



OPEN

Revisiting effects of teacher characteristics on physiological and psychological stress: a virtual reality study

Lisa Bardach¹✉, Yizhen Huang^{2,4}, Eric Richter², Robert M. Klassen³, Thilo Kleickmann⁴ & Dirk Richter²

Identifying personal characteristics associated with teachers' stress is a longstanding research goal with important implications for practice. The present work revisits the effects of individual characteristics in terms of neuroticism, classroom management self-efficacy, and cognitive (reasoning) abilities on stress using virtual reality (VR). Relying on a sample of 56 German pre-service teachers ($M_{age} = 22.73$, $SD_{age} = 4.93$; 50.9% females), we capitalized on a VR classroom environment that allowed the integration of experimental control and authentic teaching situations, where pre-service teachers responded to the disruptive behaviors of the student avatars. We focused on stress responses in terms of psychological stress (self-reported stress) assessed after the VR session, and physiological stress (heart rate) assessed during the VR session. A total of 30 (26) participants was assigned to a condition with higher (lower) levels of disruptive student behavior, referred to as higher and lower complexity condition, respectively. Results from linear mixed-effects models revealed that neuroticism positively predicted psychological and physiological stress responses in pre-service teachers, whereas classroom management self-efficacy and cognitive (reasoning) abilities were not significantly related to stress responses. Level of complexity and the interaction between complexity and individual characteristics did not have an effect. This study underlines the value of VR as a tool for psychological research and contributes to existing knowledge on teacher characteristics and stress.

Teaching can be highly stressful and demanding. Multiple studies conducted in different countries show that between 30 and 60% of teachers report experiencing high levels of stress^{1,2}. For example, the 2023 State of the American Teacher Survey found that 58% of teachers indicated that they are exposed to frequent work-related stress³. In another study, approximately 30% of teachers reported that the teaching profession is "very stressful" or "extremely stressful" (on a 5-point rating scale, ranging from "not at all stressful" to "extremely stressful"²). As compared to other professions, being a teacher has consistently been identified as one of the high-stress professions²⁻⁴. Prolonged stress is a significant health issue and has been linked to depression, anxiety disorders, and the onset of physical diseases^{5,6}. Work-related stress also represents an important factor contributing to teacher attrition⁷, with some estimates suggesting that up to half of new teachers quit their job within the first five years^{8,9}. In summary, it is important to find ways to reduce teacher stress as experiencing high levels of stress hinders individual teachers' thriving at work and negatively affects their quality of life, career-related decisions and professional well-being, and their positive interactions with students. From a socio-economic perspective, it can also be argued that teacher attrition and health problems related to teachers' work-related stress have immense financial ramifications^{10,11} and negatively affect teaching performance and students' academic success¹²; hence, decreased levels of teacher stress should lead to reduced societal costs.

Different approaches to counteract teacher stress located at different levels (context, individual) can be distinguished^{13,14}. For instance, attempts to reduce teacher stress can center on improving working conditions and limiting the number and severity of stress-inducing aspects of the work. It is also possible to focus on the individual and to help teachers develop personal resources and skills that enable them to better deal with stressful situations. Ideally, approaches to reduce teacher stress combine both the individual and contextual level (and

¹Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Walter-Simon-Straße 12, 72072 Tübingen, Germany. ²University of Potsdam, Potsdam, Germany. ³University of York, York, UK. ⁴University of Kiel, Kiel, Germany. ✉email: lisa.bardach@uni-tuebingen.de

consider their complex and dynamic interplay). However, it can, at times, be difficult or even impossible to initiate major school-level and broader socio-political changes feeding into reformations of the educational system and enhanced working conditions. Therefore, zooming in on individual characteristics relevant to stress seems worthwhile. In addition, certain features relating to the self (e.g., one's belief systems or motivations) are under individuals' control to a higher extent than are contextual features, and thus, a focus on individual features can strengthen feelings of personal agency.

Against this background, identifying individual characteristics that make teachers more versus less prone to experiencing high levels of stress becomes an important goal for psychological research with far-reaching implications for practice. For example, research can identify individual teachers who are at higher risk of suffering from stress and can inform the design of interventions for teachers to develop certain characteristics^{15,16}. The malleability of motivational and affective (teacher) characteristics has long been acknowledged¹⁷, but recent years have also seen an increase in interventions that target Big Five personality traits (e.g., conscientiousness, neuroticism, extraversion^{18,19}) which have traditionally been considered as more fixed²⁰.

The significance of individual characteristics for stress responses has been backed up by several theories and models. The Trait-State-Anxiety theory, for instance, outlines that specific personal characteristics (anxiety as a trait) prompt individuals to react with anxiety and stress in response to varying situations²¹. In a similar vein, recent conceptualizations of the Job-Demands-Resources theory¹³ acknowledge the importance of specific individual characteristics—so-called personal resources—in shaping employees' workplace experiences, including their stress responses^{22,23}. As suggested by the transactional model of stress²⁴, stress occurs if individuals perceive that external circumstances and demands exceed their personal coping resources. Thereby, Lazarus and Folkman²⁴ emphasize the interaction between the individuals and their personal characteristics and the environment. Further, according to the stress generation hypothesis, a relation between individual characteristics and exposures to stressors exists whereby individuals with certain characteristics are more likely to be exposed to stressful events and to construct events as stressful^{16,25,26}. In our study we are interested in effects of individual characteristics on stress, guided by the assumption that certain individual characteristics affect stress responses (i.e., main effects of individual characteristics). However, we also acknowledge the possibility of person x situation interactions in that individual characteristics could interact with situation characteristics (e.g., situations involving more versus fewer stressors).

The present study focused on three teacher characteristics that are conceptually and empirically relevant for teacher stress and that tap into three distinct and complementary domains of individual differences. First, neuroticism (from the Big Five personality framework^{27,28}) has been identified as a key risk factor for experiencing elevated stress levels¹⁶. Neuroticism plays an important role in determining the lens through which stressors are perceived and assigned meaning, and thus, exert effects on individual differences in how people respond to stressors. More specifically, individuals scoring high on neuroticism tend to experience higher levels of stress due to their negative filter, which makes them vulnerable to experiencing stress¹⁶.

Second, the motivation variable self-efficacy has been conceptualized as an important resource for teachers to deal with stressful situations^{23,29}. Self-efficacy captures the degree to which individuals believe they are capable of succeeding in a specific situation or reaching a specific goal³⁰. Although teachers are exposed to multiple stressors³¹, disruptive student behavior ranks high among the stress-inducing factors in teachers' professional lives^{32,33}. Therefore, higher levels of self-efficacy for classroom management³⁴ should aid in counteracting stress in teachers. Self-efficacy for classroom management, a sub-component of teaching self-efficacy, describes the extent to which a teacher feels capable of dealing with disruptions and misbehaving students, and of maintaining order in class³⁵.

Third, cognitive abilities are the most potent predictor of job performance³⁶ although research on teachers' cognitive abilities and teaching effectiveness has been less conclusive³⁷. Nonetheless, higher cognitive abilities may enable teachers to better deal with complex teaching tasks^{37,38}, which could decrease their stress levels in challenging teaching situations. In the present study, we focused on cognitive abilities in terms of reasoning abilities (assessed using a matrix reasoning task³⁹). Whereas classroom management self-efficacy and neuroticism are viable targets for interventions, we included cognitive (reasoning) abilities mainly with the aim of advancing knowledge on cognitive ability correlates in the teacher domain and not because we consider it promising to develop interventions to increase cognitive abilities⁴⁰ or select teachers and pre-service teachers into the profession or teacher education program based on their cognitive abilities.

Most research on teachers characteristics and stress has relied on survey-based studies linking teacher characteristics to their retrospectively reported psychological stress responses¹⁵. For instance, as demonstrated in a recent meta-analysis, neuroticism displayed a significant positive association with stress, which was most pronounced for self-reported psychological stress responses¹⁶. Further, an integrative review synthesized findings from existing meta-analyses and systematic reviews on teacher characteristics and concluded that teachers' self-efficacy seems consistently negatively related to psychological stress¹⁵. Moreover, laboratory research with samples other than teachers in which stress was experimentally induced has demonstrated effects of selected individual characteristics (e.g., neuroticism) on self-reported psychological stress as well as physiological stress (e.g., heart rate reactivity)⁴¹. However, even though both lines of research—retrospective survey studies and laboratory studies—have made important contributions to the current state of knowledge on (teacher) characteristics and stress, they are also limited in several regards. On the one hand, survey-based research in naturalistic settings cannot control for potential confounding factors such as differing levels of problematic student behavior in class, varying degrees of support from colleagues, and the dynamic effects of student-teacher interactions⁴². On the other hand, based on the literature on lab studies⁴¹, we argue that research on stress in the laboratory can be criticized for the reliance on stress tasks with little significance for everyday-life and the challenges that individuals encounter in their profession.

