Replication Report: Empathy, Mental Health, and Burnout in Medical Students

CID:06014353

2025-05-01

Contents

1 Background and Aim

This report attempts to replicate a study by Carrard et al. (2021), published in *Medical Teacher* (Carrard et al. (2022)). The core focus of the study is the relationship between medical students' empathy (cognitive, affective, and behavioral dimensions) and their mental health (depression, anxiety) as well as professional burnout. The original study used cross-sectional survey data from 886 medical students in years 1 to 6 at a Swiss medical school. Relevant variables were measured using self-report questionnaires, and linear regression was employed to analyze the relationships between empathy, mental health, and burnout. The replication aims to fully reproduce the key results of the original article (including Table 2 and Table 3) using R software and the publicly available original dataset on the Zenodo platform (https://zenodo.org/r ecords/5702895#.Y8OraNJBwUE), in order to assess the reproducibility and robustness of the findings.

2 Methods

The data used in this replication come from the original study's publicly available questionnaire dataset on the Zenodo platform. It includes self-reported data from 886 Swiss medical students, covering empathy (JSPE-S, QCAE, AMSP, GERT-S) across cognitive, affective, and behavioral dimensions, mental health (CES-D, STAI), and professional burnout (MBI). Data preprocessing strictly followed the original study: individuals identifying as non-binary (n = 5) were excluded. Gender and year variables were converted into factors for regression modeling. A binary variable indicating "depression risk" was created based on CES-D gender-specific cut-offs (male ≥ 16, female ≥ 21).

The main analysis has two parts: first, 10 linear regression models were built using year and gender as predictors to examine their effects on mental health and burnout dimensions; second, 25 models were constructed

to test the predictive effects of five empathy dimensions on five mental health and burnout outcomes. All models were built using the `lm()` function, with formulas generated in batch via `reformulate()`.

To ensure comparability with the original results, standardized coefficients (β) were computed using `parameters::standardize_parameters()` with `method = "basic"`, applying the following transformation:

$$\beta^* = \hat{\beta} \cdot \frac{\text{SD}(x)}{\text{SD}(y)}$$

In addition, partial η^2 (`effectsize::eta_squared()`), standard errors, F-statistics, and R^2 were extracted to fully reproduce the original result structure.

As supplementary analyses, two key descriptive tests from the original Results section were also replicated: (1) a chi-square test comparing the sample's gender distribution to the school's full student body to assess representativeness. The null hypothesis assumed no difference between sample and population gender distribution; (2) the proportion of students at clinical risk of depression, based on CES-D cut-offs, was calculated to verify key proportions reported in the article.

3 Results

The table below presents the descriptive statistics obtained in this replication. The reported variables closely match those in Table 1 of the original study, indicating that data loading and processing were accurate. Due to the large number of original variables, this report shows only key demographic and mental health variables; the full list can be found in the appendix (Table3) or figures folder. In addition, I replicated a supplementary descriptive analysis from the original study: a chi-square test comparing the sample's gender distribution with that of the full medical school population. The result showed no significant difference ($\chi^2 = 1.24$, df = 1, $p = 0.265$), consistent with the original, supporting the representativeness of the sample.

Table 1: The results of reproducing descriptive statistics (N = 886)

Variable	Mean (SD)	Percent
Sex: Female	-	68.4%
Sex: Male	-	31.04%
Sex: Non-binary	-	0.56%
Year: B1	-	27.65%
Year: B2	-	15.24%
Year: B3	-	16.14%
Year: M1	-	13.88%
Year: M2	-	14.33%
Year: M3	-	12.75%
AMSP	23.15 (4.99)	-
Depressive symptoms	18.05 (11.48)	-
GERT-S	0.72 (0.09)	-
JSPE-S	106.37 (8.78)	-
Cynicism	10.08 (4.59)	-
Academic efficacy	24.21 (4.63)	-
Emotional exhaustion	16.88 (5.26)	-
QCAE affective	34.78 (5.38)	-
QCAE cognitive	58.53 (6.57)	-
Anxiety	42.9 (11.98)	-

To replicate the findings in Table 2 of the original paper regarding the effects of year and gender on mental health and burnout, we built linear regression models for five mental health and burnout indicators (CES-D depression, STAI anxiety, and three MBI dimensions) and five empathy indicators (JSPE-S, QCAE, AMSP, GERT-S). Standardized β coefficients were used to assess the effects of year (relative to B1) and

gender (female vs. male) on each outcome. Full results are in the appendix (Table4); this section presents a visualization of the replication results (see figure below).

