**Stress Experience and Performance during an oral Exam:**

**The Role of Self-Efficacy, Threat Appraisals, Anxiety, and Cortisol**

Tobias Ringeisen[1] [corresponding author], Stephanie Lichtenfeld[2], Sandra Becker[2], & Nina Minkley[3]

[1] Berlin School of Economics and Law,

Alt Friedrichsfelde 60, 10315 Berlin, Germany;

tobias.ringeisen@hwr-berlin.de

[2] University of Munich, Leopoldstr. 13, 80802 Munich, Germany;

lichtenfeld@psy.lmu.de; Sandra.Becker@psy.lmu.de

[3] Ruhr-Universität Bochum, Behavioural Biology and Biology Education,

Universitätsstr. 150, 44780 Bochum, Germany,

nina.minkley@rub.de

Word Count including tables, references, figure captions, endnotes: 9606

**Stress Experience and Performance during an oral Exam:**

**The Role of Self-Efficacy, Threat Appraisals, Anxiety, and Cortisol**

Abstract

**Background and Objectives:** High self-efficacy may reduce emotional and physiological stress responses in the context of an examination. The present study investigated how these stress responses develop on an exam day, and sequential indirect effects between self-efficacy, threat appraisals, stress responses and performance. **Design and Methods:** The sample comprised 92 students (46 women). Self-efficacy, threat appraisals and state anxiety were assessed on a control day one week before an oral exam. Additionally, anxiety was assessed three times on the exam day. Salivary cortisol samples were collected at all time points. **Results:** Pre-exam anxiety and cortisol decreased until grades were announced. For both responses, greater levels were related to a steeper decline. However, changes in anxiety and cortisol were unrelated. Self-efficacy was negatively related to threat appraisals and anxiety on the control day. Greater threat appraisals were associated with higher pre-exam anxiety and a steeper anxiety decrease on the exam day, which in turn, was related to better performance. **Conclusions:** High levels of self-efficacy may reduce threat appraisals and anxiety in the lead up to an exam, which are related to the intensity and decline of anxiety on the exam day. A steeper decline of anxiety may be beneficial to performance.

*Keywords:* academic self-efficacy, anxiety, salivary cortisol, academic success, exams

**Disclosure of interest:** In accordance with Taylor & Francis policy and our ethical obligations as researchers, the authors report no conflict of interest.

# Stress Experience and Performance during an oral Exam:
# The Role of Self-Efficacy, Threat Appraisals, Anxiety, and Cortisol

## 1 Introduction

Taking an oral exam constitutes a social evaluative stressor for students, which may elicit cognitive, affective, and endocrinological stress responses, specifically threat appraisals, anxiety and cortisol secretion. These stress responses can have a negative effect on academic performance as they may impair working memory capacity as well as the retrieval and processing of learned information (for an overview, see Oberauer, Farrell, Jarrold, & Lewandowsky, 2016; Wolf, 2006). Researchers have made an effort to identify within-person resources, which may reduce the detrimental effects of these stress responses on exam performance. Academic self-efficacy (SE) is considered an important dispositional resource that attenuates threat appraisals, state anxiety, and cortisol secretion, and thus the negative effects these stress responses can have on performance (e.g., Bandura, 1997; Minkley, Westerholt, Kirchner, 2014). Academic SE represents self-directed competence beliefs which characterize an individual's self-perceived conviction of being capable to master performance-related tasks in educational settings, especially in exam situations (Bandura, 1992, 1997; Jerusalem & Satow, 1999; Putwain, Sander, & Larkin, 2013).

A number of laboratory studies jointly assessed appraisals, affective and/or physiological stress correlates in a test situation, for instance when participants performed mental tasks or a social-evaluative speech task (e.g., Fonseca, Blascovich, & Marques, 2014; Pruessner, Gaab, Hellhammer, Lintz, Schommer, & Kirschbaum, 1997; Tomaka, Blascovich, Kibler, & Ernst, 1997). In contrast, previous research on the effects of SE in the context of real-life exams focused either on cortisol or self-reported anxiety (Campbell & Ehlert, 2012; Dickerson & Kemeny, 2004; Schönfeld et al., 2017). Yet, studies which examine the relationships between these variables in the context of an actual exam are still lacking. Specifically, no study has examined the interplay between SE, stress responses, and performance, when anxiety and cortisol are measured repeatedly before and after a real-life exam. However, such an approach is recommended to detect the relations between changes in anxiety and cortisol, and other relevant variables over the course of an exam (Raffety, Smith, & Ptacek, 1997). Thus, the current study sought to investigate the interplay between SE, threat appraisals, anxiety, cortisol concentration, and performance in the context of an oral exam, thereby measuring anxiety and cortisol repeatedly on the exam day.

