* = .12, 95% CI [-1.56, 13.01], = .011. However, simple effects tests revealed that for URM students, higher cortisol responses predicted higher Exam 2 scores, b = 16.45, *F*(1, 213) = 5.78, *p* = .02, 95% CI [2.96, 29.94], = .026, whereas among non-URM students cortisol was not associated with Exam 2 scores, b = 5.00, *F*(1, 213) = 2.40, *p* = .36, 95% CI [-5.72, 15.71], = .004.

# S18. The Interaction Between Cortisol Responses and Other Demographic Variables Predicting Performance and Measures of Engagement

**Predicting Performance**

In addition to the final model in *Step 3*, we also tested whether socioeconomic status and gender interacted with cortisol to predict students’ academic performance. Across all models predicting different measures of academic performance, no individual interaction term was significant, and adding these two interaction terms did not significantly improve the model fit. Thus, these two interactions were not included in the final model.

**Predicting Measures of Engagement**

In addition to the final model, we also tested whether socioeconomic status and gender interacted with cortisol to predict students’ two after-class measures of engagement. Across two models predicting different measures of engagement, no individual interaction term was significant and adding these two interaction terms did not significantly improve the model fit. Thus, these two interactions were not included in the final model.

# S19. Using Averaged Cortisol to Predict Academic Performance

In the following set of analyses, we tested whether individual differences in cortisol levels are associated with academic performance. Using an average is advantageous because it provides a comprehensive and stable estimate of students’ cortisol levels during the course and provides an easily interpretable approach for testing whether students with overall higher/lower levels of cortisol in the course differed in their performance. To test this idea, we use cortisol levels aggregated across all classes as a predictor of students’ academic performance. We used the same hierarchical model-building approach described in the main text. The sample size for these analyses was 271 participants.

In *Step 1*, students’ average cortisol was not associated with their final scores, b = 9.02, *F*(1, 269) = 0.45, *p* = .51, 95% CI [-17.55, 35.58], = .002. Moreover, controlling for relevant covariates in *Step 2*, cortisol levels continued to not be associated with final scores, b = 14.06, *F*(1, 261) = 0.71, *p* = .40, 95% CI [-18.91, 47.04], = .003. As an ancillary analysis, we also tested an inverted U-shaped association between cortisol and performance. However, there was no evidence of an inverted U-shaped association between cortisol and academic performance, b = 10.53, *F*(1, 260) = 0.08, *p* = .78, 95% CI [-62.20, 83.27], < .001, and the quadratic term was dropped.

*Step 3* revealed a significant interaction between cortisol and race/ethnicity in predicting students’ final scores, b = 42.27, *F*(1, 260) = 9.50, *p* = .002, 95% CI [15.27, 69.27], = .04. Simple effects tests revealed that for URM students, cortisol predicted higher final scores, b = 80.19, *F*(1, 260) = 8.79, *p* = .003, 95% CI [26.92, 133.46], = .03. However, among non-URM students, cortisol was not associated with final scores, b = -4.35, *F*(1, 260) = 0.06, *p* = .80, 95% CI [-38.87, 30.17], < .001. Residual analyses revealed no outliers.

To test the association of cortisol with whether students dropped the course (0 = completed, 1 = dropped), we used logistic regression. The procedure was otherwise unchanged from above.

In *Step 1*, cortisol levels were not associated with whether students dropped the course, OR = 0.47, Wald χ2 (1, *N* = 271) = 0.87, *p* = .35, 95% CI [0.10, 2.29]. After including all relevant covariates in *Step 2,* cortisol remained unassociated with dropping the course, OR = 0.40, Wald χ2 (1, *N* = 271) = 0.47, *p* = .49, 95% CI [0.03, 5.39]. However, as above, *Step 3* revealed a significant interaction between cortisol and race/ethnicity, OR = 0.11, Wald χ2 (1, *N* = 271) = 4.77, *p* = .03, 95% CI [0.02, 0.80]. Simple effects tests revealed that for URM students, higher cortisol reduced the odds of dropping the course, OR = 0.02, Wald χ2 (1, *N* = 271) = 3.91, *p* = .048, 95% CI [0.00, 0.97]. For non-URM students, cortisol was not associated with dropping the course, OR = 1.57, Wald χ2 (1, *N* = 271) = 0.10, *p* = .76, 95% CI [0.09, 26.65]. Residual analyses revealed no outliers.