Research using virtual reality (VR) technology provides a remedy. VR, a new track in psychological research, uses digital technology to create complex, realistic, and immersive environments. These environments are under full experimental control, while enabling participants to behave naturally, thus boosting the ecological validity of the results⁴³. In educational psychology research, VR provides the perfect balance between experimental control and a teaching and learning situation that can be perceived as much more authentic than what can be achieved in a “regular” experiment^{44–46}. Accordingly, VR holds great promise for research on teachers and pre-service teachers and opens up myriad opportunities for situated practice in teacher education^{45,47}. However, despite the promise of VR, it should be acknowledged that it is still not a “real” teaching situation. Further, although a VR scenario may aim to confront participants with certain situation characteristics, such as authenticity and complexity, whether the participant does (or does not) subjectively construct the situation as authentic and complex is the results of a subjective interpretation or evaluation of the situation⁴⁸. Importantly for the present study, the usefulness of VR for research on teacher stress has recently been demonstrated in a number of studies. Relying on a VR classroom setting, Huang et al. showed that class size (number of student avatars) influenced psychological and physiological (heart rate) stress responses in pre-service teachers, with higher stress levels in the condition with a larger class size⁴⁵. However, studies exploiting the potential of VR to revisit effects of teacher characteristics on stress are currently lacking.

The purpose of the present study was therefore to examine the extent to which pre-service teachers’ neuroticism, classroom management self-efficacy, and cognitive (reasoning) abilities predict stress responses in a VR teaching session. Stress, a multidimensional phenomenon, manifests itself on the physiological level as well as on the level of subjective experience⁴⁹; hence, in our study, we considered both psychological (self-reported stress) and physiological stress responses (heart rate). As students’ disruptive behavior and discipline problems in class are among the most critical stressors for teachers, and especially for new teachers³², the set-up of the VR classroom was well-suited to study effects on stress. Roughly half of the sample was assigned to a condition with higher levels of disruptive student behavior and the other half of the sample was assigned to a condition with lower levels of disruptive student behavior ($n_{low} = 26$, $n_{high} = 30$). These two complexity levels differed only in the number, severity, and concurrences of disruptions, allowing for insights into whether effects of teacher characteristics on stress generalize across situations with more versus less complexity, and to test individual characteristics x condition interactions. Both conditions have been justified in previous studies^{44,45} for subjective perceptions of authenticity and complexity with samples of similar demographics and teaching tasks alike. Aside from the levels of disruptive student behavior, all other contextual factors remained consistent between the two conditions.

Based on theory and prior empirical findings, we hypothesized that neuroticism should make pre-service teachers more vulnerable to experiencing stress. Hence, we hypothesized that neuroticism should positively predict physiological and psychological stress (i.e., positive stress-exacerbating effects)². By contrast, we assumed that classroom management self-efficacy and cognitive (reasoning) abilities should function as personal resources that make individuals less likely to experience stress. Specifically, we hypothesized that classroom management self-efficacy^{23,29} and cognitive (reasoning) abilities^{37,38} should negatively predict physiological and psychological stress (i.e., negative stress-reducing effects). In addition, we examined effects of situational characteristics (complexity, i.e., higher versus lower levels of disruptive student behavior) on stress and tested whether individual characteristics interact with situation characteristics²⁴.

Methods

Participants

The current study focuses on pre-service teachers who are enrolled in university-based teacher education programs and have little prior teaching experience. Participants were recruited from a weekly seminar on classroom management held at a public German university. The seminar was a regular course offered every semester and we used data from two cohorts from the same academic year. The seminar forms part of the module on educational theories in the teacher training program; it did not contain a field practicum component and therefore provided no opportunity for real-life teaching ($N = 56$, $M_{age} = 22.73$, $SD_{age} = 4.93$; 50.9% female, 98.2% Bachelor students, $M_{teaching\ hour} = 4.2$). These demographic features are representative of this population^{44,50}. Due to its accessibility and representation of the target population, this sample was chosen. Furthermore, a focus on pre-service teachers is particularly valuable as research in this population can inform possible preventive measures that could be taken early on because they are yet to start their careers. Although the VR session was presented as an integral part of the seminar, participants could choose freely whether to participate in the VR session and the study with no further incentives. The study was approved by the ethnic review board of the university. Informed consent for study participation was obtained from all participants.

VR environment and equipment

Our VR classroom mimics a typical secondary classroom in Germany with a class size of 30 student avatars⁵¹. Student avatars possess diverse physical characteristics (see Fig. 1), and their behaviors were scripted (see Supplementary Table S1). Pre-service teacher participants (with similar demographic characteristics as our sample) from previous studies^{44,45} indicated that this VR classroom was authentic and convincing. Participants experienced the VR classroom through the HTC VIVE Pro Eye system which allows users to move freely in the real-world environment. The headset has a display resolution of 1440×1600 pixels per eye, with a 110° field of view.

Procedure and measures

Participants completed a questionnaire containing measures for classroom management self-efficacy, neuroticism, and worked on a test assessing cognitive (reasoning) abilities prior to the VR session (see Fig. 2). *Classroom management self-efficacy* was assessed using four items from the German Teacher’s Sense of Efficacy Scale⁵² (e.g.,



Figure 1. View in the VR Classroom. All views were displayed from the participant’s visual perspective. Top: front view from the teacher’s desk; bottom left: student avatar performs off-task behavior; bottom right: on-task behavior.

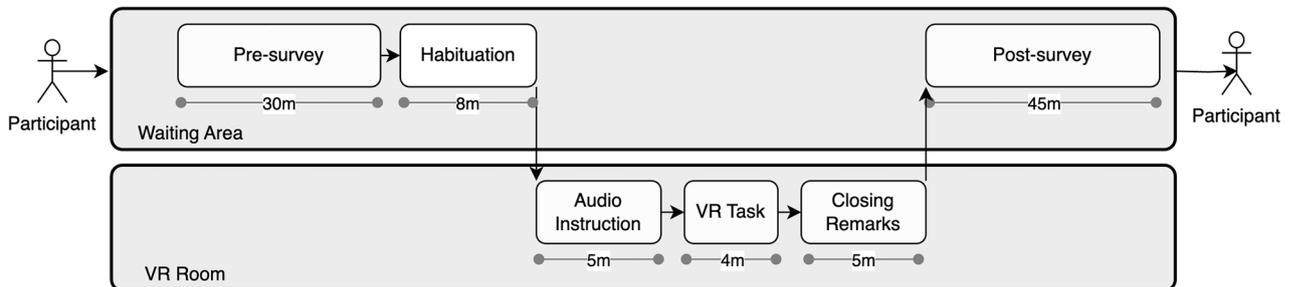


Figure 2. Experiment procedure. Physiological stress (heart rate) was continuously measured from habituation until the end of the VR task.

“I am confident in my ability to control disruptive behavior in the classroom.”; $\alpha = .81$), which were rated on a 5-point scale (1 = not convinced at all to 7 = fully convinced). We measured *neuroticism* using the three-item scale from the Big Five Inventory-SOEP⁵³ (e.g., “I see myself as somebody who worries a lot,” $\alpha = .69$). The items were rated on a 7-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree. We used the matrix reasoning test from the International Cognitive Ability Resource³⁹ (11 items, sum score of correct answers) as a proxy of participants’ *cognitive (reasoning) abilities* (for large validation studies, see ^{39,54}). The test contains Matrix stimuli that are similar to those used in Raven’s Progressive Matrices^{39,54}. Stimuli are 3×3 arrays of geometric shapes, and one of the nine shapes is missing. Participants were instructed that they should identify which of the six geometric shapes that are presented as response choices will best complete the stimuli. Participants also reported their semesters of study (“What semester of your teaching degree program are you in?”) and *VR experience* (“What experience do you have with virtual reality?”, response options: 1 = none, 2 = a bit, 3 = a lot) which

were used as random effects in the linear mixed-effects model (see Analyses below for details). Gender was also reported and used as a covariate to control for potential gender differences in stress perceptions. The participants had 30 min time to complete the measures reported above, including the cognitive abilities (reasoning) tests.