The figure shows that most mental health indicators (e.g., depressive symptoms, anxiety, emotional exhaustion) are negatively associated with higher years, suggesting improved mental states over time. Female students show positive standardized βs for depression, anxiety, and emotional exhaustion, indicating more reported negative emotional experiences. Females also score significantly higher on nearly all empathy dimensions, with the largest difference in QCAE affective empathy ($\beta = 0.37^{***}$).

These results are consistent with the conclusions in Table 2 of the original paper, supporting the authors' interpretation that higher-year students have better mental health and female students show higher empathy. Overall, the regression analysis was successfully replicated.
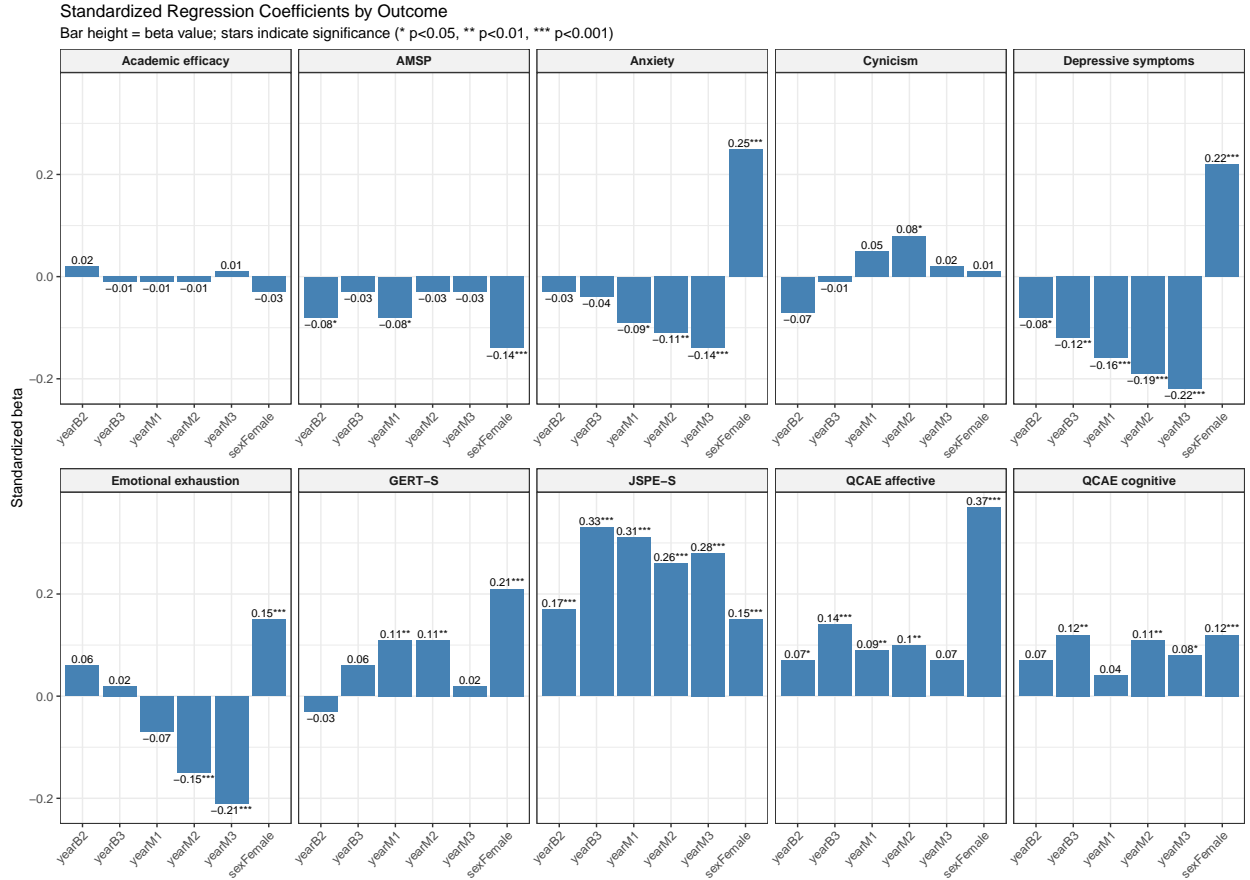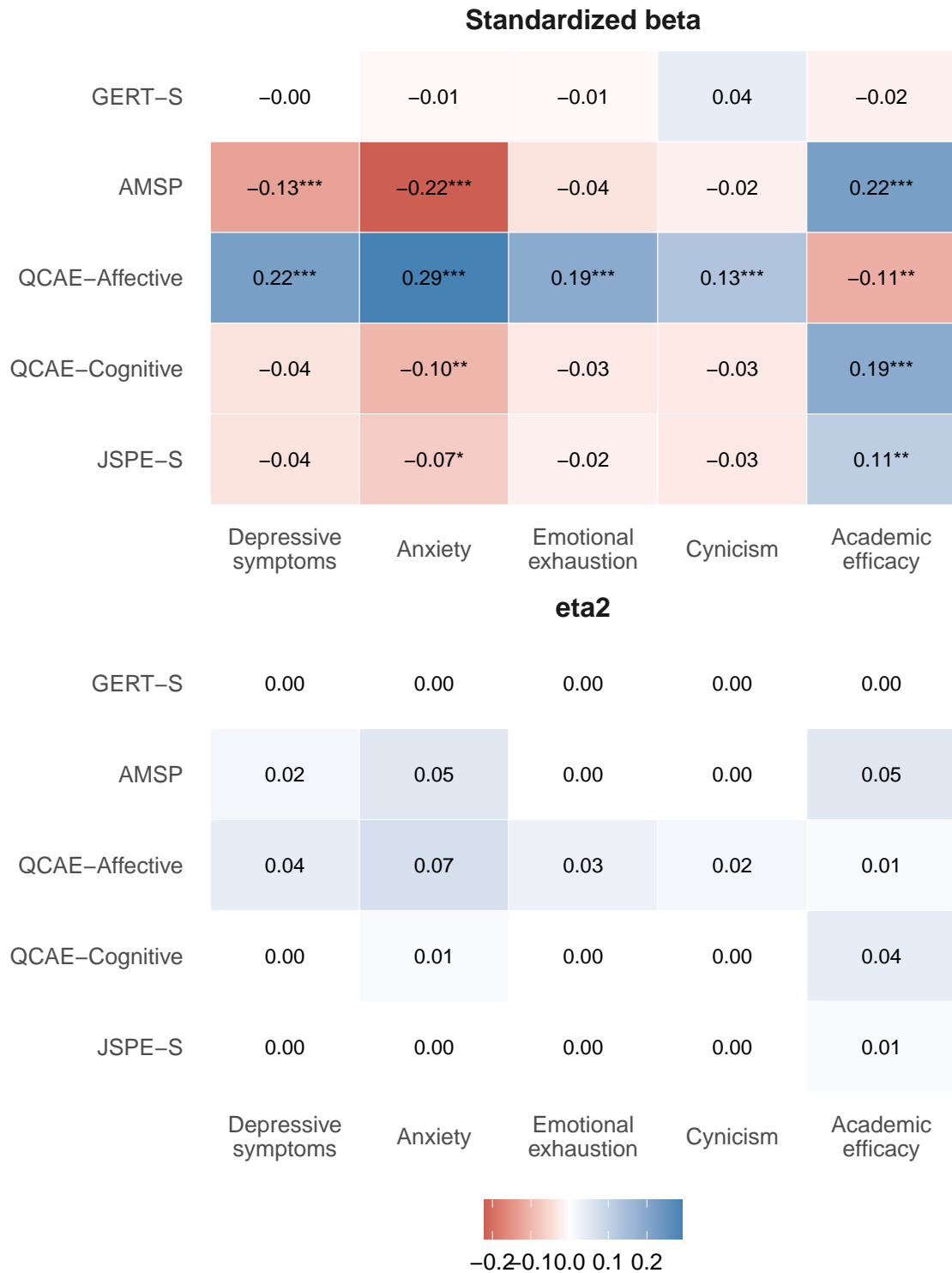