# S20. Residual Analyses to Detect Potential Outliers

Residual analysis from the model predicting students’ final class scores from their averaged cortisol levels across Saliva Class 2-4 (controlling for Saliva Class 1) did not reveal any problematic cases, as noted in the main text. Three participants had potentially problematic centered leverage values (i.e., > 3p/n, where p is the number of parameters including the intercept and n is the sample size; Hoaglin & Welsch, 1978) and 6 participants had studentized deleted residuals exceeding |3|. However, these observations did not appear to be particularly influential (i.e., Cook’s D < 0.06) and their removal did not affect the direction of coefficients or pattern of statistical significance reported in the main text.

Residual analysis from the model predicting whether students dropped the class from their averaged cortisol levels across Saliva Class 2-4 (controlling for Saliva Class 1) also did not reveal any problematic cases. Fourteen participants had potentially problematic centered leverage values and one participant had a value of Cook’s D exceeding 1, but none had studentized deleted residual values > |3|. Although this suggests little reason to be concerned with outliers, we nonetheless conducted the analysis again excluding the 14 cases with potentially problematic leverage values. Doing so did not affect the direction of coefficients or pattern of statistical significance reported in the main text.

# S21. Association Between Social Identity Threat,

# Cortisol Responses and Performance for URM Students

We conducted moderated mediation to test the association between social identity threat, cortisol, and performance for URM students. The independent variable was social identity threat, and the mediator was cortisol responses around the exam period. We tested the same outcomes as elsewhere: students’ final scores, whether they dropped the class, and Exam 1 and 2. The moderator was URM status, which was included as a predictor of performance from cortisol (other paths were not moderated by URM status).[[1]](#footnote-1) We controlled for cortisol from Saliva Class 1 and included the same set of covariates that were included elsewhere as predictors of the mediator and outcome. Indirect effects were evaluated using 95% confidence intervals estimated from 10,000 bootstrapped samples.

Overall, results partially suggest that higher cortisol responses for URM students (but not for non-URM students) mediated the association between perceiving higher social identity threat and better performance outcomes. However, this association was only significant for models predicting final scores and Exam 2 (but results also trended in the same direction in models predicting other DVs).

**Final Scores**

Among URM students, higher cortisol responses (i.e., cortisol averaged across Saliva Class 2-4) significantly mediated the association between higher social identity threat and better performance, indirect effect = 3.03, 95% CI [0.81, 6.07]. However, there was no such effect for non-URM students, indirect effect = 0.41, 95% CI [-0.92, 1.94] (see Figure S2 below). The overall moderated mediation was significant, index = 2.57, 95% CI [0.53, 5.35].

**Dropping the Course**

Higher cortisol responses (i.e., cortisol averaged across Saliva Class 2-4) did not significantly mediate the association between higher social identity threat and lower rates of dropping the course for both URM students, indirect effect = -1.30, 95% CI [-0.59, 0.04], and non-URM students, indirect effect = 0.02, 95% CI [-0.19, 0.19]. The overall moderated mediation was not significant, index = -1.15, 95% CI [-0.53, 0.04]. However, the pattern trended such that higher social identity led to lower rates of dropping the course via higher levels of cortisol.

**Exam 1 Performance**

Higher cortisol responses (i.e., cortisol assessed in Saliva Class 2) did not significantly mediate the association between higher social identity threat and better performance for both URM students, indirect effect = 0.24, 95% CI [-0.12, 0.80], and non-URM students, indirect effect = -0.07, 95% CI [-0.39, 0.12]. The overall moderated mediation was not significant, index = 0.31, 95% CI [-0.10, 0.99]. However, the pattern trended such that higher social identity led to better performance in the course via higher levels of cortisol.