During the individual VR teaching session that took place in a quiet room with one experimenter, participants first followed a 5-min audio instruction to familiarize themselves with the VR classroom. They then gave a 4-min lecture about COVID-19 vaccinations during which they needed to respond to typical classroom disruptions. The lecture topic and content were pre-selected and the same for all participants, who were given the presentation slides and lesson plan a week before the experiment. This procedure was designed based on piloting and prior studies⁴⁵. Student avatars would engage in both on- and off-task behaviors (Fig. 1). The disruptions were chosen from a collection of common disruptive behaviors that were evaluated by teachers as being obviously unrelated to the lesson and causing substantial interruptions to the class^{55,56}. Avatar behaviors were independent from the teachers' actions. Participants could react to disruptions by moving closer to students, calling their names, making eye contact, or in any way they deemed suitable.

Participants in the first semester (i.e., the first cohort) were assigned to VR scenarios of lower complexity level while the participants from the second semester (i.e., the second cohort) were assigned to a higher complexity level ($n_{low} = 26$, $n_{high} = 30$). The high and low complexity level differed in the a) total number of disruptions (30 vs. 15); b) the percentage of significant disruptions, such as hitting and throwing paper balls (10 out of 30 vs. 3 out of 15); and c) the number of occasions when multiple disruptions occurred at the same time (5 vs. 1). Aside from the complexity of the classroom management scenarios, all other aspects remained consistent between two levels. Combining both complexity conditions within one study increased statistical power due to the larger sample and allowed to explore whether effects of cognitive and non-cognitive characteristics on stress generalize across situations with more versus less stressors.

After the VR session, participants rated their immediate psychological stress responses associated with the VR session using four items from Dundee Stress State Questionnaire⁵⁷ (e.g., "I felt tense," $\alpha = .90$). The items were rated on a 7-point scale, ranging from 1 = strongly disagree to 7 = strongly agree. To capture *physiological stress*, we measured the average heart rate (HR) in beats per minute (BPM) during the VR teaching task after controlling for the baseline HR for each participant. Specifically, HR was measured at 0.3 s intervals using Polar OH1—an optical HR sensor worn around the arm that has been validated with electrocardiography⁵⁸. We evaluated each participant's baseline HR to adjust for individual variability in HR induced by variables unrelated to the experimental manipulation⁵⁹. The baseline HR was the average BPM during the last 3 min of an 8-min standardized habituation phase⁶⁰ in which participants were told to sit quietly, avoid moving, and keep their hands on the chair and feet flat on the floor. Finally, the average HR during the VR teaching task was centered on the baseline HR to obtain the measure of physiological stress. HR has already been successfully examined and discussed as a valid measure of stress in various studies⁶¹.

Analyses

Linear mixed effects models

We relied on linear mixed-effect models to address our hypotheses. In a traditional linear model, the response is modeled as a linear combination of predictors with weights that are referred to as effects or coefficients⁶². In contrast, a linear mixed-effects model (LMM) takes into account both fixed and random effects, in other words, effects that are constant or varying for groups in a population^{63,64}. As reviewed by Kliegl and colleagues⁶⁵, LMM is superior to analyses of variance (ANOVAs) because of its advantages such as tolerance of an unbalanced experimental design due to missing data⁶⁶, and researchers' capability to examine effects of both categorical factors and continuous variables (covariates), as well as their interactions, with a significant gain in statistical power^{64,67}. LMM is particularly robust against missing data. In a simulation with 1000 runs of a mixed-effects model, Baayen and colleagues⁶⁷ found that the main effect was almost always detected with and without 20% missing data (missing at random). Its power is only minimally reduced in the situation of missing data. Despite the fact that power is at its maximum, the Type I error rate is within the nominal range. To use LMM, we examined the Q-Q plot of the residuals as well as the plot of residuals plotted against fitted values. We found that the homoscedasticity and normality assumption was met.

In the present study, we treated effects of classroom complexity, neuroticism, classroom management self-efficacy, and cognitive (reasoning) abilities as fixed effects, while semester and VR experience were treated as random effects. Semester and VR experience in this design were "random" in the sense that they were assumed to be a random sample of the semester and VR experience level that might have been included in the study⁶⁸. By modeling random effects along with the fixed effects, we can obtain an estimate of the mean difference in complexity, neuroticism, classroom management self-efficacy, and cognitive (reasoning) abilities; as well as an estimate of the variances surrounding that difference, due to the individual features of each semester or VR experience⁶⁹. Models were fitted with restricted maximum likelihood (REML) using the lmer function from the lme4 package⁷⁰ in R 4.1.2⁷¹. We included Akaike Information Criterion (AIC; decreases with goodness of fit) and the Bayesian Information Criterion (BIC; decreases with goodness of fit) as goodness of fit indices⁷¹.

Models tested in the current study

We constructed and tested three sets of LMM models for each stress reaction: Model 1 (M1) included only the fixed effects of interested variables; Model 2 (M2) added the control variable of gender; Model 3 (M3) additionally included the interaction term neuroticism/cognitive (reasoning) abilities/self-efficacy \times complexity to examine the potential effects of individual characteristics \times situational conditions. To clarify whether the introduction of further components into the model would significantly improve model fit, we conducted the likelihood ratio (LR) test to compare the three models.

Effect size calculation for fixed-effects terms

The effect size estimate (d) that extended Cohen's d^{73} to mixed-effects model was conceived by Judd and colleagues⁶⁸. This estimate might be utilized for designs where the two random factors are crossed and designs where one random factor is nested within the other. Please see Equation 5 in Judd and colleagues' original paper for the mathematical expression of this definition. An example of the calculation can be seen in Brysbaert and Stevens's⁷⁴ replication study.

Missing data handling

Using Little's⁷⁵ missing completely at random (MCAR) test, we performed missing value analyses to uncover potential patterns in missing data that might influence the analysis. In the case of a non-significant Little's MCAR test, observed data is considered to be a random sample from all data, allowing for a complete-case analysis with no bias imposed⁷⁶. In the current study, we obtained incomplete HR recordings in 8 observations due to technological failure (14.29%) and these were excluded from the analysis. Same for psychological stress, one participant was excluded from the analysis due to an incomplete record. MCAR test yielded non-significant results for all variables included in our models, therefore the data were considered to be missing completely at random ($\chi^2 = 33.27$, $df = 8$, $p = .21$). Additionally, the missing data is not likely to bias the estimation due to the robustness of LMM against missing data⁶⁷.

Ethical approval statement

This study adhered to all national and international regulations for protecting human subjects. The study was approved by the ethnic review board of the University of Potsdam. Informed consent for study participation was obtained from all participants.

Results

Descriptive statistics and bivariate correlations of all variables are shown in Table 1. Data were analyzed using linear mixed-effects modeling (LMM) to evaluate the main effect of neuroticism, classroom management self-efficacy, cognitive (reasoning) abilities, and classroom complexity on stress (fixed effects), while accounting for covariate effects that could be generalized to other semesters and VR experiences (random effects; see Methods for details). Three LMM models were tested for physiological and psychological stress responses, respectively: Model 1 (M1) included only the fixed effects of the variables of interest. In Model 2 (M2), we added the control variable of gender. Lastly, Model 3 (M3) additionally contained the interaction terms neuroticism/cognitive (reasoning) abilities/self-efficacy \times complexity. Likelihood ratio (LR) tests were conducted to compare the three models and to examine whether the introduction of further components into the model significantly improves model fit. Effect sizes were calculated following standard procedures for mixed-effects models⁷⁴ (see Methods for details). All analysis files and the anonymized data set are available at the Open Science Framework (OSF, <https://osf.io/vkbwx/>).

Effects on physiological stress

First, we compared the goodness of fit between models. The change in log-likelihood was not significant when comparing M2 to M1 ($\Delta\chi^2(2) = 4.06$, $p = .19$), and M3 to M2 ($\Delta\chi^2(2) = 1.86$, $p = .39$), indicating that the goodness of fit did not significantly improve with the inclusion of the covariate of gender ($\beta = 0.15$, $p = .59$), nor the interaction terms of neuroticism/self-efficacy/cognitive (reasoning) abilities \times complexity ($\beta = -0.35 / -0.24 / 3.50$, $p = .26 / .47 / .36$). Thus, we chose to use M1 as the final model.