### *1.1 Conceptualizing the Interplay between Self-Efficacy, Stress, and Performance*

While several theories and theoretical models have been developed to conceptualize the interplay of beliefs, stress responses and performance in the context of examinations each model puts an emphasis on different processes. For instance, the control-value theory of achievement emotions (e.g., Pekrun, 2006; Pekrun, Frenzel, Goetz, & Perry, 2007) highlights the importance of control and value appraisals while the biopsychosocial model of challenge and threat (e.g., Blascovich, 2008; Fonseca et al., 2014) specifies under which conditions appraisals of threat and challenge are related to neuroendocrine stress responses, which in turn can differentially affect performance. However, control-value theory neither specifies how exam-related stress responses may develop over time nor addresses the role of threat appraisals in this context. On the other hand, the biopsychosocial model does not address SE beliefs.

Therefore, the present study was based on the transactional model of test anxiety. This model seems most suitable as it conceptualizes the associations between dispositional antecedents like SE, as well as changes in exam-related stress responses and performance in the context of an exam (Lowe, Lee, Witteborg, Prichard, Luhr, Cullinan et al., 2008; Smith & Lazarus, 1993; Spielberger & Vagg, 1995; Zeidner, 1998, 2007). According to this model, students respond with heightened state anxiety and endocrinological stress responses in the form of cortisol secretion, when they perceive performance-related evaluations as threatening. Concerning the development of exam-related stress responses the model further suggests that threat should be high shortly before the exam as students do not know how and what kind of questions they will be asked. Over the course of the exam, students become increasingly familiar with the situation, the way of questioning, as well as with the contents the examiner is aiming at, which goes hand in hand with a decrease in threat appraisals. Accompanying stress responses such as levels of anxiety or cortisol should likewise decline gradually from before until after the exam. Once the exam is completed, the requirements are fully clear but students may still experience apprehension about their performance until grades are announced. After grades are posted, anxiety and cortisol should further decrease as the stressful event is over.

In the framework of the transactional model of test anxiety, SE is considered a key personality variable, which is assumed to attenuate exam-related stress responses. The model proposes direct effects of SE on stress responses, as well as indirect effects of SE through stress responses on performance. Specifically, high SE should be associated with lower threat appraisals, lower anxiety symptoms, and lower cortisol secretion. In turn, a decrease in stress responses should be associated with better academic performance. The effects of SE on the stress responses are supposed to be inversely proportional; the greater the individuals' SE beliefs, the lower the anxiety levels and the lower the neuroendocrine stress responses (Bandura, 1992).

### *1.2 Changes in Anxiety and Cortisol in the Context of Exams*

Many studies rendered support for the above-mentioned change patterns with regard to self-reported anxiety. Compared to a control day, anxiety levels were found to be high in close proximity to the exam and drop thereafter, with further decreases after grades were announced (e.g., Bossong, 1999; Carver & Scheier, 1994; Ringeisen & Buchwald, 2010). Still, a comparison of results in respect to this pattern is difficult since assessment intervals varied considerably across studies, ranging from a few days to several weeks. To detect possible changes across shorter time intervals, Raffety et al. (1997) used a longitudinal design with daily assessments of state test anxiety starting a week before until a few days after the exam. Results from this study yielded that anxiety was high in anticipation of the exam, dropped after the exam, and decreased further in the following days.

In respect to changes of cortisol concentrations, research results are mixed. When someone faces a stressor like an exam, the secretion of cortisol is activated via the hypothalamus pituitary adrenal axis (HPA-axis), which leads to a measurable cortisol increase in the saliva about 20 minutes later (Foley & Kirschbaum, 2010). While the majority of studies found that compared to a control day cortisol concentrations are greater shortly before the exam and declined after the exam was completed (Dickerson, & Kemeny, 2004; Lacey et al. 2000; Verschoor, & Markus 2011; Spangler, Pekrun, Kramer, & Hofmann, 2002), a few studies did not detect any changes in cortisol concentrations, or solely a decrease after the exam (Glaser, Pearl, Kiecolt-Glaser, & Malarkey, 1994; Vedhara, Hyde, Gilchrist, Tytherleigh, & Plummer, 2000). Still, others found cortisol to further increase after the exam (Preuß, Schoofs, Schlotz, & Wolf, 2010; Schoofs, Hartmann, & Wolf, 2008).