Figure 1: Standardized regression coefficients (beta) by outcome, with significance stars

To replicate the regression analyses in Table 3 of the original study, we built linear models for five mental health and burnout outcomes (CES-D, STAI, and three MBI dimensions), using five empathy measures (JSPE-S, QCAE cognitive, QCAE affective, AMSP, GERT-S) as predictors. Year and gender were controlled, consistent with the original setup.

The figure below shows the standardized β coefficients and effect sizes (η^2). Color indicates the direction and magnitude of β; text labels show exact β values with significance levels (* p<.05, ** p<.01, *** p<.001). The full regression table with standard errors, F-statistics, and R^2 is in the appendix (Table5).

The results closely replicate the core findings of the original study: 1. QCAE affective empathy is significantly related to all outcomes, with positive β values, indicating students with higher affective empathy report more mental health problems, though effect sizes are small ($\eta^2 < .05$). 2. In

contrast, AMSP behavioral empathy predicts lower depression and anxiety, and higher academic efficacy (i.e., less burnout), with slightly stronger effects (max $\eta^2 = .05$). 3. JSPE-S and QCAE cognitive empathy are significantly linked to reduced anxiety and higher academic efficacy, but effects are weaker. 4. GERT-S emotion recognition shows no significant relation with any outcome.

Standardized beta

	Depressive symptoms	Anxiety	Emotional exhaustion	Cynicism	Academic efficacy
GERT–S	−0.00	−0.01	−0.01	0.04	−0.02
AMSP	−0.13***	−0.22***	−0.04	−0.02	0.22***
QCAE–Affective	0.22***	0.29***	0.19***	0.13***	−0.11**
QCAE–Cognitive	−0.04	−0.10**	−0.03	−0.03	0.19***
JSPE–S	−0.04	−0.07*	−0.02	−0.03	0.11**

eta2

	Depressive symptoms	Anxiety	Emotional exhaustion	Cynicism	Academic efficacy
GERT–S	0.00	0.00	0.00	0.00	0.00
AMSP	0.02	0.05	0.00	0.00	0.05
QCAE–Affective	0.04	0.07	0.03	0.02	0.01
QCAE–Cognitive	0.00	0.01	0.00	0.00	0.04
JSPE–S	0.00	0.00	0.00	0.00	0.01

−0.2 −0.1 0.0 0.1 0.2

To replicate the descriptive analysis of CES-D depression risk in the original study, we classified the sample using gender-specific cut-offs (male ≥ 16, female ≥ 21). About 40.98% of students in the replication sample are at clinical depression risk, close to the 40.18% reported in the original. A breakdown by year shows a general decline in risk across years, closely matching the original trend. The table below compares depression risk proportions by year:

Table 2: Comparison of CES-D clinical risk percentages by year (Reproduction vs. Original study)

Year	Reproduced (%)	Original (%)	Difference (%)
B1	52.46	51.02	1.44
B2	47.41	45.93	1.48
B3	39.86	39.16	0.70
M1	35.25	34.96	0.29
M2	32.54	33.07	-0.53
M3	25.23	24.78	0.45

4 Discussion and Improvement suggestions

The replication process in this study went smoothly overall, thanks to the original paper's detailed model descriptions, data access, and preprocessing steps, showing good reproducibility and research transparency. Although I used a different analysis tool (R instead of Stata), I was still able to successfully replicate the key regression results by following the publicly available data and methods, demonstrating the robustness of the original findings. This kind of cross-platform reproducibility is valuable in medical education research, increasing confidence in future intervention design and practical application.