Next, we examined participants' physiological stress during the teaching task in VR. Participants' HR during this task was on average 16.90 BPM higher than the baseline HR (Table 1). The effects of neuroticism, classroom management self-efficacy, cognitive (reasoning) abilities, and complexity were examined with M1. As shown in the model summary (Table 2, also see Supplementary Table S2 for the full model summary), neuroticism had a significant and positive effect on physiological stress ($\beta = 0.40$, $p = .01$) with a large effect size ($d = 0.76$). This effect is visually represented in Fig. 3. We did not find statistically significant evidence regarding the main effects

Variable name	Range (scale range)/category	$M(SD)$	Missing (%)	1	2	3	4	5	6	7
1. Semester	1–21	4.76 (3.68)	0 (0%)		.14	.20	.06	.24	.01	–.08
2. VR experience	1–3 (1–3)	1.31 (0.51)	1 (0.02%)			–.07	.03	.17	–.02	–.10
3. ¹ Neuroticism	1.33–4.67 (1–5)	2.94 (0.78)	0 (0%)				–.38**	.10	.01	.23
4. ¹ Self-efficacy	1.25–6.00 (1–7)	3.87 (1.23)	0 (0%)					.26*	.49***	–.08
5. ¹ Reasoning abilities	1–10 (1–10)	5.95 (2.39)	1 (0.02%)						.35**	.13
6. ¹ Psychological stress	1.50–6.50 (1–7)	3.49 (1.29)	0 (0%)							.10
7. ^{1,2} Physiological stress	–10.72–62.40	16.90 (14.23)	8 (14.29%)							

Table 1. Descriptive statistics of all included variables and bivariate correlations (fields 1–7=). $N = 56$. ²Variables displayed in original scale in this table but were z-score normalized before entering in the mixed model. ²Physiological stress was measured using participants' heart rate and is the average BPM (beats per minute) during VR instruction after controlling for baseline heart rate. * $p < .05$; ** $p < .01$; *** $p < .001$.

Terms	¹ Physiological stress			² Psychological stress		
	$\hat{\beta}$	$SE(\hat{\beta})$	<i>t</i>	$\hat{\beta}$	$SE(\hat{\beta})$	<i>t</i>
Intercept	-0.43	0.50	-0.86	-0.10	0.18	-0.51
Complexity (large–small)	0.26	0.31	0.83	-0.13	0.28	-0.47
Neuroticism	0.40	0.15	2.59**	0.29	0.14	2.13*
Self-efficacy in classroom management	-0.41	0.43	-0.95	-0.28	0.18	-1.62
Cogn. abilities	-0.06	0.19	-0.32	0.002	0.14	0.01

Table 2. LMM model summary of Model 1. * $p < .05$. ** $p < .01$, *** $p < .001$. ¹ $n = 47$, AIC = 150.05, BIC = 178.49; ² $n = 55$, AIC = 169.89, BIC = 202.47. Cogn. abilities = Cognitive (reasoning) abilities. Physiological stress was operationalized by heart rate.

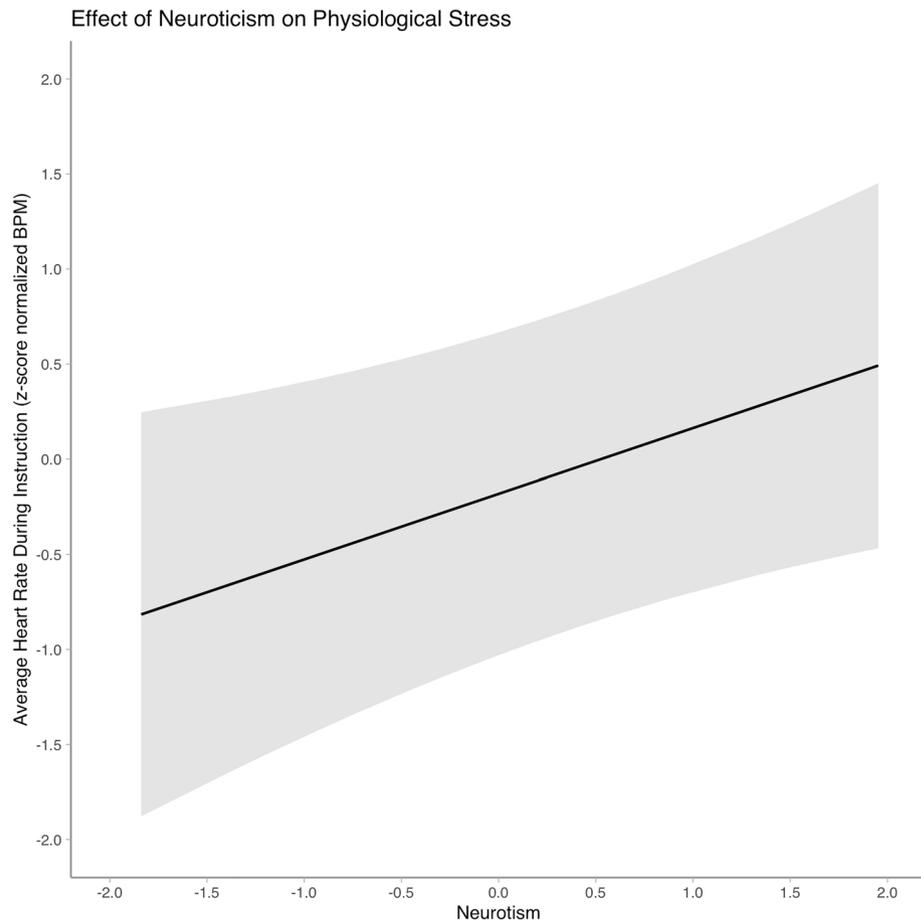


Figure 3. Effect of neuroticism on physiological stress (operationalized by heart rate).

of classroom management self-efficacy ($\hat{\beta} = -0.41$, $p = .50$), cognitive (reasoning) abilities ($\hat{\beta} = -0.06$, $p = .79$), and complexity ($\hat{\beta} = 0.26$, $p = .42$).

Effects on psychological stress

Similarly, for psychological stress, the goodness of fit did not differ significantly between M2 and M1 ($\Delta\chi^2(2) = 1.77$, $p = .41$), as well as between M3 and M2 ($\Delta\chi^2(2) = 2.15$, $p = .29$). The effect of the covariate of gender ($\hat{\beta} = 0.36$, $p = .16$), and the interaction terms neuroticism/self-efficacy/cognitive (reasoning) abilities \times complexity ($\hat{\beta} = 0.24/0.72/-0.14$, $p = .42/.08/.66$) were not significant in relation to psychological stress. Therefore, we chose M1 as the final model to examine the fixed effects of neuroticism, classroom management self-efficacy, cognitive (reasoning) abilities, and complexity on psychological stress.

As shown in Table 2 (also see Supplementary Table S3 for the full model summary), similar to physiological stress, neuroticism had a significant and positive effect on psychological stress ($\hat{\beta} = 0.29$, $p = .04$) with a medium to large effect size ($d = 0.61$). Figure 4 portrays this effect visually. The main effects of classroom management