A number of studies investigated associations between cortisol concentrations and subjective-affective stress responses, in particular anxiety, in the context of social-evaluative stressors (for a review see Campbell & Ehlert, 2012). Studies investigating these relationships likewise produced mixed results: Some authors found low to moderate associations between cortisol concentrations and subjective-affective stress responses, usually after aggregating repeated assessments across multiple days (e.g., Lindahl, Theorell, & Lindblad, 2005; Spangler et al., 2002), The majority of studies, however, could not find any relationship (Buchanan, al'Absi, & Lovallo, 1999; Campbell & Ehlert, 2012; Minkley et al., 2014; Weekes, Lewis, Patel, Garrison-Jakel, Berger et al., 2006).

### *1.3 The Interplay between Self-Efficacy, Exam-Related Stress, and Academic Performance*

In the context of real-life exams, no study has examined the interplay of SE, cognitive, affective and endocrinological stress responses, and performance. Instead, existing studies which incorporated the assessment of SE focused either on cortisol or self-reported anxiety (Schönfeld et al., 2017; Zeidner, 1998, 2007). In line with the transactional model of test anxiety, a few studies yielded that high SE was associated with lower anxiety and better performance, with anxiety serving as a mediator between SE and performance (e.g., Bonaccio & Reeve, 2010; Mills, Pajares, & Herron, 2006; Putwain et al., 2013; Schnell, Ringeisen, Raufelder, & Rohrmann, 2015).

In respect to the relationship between SE and physiological stress measures such as cortisol concentrations, laboratory studies yielded evidence for negative linear associations between both variables (Schönfeld et al., 2017). In these studies, participants were usually confronted with variations of the Trier Stress Test consisting of a social-evaluative speech task and a mental arithmetic task. SE measured by self-report was negatively associated with the acute cortisol response, especially after aggregating across multiple assessments (e.g, Nierop, Wirtz, Bratsikas, Zimmermann, & Ehlert, 2008; Pruessner et al., 1997). Outside the context of social-evaluative stress, experimental studies also yielded a linear dampening effect of SE on the cortisol response, for instance when patients dealt with a phobic stressor (Bandura, Taylor, Williams, Mefford, Barchas, & Det, 1985; Wiedenfeld, O'Leary, Bandura, Brown, Levine, & Raska, 1990). Other studies, however, failed to identify associations between SE beliefs and cortisol in the context of social evaluative stressors. These studies comprised experimental studies based on the Trier Stress Test (Schommer, Kudielka, Hellhammer, & Kirschbaum, 1999), studies in the context of real-life exams (Schoofs et al., 2008), and a combination of experimental and real-life stressors (van Eck, Nicolson, Berkhof, & Sulon, 1996).

Furthermore, and in contrast to implications from neuropsychological studies (Wolf, 2006), no direct associations between cortisol and performance were found, even in experimental studies (Dickerson & Kemeny, 2004; Schoofs et al., 2008). Methodological reasons may account for inconsistent patterns with regard to the reactivity of cortisol, and/or its associations with SE: The cortisol assessment intervals varied across studies, ranging from 30 minutes to a few days (Dickerson & Kemeny, 2004). Changes in cortisol are primarily expected in close proximity to the exam, which calls for multiple assessments on the exam day. However, most studies measured cortisol only once before and after an exam, and seldom in combination with state anxiety (Preuß et al., 2010; Schoofs et al. 2008).

### *1.4 Current Study*

To date, the relations between SE, anxiety and cortisol, and performance are still to be examined in the context of a real-life exam. Moreover, it remains unclear whether levels of and/or changes in state anxiety and cortisol concentrations are linked to one another. In response, the current study investigated whether anxiety and cortisol exhibit similar change patterns from before until after an oral exam, on the exam day. Furthermore, we investigated the relations between changes in both stress responses and SE, threat appraisals, and performance, thereby examining sequential indirect effects. The study implemented three parallelized assessments of anxiety and cortisol on the exam day. In addition. SE, threat appraisals, and baseline levels of anxiety and cortisol were assessed on a control day a week before the exam. By means of growth curve modeling we analyzed levels and change trajectories of anxiety and cortisol on the exam day, as well as corresponding associations with self-efficacy and threat appraisals controlling for the level of anxiety and cortisol on the control day. Specifically, we examined the following two hypotheses:

*Hypothesis 1: Cortisol Concentrations and State Anxiety.*

We expected that taking an oral exam constitutes a stressor for students, which would be indicated by greater levels of cortisol concentration and state anxiety on the exam day, as compared to a control day. Moreover, we hypothesized similar change patterns in both stress responses on the exam day: From before until after the exam, cortisol and state anxiety should gradually decline until the grade is announced. Due to inconsistent findings, we did not specify a hypothesis regarding the relation of anxiety and cortisol (intercepts) and their change gradients (slopes) but tested whether these variables were associated.