Several key technical points emerged during the replication. For example, when fitting initial linear models with unstandardized data, the resulting β coefficients differed from the original, although effect sizes (η^2) and F-values remained consistent. Upon further review, I confirmed that the original study used standardized β coefficients. I therefore applied `parameters::standardize_parameters(method = "basic")` in R to standardize the results. This method does not transform the data directly but uses the following formula to adjust coefficients based on the ratio of standard deviations:

$$\beta^* = \hat{\beta} \cdot \frac{\text{SD}(x)}{\text{SD}(y)}$$

This approach is logically equivalent to Stata's `, beta` option. After standardization, the β values matched the original exactly, confirming the authors' method and highlighting the importance of coefficient definitions in cross-platform replication.

For the CES-D positive rate (depression risk), I followed the gender-specific cut-offs noted in the paper (16 for males, 21 for females), and tested both ">" and ">" thresholds. In both cases, the calculated rate differed from the original by about 1%, though the trend was fully consistent. I suspect this difference may stem from how the authors handled scores at the cut-off (exactly 16 or 21), which was not explained in the paper. One possibility is that decisions were made based on clinical judgment or questionnaire logic; another is that they used adjusted or standardized scores (e.g., Z-scores, T-scores, scale scores, or totals excluding some items) rather than raw CES-D totals. As the paper lacks detail, such differences likely reflect unobservable decisions rather than replication error.

Building on the replication, I also optimized the result presentation by adding visual outputs. Compared to tables alone, bar plots and heatmaps more clearly show the direction and significance of standardized β coefficients, and allow more efficient comparison across models. This improves readability and communication, aligning with current expectations for interpretability and user experience in research reporting.

5 Reflection and Ethical Considerations

The data used in this replication come from an ethically approved project (approved by the Ethics Committee of the Canton of Vaud, Switzerland, No. 2020-02474). The original research team ensured participants' informed rights and autonomy through appropriate procedures in recruitment, compensation, and informed consent. In this replication, I also followed ethical standards by using only public data, avoiding access to raw sensitive scores, and conducting analyses strictly based on the original cut-off values.

In reflection, we believe that if the original study had provided more details on scale scoring (e.g., CES-D item scores and processing logic), it would improve the transparency and accuracy of cut-off determinations. During preprocessing, the original study excluded non-binary participants (n = 6). While this aligns with conventional modeling, it raises questions about inclusivity. Future research could consider alternative handling methods for rare categories (e.g., category merging, small-sample modeling, or weighted estimates) to enhance representativeness.

In addition, the current models are all main-effect models. Future studies could explore interaction effects, multiverse analysis, or multilevel modeling to better capture complex relationships and improve explanatory power.

6 Appendix A.

Table 3: Full descriptive statistics for all categorical variables (replication)

Variable	Mean (SD)	Percent
Consulted psy last year: No	NA	77.54%
Consulted psy last year: Yes	NA	22.46%
Has paid job: No	NA	65.12%
Has paid job: Yes	NA	34.88%
Has partner: No	NA	43.68%
Has partner: Yes	NA	56.32%
Mother tongue: Arab	NA	0.34%
Mother tongue: English	NA	2.48%
Mother tongue: French	NA	80.93%
Mother tongue: German	NA	3.5%
Sex: Female	NA	68.4%
Sex: Male	NA	31.04%
Sex: Non-binary	NA	0.56%
Year: B1	NA	27.65%
Year: B2	NA	15.24%
Year: B3	NA	16.14%
Year: M1	NA	13.88%
Year: M2	NA	14.33%
Year: M3	NA	12.75%
age	22.38 (3.3)	NA
amsp	23.15 (4.99)	NA
cesd	18.05 (11.48)	NA
erec_mean	0.72 (0.09)	NA
glang 102	NA	3.05%
glang 104	NA	0.45%
glang 106	NA	0.68%

glang 108		NA	0.11%
glang 114		NA	0.11%
glang 118		NA	0.23%
glang 120		NA	0.23%
glang 121		NA	1.47%
glang 54		NA	0.11%
glang 60		NA	0.34%
glang 63		NA	0.56%
glang 90		NA	5.08%
glang 92		NA	0.11%
glang 95		NA	0.11%
glang 98		NA	0.11%
health		3.78 (1.06)	NA
jspe		106.37 (8.78)	NA
mbi_cy		10.08 (4.59)	NA
mbi_ea		24.21 (4.63)	NA
mbi_ex		16.88 (5.26)	NA
qcae_aff		34.78 (5.38)	NA
qcae_cog		58.53 (6.57)	NA
stai_t		42.9 (11.98)	NA
stud_h		25.29 (15.93)	NA