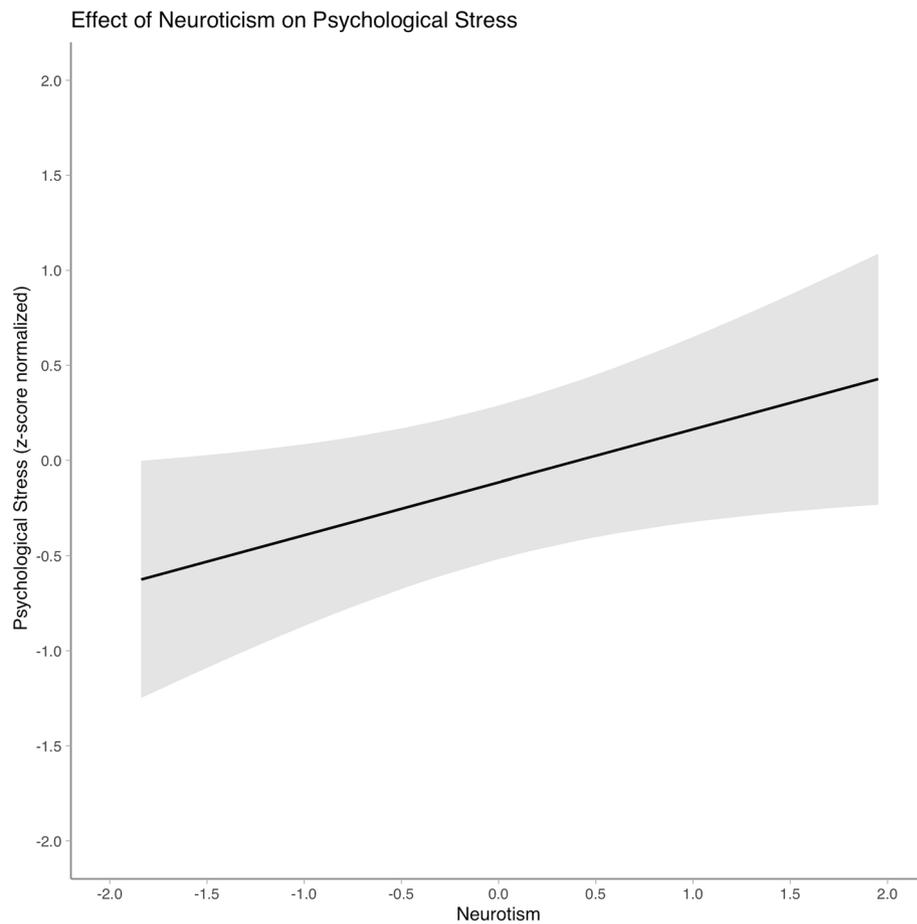


Figure 4. Effect of neuroticism on psychological stress.

self-efficacy ($\hat{\beta} = -0.28, p = .16$), cognitive (reasoning) abilities ($\hat{\beta} = 0.002, p = .99$), and of complexity ($\hat{\beta} = -0.13, p = .65$) were not statistically significant.

Discussion

Capitalizing on a VR classroom environment, this study aimed to investigate effects of pre-service teachers' individual characteristics in terms of neuroticism, classroom management self-efficacy, and cognitive (reasoning) abilities on psychological and physiological stress responses. The results showed that neuroticism positively predicted psychological and physiological stress, whereas cognitive (reasoning) abilities and classroom management self-efficacy were not significantly related to either form of stress. We draw four major conclusions from our findings.

First, the results underscore the importance of neuroticism in research on stress¹⁶. As hypothesized, higher levels of neuroticism made pre-service teachers more vulnerable to experiencing stress in a fully standardized VR experimental environment. Educational psychologists are sometimes reluctant to include Big Five personality traits such as neuroticism in their research⁷⁷. This likely stems from the misconception that Big Five personality traits are, unlike motivational constructs, stable, and therefore not a useful target for interventions. However, mounting evidence from personality psychology indicates that Big Five personality dimensions change over the lifespan⁷⁸, are malleable, and can be altered via interventions, if people are motivated to change specific personality aspects^{18,19}.

Our findings hold implications for research as they point towards the promise of focusing on neuroticism in research on stress in challenging teaching situations. The findings also have important implications for educational practice, for example, for teacher education programs and interventions for practicing teachers aiming to counteract neuroticism. People do, in general, want to become less neurotic. To illustrate, of all participants who signed up for a Big Five personality change intervention, most participants wanted to decrease in neuroticism (26.7%)¹⁹. Personality change interventions, which can be delivered digitally, could thus be used to support pre-service teachers and teachers to achieve their personal personality-change goals, such as reducing neuroticism levels¹⁹. Big Five personality change interventions typically include microinterventions (in terms of specific tools and techniques) that assist people in modifying or changing behaviors and experiences, and to maintain the change process^{18,19}. In addition, mindfulness interventions have been shown to decrease neuroticism⁷⁹, and could be more systematically integrated into teacher education programs and implemented in schools.

Second, there are theoretically sound reasons to assume that cognitive abilities should matter for how individuals, including teachers, deal with challenging situations^{37,38}, which could affect their stress responses in such situations. Hence, we had hypothesized that higher cognitive (reasoning) abilities should predict lower levels of stress. However, according to our findings, pre-service teachers' cognitive (reasoning) abilities did not play a relevant role for their psychological or physiological stress responses. Nonetheless, this finding adds to the scarce evidence base on teacher cognitive abilities and their correlates^{37,80}. Future VR research on stress following up on our study could focus on potential links between teachers' cognitive abilities and specific tasks within the VR environment that are even more "cognitively loaded" than the teaching task employed in our study (e.g., diagnostic decisions³⁸).

Third, self-efficacy beliefs, and especially teaching-specific self-efficacy beliefs, have been portrayed as one of the most relevant teacher characteristics (e.g., regarding links to teachers' occupational well-being, instructional approaches, and relations to student outcomes^{81–83}) and have been conceptualized as personal resources that can counteract stress^{23,83}. In our study, we had thus hypothesized that self-efficacy should be negatively related to stress. However, we did not obtain a significant effect of classroom management self-efficacy on stress experienced in the VR session. The reasons underlying this finding are hard to establish⁸⁴. For instance, it could be that some of those pre-service teachers with higher classroom management self-efficacy also had higher expectations regarding their performance in the challenging teaching situation, which may have even caused stress. It may also be that classroom management self-efficacy does not exert an immediate effect within a stressful situation, which could have been captured in our study. Instead, effects of classroom management self-efficacy on stress responses may take longer to unfold and may to a larger extent depend on the broader social context⁸². More research systematically contrasting effects of self-efficacy on stress in "real-life" longitudinal studies¹⁴ and in VR environments is needed to gain clarity. Nonetheless, jointly, the non-significant effect for classroom management self-efficacy and the significant effect for neuroticism indicate that general behavioral tendencies (neuroticism) and not teaching-related characteristics (classroom management self-efficacy) seem more relevant to stress, even in a VR situation with a clear connection to classroom management (i.e., dealing with disruptive students). Against the background that research on teachers often strongly favors teaching-specific features, this is an important insight, which was made possible by the inter-theoretical integrative approach of our study focusing on individual characteristics from different research traditions. It is for future research to replicate our findings and to figure out whether neuroticism (and not teaching-specific self-efficacy beliefs) consistently predicts stress in other teaching situations.

Fourth, it should be mentioned that being assigned to one of the two complexity conditions (high versus low levels of disruptions) did not significantly predict stress. Further, we did not obtain significant interactions between individual characteristics and complexity levels. These findings speak to the fact that the investigated characteristics affected stress responses that would be generalizable to instructional situations with different complexity levels. Overall, our results are more in line with unidirectional perspectives on stress²¹ than with bidirectional perspectives presuming interactions between the person and the environment²⁴, but replications using the same as well as combinations of other stressors are clearly required. Additionally, the finding that psychological and physiological stress reactions were low in correlation (see Table 1) but shared similar relationships with preservice teachers' characteristics corresponds with earlier findings⁸⁵ indicating that measuring stress with indicators of different modes is key for effects validation.

Fifth, our study underlines the value of using VR in psychological research in general and research on teacher characteristics and stress in particular. Prior survey-based research in "naturalistic" settings can be criticized for not being able to account for many confounding factors that may cause teacher stress, in addition to effects of individual characteristics. On the other hand, research in the lab often relies on artificial stress tasks that do not resemble the tasks that individuals encounter in their work life on a regular base and lack ecological validity. Hence, the reliance on VR in the present study helped to overcome significant limitations of prior research on (teacher) stress by balancing the teaching tasks with the precision afforded by experimental control⁴⁴, thus combining the best from both worlds⁴³. Moreover, it is interesting to note that in our study, neuroticism was significantly related to both psychological and physiological stress, whereas the meta-analysis by Luo et al.¹⁶ revealed significant links between neuroticism and psychological stress responses, but not physiological stress responses. Of course, replications of our work are needed; however, our findings may serve as an indication that VR environments that allow participants to experience work-like situations are particularly well suited to study effects of neuroticism on stress.