*Hypothesis 2: Associations between Self-Efficacy, Threat, Anxiety, Cortisol, and Performance.*

We expected SE to be negatively associated with cortisol and anxiety, directly and indirectly via threat appraisals. Threat appraisals should be positively related to cortisol and anxiety. Further, we hypothesized that a better performance during the exam would be associated with lower threat appraisals, lower anxiety levels, and a steeper anxiety decrease on the exam day. In terms of sequential effects, we expected the indirect paths from SE through threat appraisals and anxiety levels (control day) on anxiety (exam day), and the indirect paths from SE through anxiety on performance to be significant. Due to inconsistent findings, we did not specify a hypothesis regarding the relations between performance and cortisol but tested whether these variables were related.

## 2 Method

### *2.1 Participants*

Participants were 92 students (n = 46 women) between the ages of 20 and 36 ($M_{age}$ = 24.53; *SD* = 3.07) from a German university. The majority of students had been raised in Germany (80%, n=74). The remaining students indicated Russian, Chinese, or Vietnamese as cultural backgrounds. All students had been living in Germany for at least two years. Foreign students verified their language skills with a C2-certificate to pursue their studies in German.

### *2.2 Procedure*

The study was conducted as part of a larger study (see also Bermeitinger, Hellweg, Andree, Roick, & Ringeisen, 2018; Roick & Ringeisen, 2017) in a regular psychology module. Participants attended a weekly lecture and accompanying tutorials. To complete the module, all students had to pass an oral exam, which was administered by the same examiner over the course of 14 days. Prior to the study, students were thoroughly informed about the voluntary nature of their participation. All students provided written informed consent. The study was conducted in accordance with the Declaration of Helsinki and was approved by the local ethics commission.

SE, threat appraisals, demographics (gender, age, cultural background), and medical information (weight, height, wake-up time, chronic diseases, medication intake, smoking, average cigarette consumption per day, the use of oral contraceptives) were assessed on a control day a week before the exam (t1). On the exam day, there were three measurement points (cf. Preuß et al., 2010; Schoofs et al. 2008): 30 minutes before the exam (t2), directly after the exam (which lasted about 30 minutes) but before announcement of the grade (t3), and another 30 min later after announcement of the grade to the examinee (t4). State anxiety and cortisol concentration were assessed at all four time points[1]. Except for cortisol concentrations, all study variables were assessed by means of paper-and-pencil questionnaires. The student's grades were determined by the examiner once the exam was completed. Students were orally informed about their grades right thereafter. The grades were then reported anonymously by the student in the questionnaire. All students participated in the entire study, though there were single missing values for some items.

The starting time of each oral exam varied between 9 am and 4 pm for practical reasons. The topics were identical for all exams although the questions varied slightly between examinees to ensure an adaptive exam format. To take the varying starting times of the oral exams into account, we employed an intraindividual control design since the cortisol concentration does not only change with the onset of a stressor (acute cortisol response) but also varies throughout the

[1] Aside from the control day, we initially planned to measure both stress responses at intervals of 30 min on the exam day: 30 min before, right before, directly after the exam (which lasted about 30 minutes) but before announcement of the grade, and another 30 min later after announcement of the grade to the examinee. The ethics commission of the university, however, considered an assessment right before the exam as potentially performance-hindering for students and required us to omit the measurement point right before the exam.

day (diurnal cortisol pattern) (Dickerson & Kemeny, 2004; Foley & Kirschbaum, 2010). This pattern is characterized by a peak cortisol concentration about 30 min after awakening followed by a continuous decrease of the cortisol concentration (Hellhammer, Wüst, & Kudielka, 2009). As part of the intraindividual control design, students provided a saliva sample on the control day at the same time the exam would start a week later, in addition to providing three saliva samples on the exam day. This way, individual baseline cortisol concentrations and individual diurnal cortisol patterns could be controlled in subsequent data analyses.

### *2.3 Measurements*

The measures to assess self-efficacy, threat appraisals, and state anxiety represent well-established instruments that were validated on German students.