Table 4: Regression results by outcome (replication)

Predictor	beta	SE	eta2	F	R^2
CES-D (Depressive symptoms)					
yearB2	-0.08*	1.17	0.06	17.82	0.11
yearB3	-0.12**	1.15	0.06	17.82	0.11
yearM1	-0.16***	1.21	0.06	17.82	0.11
yearM2	-0.19***	1.19	0.06	17.82	0.11
yearM3	-0.22***	1.25	0.06	17.82	0.11
sexFemale	0.22***	0.79	0.05	17.82	0.11
STAI-T (Anxiety)					
yearB2	-0.03	1.23	0.03	14.34	0.09
yearB3	-0.04	1.21	0.03	14.34	0.09
yearM1	-0.09*	1.27	0.03	14.34	0.09
yearM2	-0.11**	1.26	0.03	14.34	0.09
yearM3	-0.14***	1.32	0.03	14.34	0.09
sexFemale	0.25***	0.84	0.06	14.34	0.09
MBI-EX (Emotional exhaustion)					
yearB2	0.06	0.54	0.08	16.16	0.10
yearB3	0.02	0.53	0.08	16.16	0.10
yearM1	-0.07	0.55	0.08	16.16	0.10
yearM2	-0.15***	0.55	0.08	16.16	0.10
yearM3	-0.21***	0.57	0.08	16.16	0.10
sexFemale	0.15***	0.36	0.03	16.16	0.10
MBI-CY (Cynicism)					

Table 4: Regression results by outcome (replication) *(continued)*

Predictor	beta	SE	eta2	F	R^2
yearB2	-0.07	0.49	0.02	2.55	0.02
yearB3	-0.01	0.48	0.02	2.55	0.02
yearM1	0.05	0.51	0.02	2.55	0.02
yearM2	0.08*	0.50	0.02	2.55	0.02
yearM3	0.02	0.52	0.02	2.55	0.02
sexFemale	0.01	0.33	0.00	2.55	0.02
MBI-EA (Academic efficacy)					
yearB2	0.02	0.50	0.00	0.32	0.00
yearB3	-0.01	0.49	0.00	0.32	0.00
yearM1	-0.01	0.51	0.00	0.32	0.00
yearM2	-0.01	0.51	0.00	0.32	0.00
yearM3	0.01	0.53	0.00	0.32	0.00
sexFemale	-0.03	0.34	0.00	0.32	0.00
JSPE-S					
yearB2	0.17***	0.87	0.13	25.72	0.15
yearB3	0.33***	0.86	0.13	25.72	0.15
yearM1	0.31***	0.90	0.13	25.72	0.15
yearM2	0.26***	0.89	0.13	25.72	0.15
yearM3	0.28***	0.93	0.13	25.72	0.15
sexFemale	0.15***	0.59	0.02	25.72	0.15
QCAE-Cognitive					
yearB2	0.07	0.70	0.01	4.36	0.03
yearB3	0.12**	0.69	0.01	4.36	0.03
yearM1	0.04	0.72	0.01	4.36	0.03
yearM2	0.11**	0.71	0.01	4.36	0.03
yearM3	0.08*	0.75	0.01	4.36	0.03
sexFemale	0.12***	0.48	0.02	4.36	0.03
QCAE-Affective					
yearB2	0.07*	0.54	0.01	24.74	0.15
yearB3	0.14***	0.53	0.01	24.74	0.15
yearM1	0.09**	0.55	0.01	24.74	0.15
yearM2	0.1**	0.55	0.01	24.74	0.15
yearM3	0.07	0.57	0.01	24.74	0.15
sexFemale	0.37***	0.36	0.14	24.74	0.15
AMSP					
yearB2	-0.08*	0.53	0.01	3.80	0.03
yearB3	-0.03	0.52	0.01	3.80	0.03
yearM1	-0.08*	0.55	0.01	3.80	0.03
yearM2	-0.03	0.54	0.01	3.80	0.03
yearM3	-0.03	0.57	0.01	3.80	0.03
sexFemale	-0.14***	0.36	0.02	3.80	0.03
GERT-S					
yearB2	-0.03	0.01	0.02	10.45	0.07
yearB3	0.06	0.01	0.02	10.45	0.07