Several limitations and promising directions for future research should be noted. Our focus on the specific characteristics investigated in this study necessarily excluded other potentially relevant characteristics (e.g., other motivational constructs, such as causal attributions or interest, other Big Five personality traits¹⁶). In addition, even though the inclusion of both psychological and physiological stress responses is a strength of our study, considering a range of conventional stress measures (e.g., for physiological stress: heart rate variability indicators and salivary cortisol levels^{8,86}) can provide an even more comprehensive understanding of the link between individual characteristics and stress. The dynamic nature of HR also calls for finer examinations of changes over time to capture time-lagged effects and the associations between time-varying contextual conditions with the cardiovascular data. One of the limitations of measuring HR in a room-scale VR simulation was the confounding influence of mental process and physical movement on HR measures. Future study may replicate the effect with more stationary setup to avoid potential movement artifacts. Further, we relied on measuring acute psychophysiological stress reactions in one short VR session, and it would have been desirable to investigate the exposure to multiple challenging VR situations and related changes in stress responses over time. Chronic stress reactions, which are long-term physiological and biochemical changes together with psychosomatic symptoms⁸⁷ are also important to study in relation to individual characteristics. Relatedly, conceptual replications of our work using ecological momentary assessments to assess teachers' daily hassles in naturalistic classroom settings⁸⁸ could

profitably complement VR-based research on stress. Also, the association between stress and teacher characteristics is not yet established for in-service teachers because the sample solely consisted of pre-service teachers who had not yet taught in actual classrooms. Further, some of our constructs were assessed with a few items only and had reliabilities slightly below .70. Hence, future research should use longer and more differentiated measures (e.g., distinct neuroticism facets). Another issue worth noting is that in our study classroom management self-efficacy and cognitive (reasoning) abilities had relatively lower standard deviations, which may have impacted the ability to detect significant effects. The cognitive abilities (reasoning) test was administered on the same date as the VR session. Even though it was not a high-stakes test situation (and no time limit was set for the test alone) and we do not think that the test caused a lot of stress, we cannot rule out that the test evoked stress in (some) participants. Hence, future VR research on cognitive abilities could administer the test at a separate day (or systematically vary time of administration to empirically examine whether working on a cognitive test prior to the VR session significantly affects stress). Finally, future work may want to adopt a within-person design in which participants complete both levels of complexity. However, in our study, we opted for a between-person design to reduce the burden for participants due to limited space and time resource, and because the issue relating to high versus low levels of complexity was not the main interest of our study. In our study, we were mainly interested in the effects of teacher characteristics on physiological and psychological stress.

To conclude, VR opens unprecedented opportunities for psychological research and the study of teaching and learning processes^{89,90}. Our study adds to this emerging line of research by demonstrating that pre-service teachers' neuroticism predicted psychological and physiological stress in a VR classroom. Classroom management self-efficacy and cognitive (reasoning) abilities were not significantly related to stress. Our results provide a basis for future studies to further exploit the potential of VR in order to examine characteristics relevant to teachers' stress responses.

Data and code availability

The anonymized dataset and all analysis files are available on the Open Science Framework (OSF, <https://osf.io/vkbwx/>). The source code of our virtual reality classroom (https://gitup.uni-potsdam.de/mm_vr/vr-klassenzimmer) is shared under GNU Affero General Public License.

Received: 3 May 2023; Accepted: 8 December 2023

Published online: 14 December 2023

References

- Gallup. *State of America's schools: The path to winning again in education*. <https://www.gallup.com/education/269648/state-america-schools-report.aspx> (2014).
- Kyriacou, C. Teacher stress and burnout: Methodological perspectives in *International Encyclopedia of the Social & Behavioral Sciences* (eds. Wright, J. D.) 72–74 (Elsevier, 2015).
- Doan, S., Steiner, E. D., Pandey, R., & Woo, A. Teacher well-being and intentions to leave: Finding from the 2023 State of the American Teacher Survey. *American Educator Panels*. https://www.rand.org/pubs/research_reports/RRA1108-8.html (RAND Corporation, 2023).
- Schaarschmidt, U. & Kieschke, U. Gerüstet für den Schulalltag in *Psychologische Unterstützungsangebote für Lehrerinnen und Lehrer [Ready for school life in psychological supporting proposals for teachers]* (Beltz, 2007).
- Cohen, S., Janicki-Deverts, D. & Miller, G. E. Psychological stress and disease. *JAMA* **298**, 1685 (2007).
- O'Connor, D. B., Thayer, J. F. & Vedhara, K. Stress and health: A review of psychobiological processes. *Annu. Rev. Psychol.* **72**, 663–688. <https://doi.org/10.1146/annurev-psych-062520-122331> (2021).
- Klassen, R. M. & Chiu, M. M. Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *J. Educ. Psychol.* **102**, 741–756 (2010).
- Sims, S. & Jerrim, J. TALIS 2018: teacher working conditions, turnover and attrition. *Statistical working paper* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/873922/Teaching_and_Learning_International_Survey_2018_March_2020.pdf (2020).
- Alliance for Excellent Education. *Teacher attrition costs United States up to \$2.2 billion annually, says new alliance report*. https://all4ed.org/press_release/teacher-attrition-costs-united-states-up-to-2-2-billion-annually-says-new-alliance-report/ (2017).
- OECD. TALIS 2018 Results (Volume II) Teachers and School Leaders as Valued Professionals. <https://doi.org/10.1787/19cf08df-en> (OECD Publishing, 2020).
- Künzi, K. & Oesch, T. Berufsbedingte Krankheitskosten der Lehrpersonen. Teilprojekt des LCH im Rahmen des Projekts Gesundheit der Lehrpersonen. Schlussbericht [Work-related illness costs of teachers. Sub-project of the LCH within the framework of the project Health of Teachers Final report]. https://www.buerobass.ch/fileadmin/Files/2016/LCH_2016_KrankheitskostenLehrpersonen.pdf (2016).
- Schleicher, A. Valuing our teachers and raising their status: How communities can help. <https://doi.org/10.1787/9789264292697-en> (OECD Publishing, 2018).
- Demerouti, E., Bakker, A. B., Nachreiner, F. & Schaufeli, W. B. The job demands- resources model of burnout. *J. Appl. Psychol.* **86**, 499–512. <https://doi.org/10.1037/0021-9010.86.3.499> (2001).
- Jögi, A., Aulén, A., Pakarinen, E. & Lerkkanen, M. Teachers' daily physiological stress and positive affect in relation to their general occupational well-being. *Br. J. Educ. Psychol.* **93**, 368–385. <https://doi.org/10.1111/bjep.12561> (2023).
- Bardach, L., Klassen, R. M. & Perry, N. E. Teachers' psychological characteristics: Do they matter for teacher effectiveness, teachers' well-being, retention, and interpersonal relations? An integrative review. *Educ. Psychol. Rev.* **34**, 259–300. <https://doi.org/10.1007/s10648-021-09614-9> (2022).
- Luo, J., Zhang, B., Cao, M. & Roberts, B. W. The stressful personality: A meta-analytical review of the relation between personality and stress. *Pers. Soc. Psychol. Rev.* **27**, 128–194. <https://doi.org/10.1177/10888683221104002> (2022).
- Tschannen-Moran, M. & McMaster, P. Sources of self-efficacy: Four professional development formats and their relationship to self-efficacy and implementation of a new teaching strategy. *Elem. Sch. J.* **110**, 228–245. <https://doi.org/10.1086/605771> (2009).
- Allemand, M. & Flückiger, C. Personality change through digital-coaching interventions. *Curr. Dir. Psychol. Sci.* **31**, 41–48. <https://doi.org/10.1177/09637214211067782> (2022).
- Stieger, M. et al. Changing personality traits with the help of a digital personality change intervention. *Psychol. Cogn. Sci.* **118**, e2017548118. <https://doi.org/10.1073/pnas.2017548118> (2021).