**Exam 2 Performance**

Among URM students, higher cortisol responses (i.e., cortisol averaged across Saliva Class 2-4) did significantly mediate the association between higher social identity threat and better performance outcomes, indirect effect = 0.65, 95% CI [0.06, 1.41]. There was no significant indirect effect for non-URM students, indirect effect = 0.19, 95% CI [-0.25, 0.67]. The overall moderated mediation was not significant, index = 0.46, 95% CI [-0.14, 1.23].

# Table S1:Predicting Averaged Exam Scores

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table S1**  *Regression Coefficients Predicting Averaged Exam Scores Step 1-3* | | | | | | | | | | | | | | | | |
|  | |  | | --- | | Step 1 | | | | | | |  | | --- | | Step 2 | | | | | | |  | | --- | | Step 3 | | | | | | |
| Variable | b | se(b) | | CI(95%) | | b | se(b) | | CI(95%) | | b | | se(b) | CI(95%) | | |
| |  | | --- | |  | | |  | | --- | |  | | | | | | |  | | --- | |  | | | | | | |  | | --- | |  | | | | | | |
| Intercept | 72.87\*\*\* | | 0.84 | 71.21 | 74.53 | 24.05\*\*\* | | 5.38 | 13.44 | 34.66 | 23.59\*\*\* | 5.35 | | 13.05 | 34.12 |
| Saliva Class 1 Cortisol1 | -0.95 | | 3.77 | -8.38 | 6.49 | -3.55 | | 3.03 | -9.51 | 2.41 | -2.79 | 3.03 | | -8.76 | 3.17 |
| Saliva Class 2-4 Cortisol1 | 2.65 | | 4.24 | -5.70 | 10.99 | 6.58 | | 3.67 | -0.66 | 13.81 | 8.46\* | 3.76 | | 1.05 | 15.86 |
| Race/ethnicity 2 |  | |  |  |  | -2.42\*\* | | 0.77 | -3.94 | -0.89 | -2.48\*\*\* | 0.77 | | -3.99 | -0.96 |
| Gender3 |  | |  |  |  | -0.76 | | 0.69 | -2.12 | 0.59 | -0.86 | 0.69 | | -2.21 | 0.49 |
| SES4 |  | |  |  |  | 1.50 | | 0.87 | -0.22 | 3.22 | 1.53 | 0.87 | | -0.17 | 3.24 |
| Prior performance |  | |  |  |  | 14.31\*\*\* | | 1.43 | 11.49 | 17.12 | 14.37\*\*\* | 1.42 | | 11.57 | 17.16 |
| Missing prior performance5 |  | |  |  |  | 0.79 | | 3.61 | -6.32 | 7.90 | 1.90 | 3.62 | | -5.24 | 9.03 |
| Time since awakening6 |  | |  |  |  | -0.16 | | 0.40 | -0.95 | 0.63 | -0.12 | 0.40 | | -0.90 | 0.66 |
| Medication7 |  | |  |  |  | 1.21 | | 2.29 | -3.30 | 5.73 | 0.61 | 2.29 | | -3.91 | 5.13 |
| Brush teeth8 |  | |  |  |  | 0.19 | | 2.44 | -4.61 | 4.99 | 0.24 | 2.42 | | -4.52 | 5.01 |
| Cortisol X race/ethnicity |  | |  |  |  |  | |  |  |  | 5.79\* | 2.80 | | 0.27 | 11.31 |
| |  | | --- | |  | | |  | | --- | |  | | | | | | |  | | --- | |  | | | | | | |  | | --- | |  | | | | | | |
| *N* = 234 | *R*2 = .002 | | | | | *R*2 = .46\*\*\* | | | | | *R*2 = .47\* | | | | | |