Table 4: Regression results by outcome (replication) *(continued)*

Predictor	beta	SE	eta2	F	R^2
yearM1	0.11**	0.01	0.02	10.45	0.07
yearM2	0.11**	0.01	0.02	10.45	0.07
yearM3	0.02	0.01	0.02	10.45	0.07
sexFemale	0.21***	0.01	0.05	10.45	0.07

Table 5: Relationship between empathy, mental health, and burnout (replication).

Empathy	Outcome	Beta	p stars	SE	η^2 (partial)	F statistic	R^2
JSPE-S	Depressive symptoms	-0.04	NA	0.05	0.00	15.42	0.11
JSPE-S	Anxiety	-0.07	*	0.05	0.00	12.93	0.09
JSPE-S	Emotional exhaustion	-0.02	NA	0.02	0.00	13.87	0.10
JSPE-S	Cynicism	-0.03	NA	0.02	0.00	2.32	0.02
JSPE-S	Academic efficacy	0.11	**	0.02	0.01	1.69	0.01
QCAE-Cognitive	Depressive symptoms	-0.04	NA	0.06	0.00	15.49	0.11
QCAE-Cognitive	Anxiety	-0.10	**	0.06	0.01	13.69	0.10
QCAE-Cognitive	Emotional exhaustion	-0.03	NA	0.03	0.00	13.97	0.10
QCAE-Cognitive	Cynicism	-0.03	NA	0.02	0.00	2.28	0.02
QCAE-Cognitive	Academic efficacy	0.19	***	0.02	0.04	5.02	0.04
QCAE-Affective	Depressive symptoms	0.22	***	0.07	0.04	22.15	0.15
QCAE-Affective	Anxiety	0.29	***	0.07	0.07	24.30	0.16
QCAE-Affective	Emotional exhaustion	0.19	***	0.03	0.03	18.66	0.13
QCAE-Affective	Cynicism	0.13	***	0.03	0.02	4.22	0.03
QCAE-Affective	Academic efficacy	-0.11	**	0.03	0.01	1.64	0.01
AMSP	Depressive symptoms	-0.13	***	0.07	0.02	17.78	0.12
AMSP	Anxiety	-0.22	***	0.08	0.05	19.78	0.14
AMSP	Emotional exhaustion	-0.04	NA	0.03	0.00	14.09	0.10
AMSP	Cynicism	-0.02	NA	0.03	0.00	2.26	0.02
AMSP	Academic efficacy	0.22	***	0.03	0.05	6.44	0.05
GERT-S	Depressive symptoms	0.00	NA	4.06	0.00	15.26	0.11
GERT-S	Anxiety	-0.01	NA	4.29	0.00	12.29	0.09
GERT-S	Emotional exhaustion	-0.01	NA	1.87	0.00	13.84	0.10
GERT-S	Cynicism	0.04	NA	1.70	0.00	2.41	0.02
GERT-S	Academic efficacy	-0.02	NA	1.73	0.00	0.31	0.00

References

Carrard, Valerie, Céline Bourquin, Sylvie Berney, Katja Schlegel, Jacques Gaume, Pierre-Alexandre Bart, Martin Preisig, Marianne Schmid Mast, and Alexandre Berney and. 2022. "The Relationship Between Medical Students' Empathy, Mental Health, and Burnout: A Cross-Sectional Study." *Medical Teacher* 44 (12): 1392–99. https://doi.org/10.1080/0142159X.2022.2098708.