20. Jach, H., Bardach, L. & Murayama, K. How personality matters for education research. *Educ. Psychol. Rev.* **35**, 94. <https://doi.org/10.1007/s10648-023-09807-4> (2023).
21. Spielberg, C. Anxiety as an emotional state in *Anxiety: Current Trends in Theory and Research* (ed. Spielberg, C.) 35–46 (Academic Press, 1972).
22. Xanthopoulou, D., Bakker, A. B., Demerouti, E. & Schaufeli, W. B. The role of personal resources in the job demands-resources model. *Int. J. Stress Manag.* **14**, 121–141. <https://doi.org/10.1037/1072-5245.14.2.121> (2007).
23. Granziera, H., Collie, R. & Martin, A. Understanding teacher wellbeing through job demands-resources theory in *Cultivating Teacher Resilience* (ed. Mansfield, C. F.) 229–244; https://doi.org/10.1007/978-981-15-5963-1_14 (Springer Singapore, 2020).
24. Lazarus, R. S., & Folkman, S. *Stress, Appraisal, and Coping*. (Springer, 1984).
25. Bolger, N. & Zuckerman, A. A framework for studying personality in the stress process. *J. Pers. Soc. Psychol.* **69**, 890–902. <https://doi.org/10.1037/0022-3514.69.5.890> (1995).
26. Hammen, C. Stress generation in depression: Reflections on origins, research, and future directions. *J. Clin. Psychol.* **62**, 1065–1082. <https://doi.org/10.1002/jclp.20293> (2006).
27. Costa, P. T. & McCrae, R. R. Revised NEO personality inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. *Psychol. Assess. Resour.* **396** (1992).
28. Roberts, B. W. Back to the future: Personality and assessment and personality development. *J. Res. Pers.* **43**, 137–145. <https://doi.org/10.1016/j.jrp.2008.12.015> (2009).
29. Dicke, T., Stebner, F., Linninger, C., Kunter, M. & Leutner, D. A longitudinal study of teachers' occupational well-being: Applying the job demands-resources model. *J. Occup. Health Psychol.* **23**, 262–277. <https://doi.org/10.1037/ocp0000070> (2018).
30. Bandura, A. *Self-efficacy: The exercise of control*. (W. H. Freeman and Company, 1997).
31. Kärner, T. & Höning, J. Teachers' experienced classroom demands and autonomic stress reactions: Results of a pilot study and implications for process-oriented research in vocational education and training. *Empir. Res. Vocat. Educ. Train.* **13**, 8. <https://doi.org/10.1186/s40461-021-00113-3> (2021).
32. Scherzinger, M. & Wettstein, A. Classroom disruptions, the teacher–student relationship and classroom management from the perspective of teachers, students and external observers: A multimethod approach. *Learn. Environ. Res.* **22**, 101–116. <https://doi.org/10.1007/s10984-018-9269-x> (2019).
33. Emmer, E. T. & Stough, L. M. Classroom management: A critical part of educational psychology, with implications for teacher education. *Educ. Psychol.* **36**, 103–112. <https://doi.org/10.4324/9781410608130-5> (2001).
34. Tschannen-Moran, M. & Hoy, A. W. Teacher efficacy: Capturing an elusive construct. *Teach. Teach. Educ.* **17**, 783–805. [https://doi.org/10.1016/S0742-051X\(01\)00036-1](https://doi.org/10.1016/S0742-051X(01)00036-1) (2001).
35. Emmer, E. T. & Hickman, J. Teacher efficacy in classroom management and discipline. *Educ. Psychol. Meas.* **51**, 755–765. <https://doi.org/10.1177/0013164491513027> (1991).
36. Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A. & Rich, B. L. Self-efficacy and work-related performance: The integral role of individual differences. *J. Appl. Psychol.* **92**, 107–127. <https://doi.org/10.1037/0021-9010.92.1.107> (2007).
37. Bardach, L. & Klassen, R. M. Smart teachers, successful students? A systematic review of the literature on teachers' cognitive abilities and teacher effectiveness. *Educ. Res. Rev.* **30**, 100312. <https://doi.org/10.1016/j.edurev.2020.100312> (2020).
38. Kaiser, J., Helm, F., Retelsdorf, J., Südkamp, A. & Möller, J. Zum Zusammenhang von Intelligenz und Urteilsgenauigkeit bei der Beurteilung von Schülerleistungen im Simulierten Klassenraum [On the relationship between intelligence and judgment accuracy in assessing student performance in the simulated classroom]. *Zeitschrift für Pädagogische Psychologie.* **26**, 251–261. <https://doi.org/10.1024/1010-0652/a000076> (2012).
39. Condon, D. M. & Revelle, W. The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intell.* **43**, 52–64. <https://doi.org/10.1016/j.intell.2014.01.004> (2014).
40. Hayes, T. R., Petrov, A. A. & Sederberg, P. B. Do we really become smarter when our fluid-intelligence test scores improve?. *Intelligence* **48**, 1–14. <https://doi.org/10.1016/j.intell.2014.10.005> (2015).
41. Bibbey, A., Carroll, D., Roseboom, T. J., Phillips, A. C. & de Rooij, S. R. Personality and physiological reactions to acute psychological stress. *Int. J. Psychophysiol.* **90**, 28–36. <https://doi.org/10.1016/j.ijpsycho.2012.10.018> (2013).
42. Nurmi, J.-E. & Kiuru, N. Students' evocative impact on teacher instruction and teacher–child relationships: Theoretical background and an overview of previous research. *Int. J. Behav. Dev.* **39**, 445–457. <https://doi.org/10.1177/0165025415592514> (2015).
43. de la Rosa, S. & Breidt, M. Virtual reality: A new track in psychological research. *Br. J. Psychol.* **109**, 427–430. <https://doi.org/10.1111/bjop.12302> (2018).
44. Huang, Y., Richter, E., Kleickmann, T., Wiepke, A. & Richter, D. Classroom complexity affects student teachers' behavior in a VR classroom. *Comput. Educ.* **163**, 104100. <https://doi.org/10.1016/j.compedu.2020.104100> (2021).
45. Huang, Y., Richter, E., Kleickmann, T. & Richter, D. Class size affects preservice teachers' physiological and psychological stress reactions: An experiment in a virtual reality classroom. *Comput. Educ.* **184**, 104503. <https://doi.org/10.1016/j.compedu.2022.104503> (2022).
46. Hasenbein, L. et al. Investigating social comparison behaviour in an immersive virtual reality classroom based on eye-movement data. *Sci. Rep.* **13**, 14672. <https://doi.org/10.1038/s41598-023-41704-2> (2023).
47. Billingsley, G., Smith, S., Smith, S. & Meritt, J. A systematic literature review of using immersive virtual reality technology in teacher education. *J. Interact. Learn. Res.* **30**, 65–90 (2019).
48. Heid, H. Situation als Konstrukt. Zur Kritik objektivistischer Situationsdefinitionen [Situation as construct. A critical review of objectivist definitions of the situation]. *Schweizerische Zeitschrift für Bildungswissenschaften.* **23**, 513–528 (2001).
49. Kärner, T., & Sembill, D. Multimodal stress assessment in working and learning contexts using physiological, observational and experience-based data in *Methods for researching professional learning and development: challenges, applications and empirical illustrations* (ed. Goller, M., Kyndt, E., Paloniemi, S. & Damşa C.) 111–136 (Springer, 2022).
50. Seufert, C. et al. Classroom management competency enhancement for student teachers using a fully immersive virtual classroom. *Comput. Educ.* **179**, 104410. <https://doi.org/10.1016/j.compedu.2021.104410> (2022).
51. Wiepke, A., Heinemann, B., Lucke, U., & Schroeder, U. Jenseits des eigenen Klassenzimmers: Perspektiven & Weiterentwicklungen des VR-Klassenzimmers [Beyond the own classroom: Perspectives & further developments of the VR classroom.]. *DELFI 2021 - Die 19. Fachtagung Bildungstechnologien [Educational technologies symposium]*. 331–336; <http://dl.gi.de/handle/20.500.12116/37031> (2021).
52. Pfitzner-Eden, F. Teacher's sense of efficacy scale (TSES) - german adapted version. *Leibniz-Institute for Psychology, Open Test Archive.* <https://doi.org/10.23668/psycharchives.6585> (2016).
53. Schupp, J. & Gerlitz, J.-Y. Big Five Inventory-SOEP (BFI-S). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)* <https://doi.org/10.6102/zis54> (2008).
54. Dworak, E. M., Revelle, W., Doebler, P. & Condon, D. M. Using the international cognitive ability resource as an open source tool to explore individual differences in cognitive ability. *Pers. Individ. Dif.* **169**, 109906. <https://doi.org/10.1016/j.paid.2020.109906> (2021).
55. Borko, H. Methodological contributions to video-based studies of classroom teaching and learning: A commentary. *ZDM Math. Educ.* **48**, 213–218. <https://doi.org/10.1007/s11858-016-0776-x> (2016).
56. Wolff, C. E., Jarodzka, H., van den Bogert, N. & Boshuizen, H. P. A. Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instr. Sci.* **44**, 243–265. <https://doi.org/10.1007/s11251-016-9367-z> (2016).