| *Note*. The dependent variable in these analyses is students’ average score on the exams that they took, rather than their final class score as calculated by the instructor and reported in the main text. *R*2 change between steps significant at *p* < .001. 1 Cortisol level averaged across class(es) and then log-transformed and mean centered. 2 Non-URM = -1, URM = 1. 3 Male = -1, Female = 1. 4 An average of standardized SES variables (parent’s income, parent’s education level, social status). 5 Participants with prior performance data = 0, Participants without prior performance data due to not consenting to release their transcript = 1. 6 Time since awakening averaged across four classes (in hours). 7 Whether students took medication 24 hours before saliva collection, averaged across the four classes in which saliva samples were collected. 8 Whether students had brushed their teeth an hour before the saliva collection, averaged across the four classes in which saliva samples were collected.  \**p ≤ .05.* \*\**p ≤ .01.* \*\*\**p ≤ .001.* | | | | | | | | | | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S2: Association Between Cortisol and Self-Reported Focus and Engagement **Table S2**  *Two-Level Models of the Association Between Cortisol and Self-Reported Focus and Engagement in 76 Participants (135 time points)* | | | | | | | | |
| Fixed effects | Estimate  [95% CI] | *t* | *p* |  | Estimate  [95% CI] | *t* | *p* |
| **Dependent variable:**  Focus |  |  |  | **Dependent variable:**  Engagement |  |  |  |
| **Independent variables:**  **Level 1**  **(Within-Person)** |  |  |  |  |  |  |  |
| Intercept | 3.22\*\*\*  [3.02, 3.41] | 31.82 | <.001 |  | 2.96  [2.73, 3.20] | 24.31\*\*\* | < 0.001 |
| Cortisol1 | 0.82\*  [0.07, 1.57] | 2.12 | 0.04 |  | 0.60  [-0.19, 1.39] | 1.47 | 0.15 |
| Time since awakening1,2 | 0.02  [-0.07, 0.11] | 0.43 | .67 |  | -0.03  [-0.13, 0.07] | -0.53 | 0.60 |
| Medication | -0.03  [-0.49, 0.43] | -0.12 | .91 |  | -0.24  [-0.71, 0.24] | -0.97 | 0.34 |
| Brush teeth | -0.24  [-0.70, 0.23] | -0.99 | .32 |  | -0.20  [-0.70, 0.29] | -0.79 | 0.43 |
| **Level 2**  **(Between-person)** |  |  |  |  |  |  |  |
| Race/ethnicity3 | -0.09  [-0.30, 0.11] | -0.89 | .38 |  | -0.06  [-0.31, 0.18] | -0.50 | 0.62 |
| Gender4 | -0.07  [-0.28, 0.14] | -0.64 | .53 |  | -0.13  [-0.38, 0.12] | -0.99 | 0.33 |
| SES5 | -0.09  [-0.33, 0.15] | -0.74 | .46 |  | -0.002  [-0.29, 0.29] | -0.01 | 0.99 |
| Cortisol X Race/ethnicity | 0.30  [-0.25, 0.86] | 1.06 | .29 |  | 0.11  [-0.49, 0.71] | 0.37 | 0.71 |
| Random effect | Variance |  |  |  |  |  |  |
| **Level 1**  **(Within- person)** |  |  |  |  |  |  |  |
| Intercept | 0.26  [0.04, 1.51] |  |  |  | 0.67  [0.47, 0.96] |  |  |
| 1 Values were aggregated for each class. 2 Time since awakening coded in hours. 3 Non-URM = -1, URM = 1. 4 Male = -1, Female = 1. 5 An average of standardized SES variables (parent’s income, parent’s education level, social status). \**p ≤ .05.* \*\**p ≤ .01.* \*\*\**p ≤ .001.* | | | | | | | | |