57. Matthews, G., Szalma, J., Panganiban, A. R., Neubauer, C. & Warm, J. S. Profiling task stress with the dundee stress state questionnaire in *Psychology of Stress: New Research* (eds. Cavalcanti, L. & Azevedo, S.) (Nova Science Publishers, 2013).
58. Hettiarachchi, I. T., Hanoun, S., Nahavandi, D. & Nahavandi, S. Validation of Polar OH1 optical heart rate sensor for moderate and high intensity physical activities. *PLOS ONE*. **14**, 1–13. <https://doi.org/10.1371/journal.pone.0217288> (2019).
59. Schubert, C. *et al.* Effects of stress on heart rate complexity—A comparison between short-term and chronic stress. *Biol. Psychol.* **80**, 325–332. <https://doi.org/10.1016/j.biopsycho.2008.11.005> (2009).
60. Goodie, J. L., Larkin, K. T. & Schauss, S. Validation of the Polar heart rate monitor for assessing heart rate during physical and mental stress. *J. Psychophysiol.* **14**, 159–164 (2000).
61. Johnston, D. *et al.* Stressors, appraisal of stressors, experienced stress and cardiac response: A real-time, real-life investigation of work stress in nurses. *Ann. Behav. Med.* **50**, 187–197. <https://doi.org/10.1007/s12160-015-9746-8> (2016).
62. Faraway, J. J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. (Chapman & Hall/CRC, 2006).
63. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/HIERARCHICAL Models*. (Cambridge University Press, 2006).
64. Meteyard, L. & Davies, R. A. I. Best practice guidance for linear mixed-effects models in psychological science. *J. Mem. Lang.* **112**, 104092. <https://doi.org/10.1016/j.jml.2020.104092> (2020).
65. Kliegl, R., Wei, P., Dambacher, M., Yan, M. & Zhou, X. Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Front. Psychol.* **1**, 1–12. <https://doi.org/10.3389/fpsyg.2010.00238> (2011).
66. Pinheiro, J. C. & Bates, D. M. *Mixed-Effects Models in S and S-PLUS*. (Springer, 2009).
67. Baayen, R. H., Davidson, D. J. & Bates, D. M. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* **59**, 390–412. <https://doi.org/10.1016/j.jml.2007.12.005> (2008).
68. Judd, C. M., Westfall, J. & Kenny, D. A. Experiments with more than one random factor: Designs, analytic models, and statistical power. *Ann. Rev. Psychol.* **68**, 601–625. <https://doi.org/10.1146/annurev-psych-122414-033702> (2016).
69. Westfall, J., Kenny, D. A. & Judd, C. M. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exp. Psychol. Gen.* **143**, 2020–2045. <https://doi.org/10.1037/xge0000014> (2014).
70. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48. <https://doi.org/10.18637/jss.v067.i01> (2015).
71. R Core Team R. *A Language and Environment for Statistical Computing* (2021).
72. Goldhammer, F. *et al.* The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *J. Educ. Psychol.* **106**, 608–626. <https://doi.org/10.25656/01:17967> (2014).
73. Cohen, J. A power primer. *Psychol. Bull.* **112**, 155–159. <https://doi.org/10.1037/0033-2909.112.1.155> (1992).
74. Brysbaert, M. & Stevens, M. Power analysis and effect size in mixed effects models: A tutorial. *J. Cogn.* **1**, 1–20. <https://doi.org/10.5334/joc.10> (2018).
75. Little, R. J. A. A test of missing completely at random for multivariate data with missing values. *J. Am. Stat. Assoc.* **83**, 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722> (1988).
76. Ibrahim, J. G. & Molenberghs, G. Missing data methods in longitudinal studies: A review. *TEST* **18**, 1–43. <https://doi.org/10.1007/s11749-009-0138-x> (2009).
77. Kell, H. J. Do teachers' personality traits predict their performance? A comprehensive review of the empirical literature from 1990 to 2018. *ETS Res. Rep. Ser.* **1–27**, 2019. <https://doi.org/10.1002/ets2.12241> (2019).
78. Roberts, B. W. & Mroczek, D. Personality trait change in adulthood. *Curr. Dir. Psychol. Sci.* **17**, 31–35. <https://doi.org/10.1111/j.1467-8721.2008.00543.x> (2019).
79. Hanley, A. W., de Vibe, M., Solhaug, I., Gonzalez-Pons, K. & Garland, E. L. Mindfulness training reduces neuroticism over a 6-year longitudinal randomized control trial in Norwegian medical and psychology students. *J. Res. Pers.* **82**, 103859. <https://doi.org/10.1016/j.jrp.2019.103859> (2019).
80. Baier, F. *et al.* What makes a good teacher? The relative importance of mathematics teachers' cognitive ability, personality, knowledge, beliefs, and motivation for instructional quality. *Br. J. Educ. Psychol.* **89**, 767–786. <https://doi.org/10.1111/bjep.12256> (2019).
81. Klassen, R. M. & Tze, V. M. C. Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educ. Res. Rev.* **12**, 59–76. <https://doi.org/10.1016/j.edurev.2014.06.001> (2014).
82. Lazarides, R., Watt, H. M. G. & Richardson, P. W. Teachers' classroom management self-efficacy, perceived classroom management and teaching contexts from beginning until mid-career. *Learn. Instr.* **69**, 101346. <https://doi.org/10.1016/j.learninstruc.2020.101346> (2020).
83. Aloe, A. M., Amo, L. C. & Shanahan, M. E. Classroom management self-efficacy and burnout: A multivariate meta-analysis. *Educ. Psychol. Rev.* **26**, 101–126. <https://doi.org/10.1007/s10648-013-9244-0> (2014).
84. Jerrim, J., Sims, S. & Oliver, M. Teacher self-efficacy and pupil achievement: much ado about nothing? International evidence from TIMSS. *Teach. Teach.* <https://doi.org/10.1080/13540602.2022.2159365> (2023).
85. Campbell, J. & Ehlert, U. Acute psychosocial stress: does the emotional stress response correspond with physiological responses?. *Psychoneuroendocrinology*. **37**, 1111–1134. <https://doi.org/10.1016/j.psyneuen.2011.12.010> (2012).
86. Delaney, J. P. A. & Brodie, D. Effects of short-term psychological stress on the time and frequency domains of heart-rate variability. *Percept. Mot. Skills* **91**, 515–524. <https://doi.org/10.2466/pms.2000.91.2.515> (2000).
87. van Dick, R. V. & Wagner, U. Stress and strain in teaching: A structural equation approach. *Br. J. Educ. Psychol.* **71**, 243–259. <https://doi.org/10.1348/000709901158505> (2001).
88. Schmid, R. F. & Thomas, J. Teachers' ambulatory heart rate variability as an outcome and moderating variable in the job demands-resources model. *Anxiety Stress Coping*. **33**, 387–402. <https://doi.org/10.1080/10615806.2020.1746286> (2020).
89. Huang, Y., Richter, E., Kleickmann, T. & Richter, D. Comparing video and virtual reality as tools for fostering interest and self-efficacy in classroom management: Results of a pre-registered experiment. *Br. J. Educ. Tech.* **54**, 467–488. <https://doi.org/10.1111/bjet.13254> (2023).
90. Hamilton, D., McKechnie, J., Edgerton, E. & Wilson, C. Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *J. Comput. Educ.* **8**, 1–32. <https://doi.org/10.1007/s40692-020-00169-2> (2021).

Acknowledgements

The first author is supported by a Jacobs Foundation Research Fellowship, and an Eliteprogramm for Post-docs Fellowship from the Baden-Württemberg Foundation. We acknowledge support from the Open Access Publication Fund of the University of Tübingen.

Author contributions

Conceptualization: L.B., Y.H., E.R., R.M., T.K., D.R.; Writing – Original Draft: L.B., Y.H.; Writing – Review & Editing: L.B., Y.H., E.R., R.M., T.K., D.R.; Formal Analysis: Y.H.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-49508-0>.

Correspondence and requests for materials should be addressed to L.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023