# Figure S1

|  |
| --- |
| **Figure S1**  *Students’ Self-Reported Daily Stress Across the Smartphone Study (21 Days)* |
|  |
| *Note.* Values are students’ average daily self-reported stress levels across the 21-day smartphone study. Each weeknight participants responded to the question, “How stressful was your day?” on an anchored scale from 1 (*not at all stressful*) to 9 (*very stressful*). The dashed vertical line indicates the date of the first exam. Saliva Class 1 occurred before the smartphone study. Saliva Class 4 was rescheduled to two different days because of a religious holiday on the scheduled class day; class time was unchanged. |

# Figure S2

|  |
| --- |
| **Figure S2**  *Moderated Mediation of Final Scores from Social Identity Threat through Cortisol Responses as a Function of Race/Ethnicity* |
|  |
| *Note*. All coefficients are unstandardized. Cortisol levels averaged across Saliva classes 2-4 and then log-transformed. We control for Saliva 1 class cortisol and all covariates used in the previous section. NS = not significant (*p* > .05).  \**p* ≤ .05. \*\**p* ≤ .01. \*\*\**p* ≤ .001. |

**References**

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. *Health Psychology, 19*(6), 586-592. https://doi.org/http://dx.doi.org/10.1037/0278-6133.19.6.586

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *Bmj*, 332(7549), 1080. <https://doi.org/10.1136/bmj.332.7549.1080>

Cheema, J. R. (2014). Some general guidelines for choosing missing data handling methods in educational research. *Journal of Modern Applied Statistical Methods, 13*(2), Article 3. <https://doi.org/10.22237/jmasm/1414814520>

Granger, D. A., Hibel, L. C., Fortunato, C. K., & Kapelewski, C. H. (2009). Medication effects on salivary cortisol: tactics and strategy to minimize impact in behavioral and developmental science. *Psychoneuroendocrinology, 34*(10), 1437-1448. <https://doi.org/10.1016/j.psyneuen.2009.06.017>

Hoaglin, D. C., & Welsch, R. E. (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician, 32*(1), 17-22. doi:10.1080/00031305.1978.10479237

Jordt, H., Eddy, S. L., Brazil, R., Lau, I., Mann, C., Brownell, S. E., King, K., & Freeman, S. (2017). Values affirmation intervention reduces achievement gap between underrepresented minority and white students in introductory biology classes. *CBE Life Sciences Education, 16*(3), ar41. <https://doi.org/https://doi.org/10.1187/cbe.16-12-0351>

Kirschbaum, C., & Hellhammer, D. H. (1989). Salivary cortisol in psychobiological research: an overview. *Neuropsychobiology, 22*(3), 150-169. https://doi.org/10.1159/000118611

Long, J. A. (2019). interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions [Computer software]. Retrieved from <https://cran.r-project.org/package=interactions>.

Malamud, D., & Rodriguez-Chavez, I. R. (2011). Saliva as a Diagnostic Fluid. *Dental Clinics of North America, 55*(1), 159-178. https://doi.org/10.1016/j.cden.2010.08.004

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2017). R package "nlme" [Computer software]. Retrieved from http://cran.r-project.org/package=nlme.

Revelle, W. (2014). Package ‘psych’ [Computer software]. Retrieved from http://cran.r-project.org/web/packages/psych/psych.pdf

Sax, L. J., Lim, G., Lehman, K., & Lonje-Paulson, L. (2018). Reversal of the Gender Gap: The biological sciences as a unique case within science, technology, engineering, and mathematics (STEM). *Journal of Women and Minorities in Science and Engineering, 24*(4), 291-324. https://doi.org/10.1615/JWomenMinorScienEng.2018019995

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package ‘ggplot2’. *Create Elegant Data Visualisations Using the Grammar of Graphics*. Version, 2(1), 1-189.

Wickham, H., François, R., Henry, L., & Müller, K. (2019a). dplyr: A grammar of data

Manipulation [Computer software]. Retrieved from https://CRAN.R-project.org/package=dplyr

1. The pattern and significance of the results do not change if we conduct models where URM status moderated all paths. [↑](#footnote-ref-1)