#### *2.3.1 Academic Self-Efficacy*

Self-efficacy was assessed using the *Academic Self-efficacy Scale* developed by Jerusalem and Satow (1999). The scale consists of eight items that measure students' subjective beliefs regarding their ability to deal with high demands related to academic performance (e.g., "Even if a lecturer doubts my abilities, I am sure that I can achieve a good performance"*).* Responses were evaluated on a Likert scale ranging from 1 ("not at all") to 4 ("a great deal"*).* Reliability was acceptable in the current study (Cronbach's alpha = .68).

#### *2.3.2 Threat Appraisals*

Threat appraisals were measured with a respective scale consisting of four items (Rakoczy, Buff, & Lipowsky, 2005). The scale captured the degree to which students perceived the upcoming exam to exceed their coping capabilities (e.g., "The exam puts me under pressure"). Responses were evaluated on a Likert scale ranging from 1 ("not at all") to 4 ("a great deal"*).* Reliability was high (Cronbach's alpha = .82).

#### *2.3.3 State Anxiety*

Exam-related state anxiety was measured by means of an adjective-based instrument (Carver & Scheier, 1994; Ringeisen & Buchwald, 2010). Participants rated the extent to which they were “anxious” (German: ängstlich), “fearful” (furchtsam), and “worried” (besorgt), using a five-point Likert scale (1 = “not at all” to 5 = “a great deal”). Cronbach’s alpha was good for all four measurement points, ranging from .79 to .82.

*2.3.4 Cortisol Assessment*

Participants’ saliva was collected via a shortened straw into polypropylene micro tubes (SafeSeal, Sarstedt). Thereafter, the samples were frozen at -20 °C, thawed, vortexed, and centrifuged for 15 min at 2500 × g (Function Line 400R, Heraeus) twice. On the day of the analysis, the supernatant was transferred in duplicate into a pre-coated microwell plate and cortisol concentration was quantified using an immunoassay kit (IBL, Hamburg, Germany). Since blood contamination (caused by gums bleed or injuries in the participants’ mouth) affects subsequent measurements, two blood-contaminated samples were discarded (Westermann, Demir, & Herbst, 2004). Analyses were conducted using a 96-well ELISA reader (Thermo Fisher). Intra-assay coefficients of variance were below 5%, and inter-assay coefficients were below 11%.

*2.3.5 Received Grade*

Consistent with the recommendations by the German Association to Foster Educational Research (Rakoczy et al., 2005; adapted from Carver & Scheier, 1994), participants reported the grade they had achieved with open response format (“Which grade did you achieve on the exam?”). Overall, student outcomes were satisfactory (*M* = 2.67, *SD* = 0.91). The German grading system ranges from 1.0 (very good) to 5.0 (insufficient). For ease of interpretation grades were reverse-coded, so that a high grade indicates high performance.

***2.4 Preparation of the Data***

Medical or demographic variables were found to influence the intensity of SE, cortisol concentration, and anxiety, as well as their interplay in the context of exams (Foley & Kirschbaum, 2010; Tibubos, Rohrmann, Hodapp, & Ringeisen, 2013; Zeidner, 1998, 2007). Thus, these variables were included as control variables in all data analysis. Special emphasis was given to gender and cultural background as their influence may be particularly strong in oral exams. A significant interaction between gender and cultural background ($F(3, 85) = 2.87$, $p < .05$, $\eta_p^2 = .09$) indicated that Non-Germans, compared to Germans, reported greater threat, lower SE, and achieved worse grades, especially Non-German males. Moreover, on average, women achieved better exam results than men. Before the data was analyzed, a number of cases had to be excluded. Because of possible confounding effects on cortisol concentrations, the saliva of students with the presence of a serious medical condition, and those who used any long-term medication (except oral contraceptives [OC]), or over-/underweight (BMI > 25 kg/m2 or < 18.5 kg/$m^2$) was not analyzed (Foley & Kirschbaum, 2010). Consequently, we examined saliva samples from 80 students and questionnaire data from all 92 students.

### *2.5 Statistical Approach*

Data was analyzed by means of repeated-measures ANCOVAs using SPSS Version 22.0, and by means of first-order latent growth curve models (Bollen & Curran, 2006) using Mplus Version 7.0 (Muthén & Muthén, 2013). To examine changes in anxiety and cortisol concentrations (Hypothesis 1), we modeled levels and change trajectories of both stress responses on the exam day, controlling for the baseline levels of anxiety and cortisol on the control day (model 1). Such approach mirrors previous studies, which consistently identified positive associations between repeated assessments of either anxiety (Bossong, 1999; Carver & Scheier, 1994; Ringeisen & Buchwald, 2010) or cortisol (e.g., Dickerson & Kemeny, 2004; Preuß et al., 2010; Lacey et al. 2000; Spangler et al., 2002) over time. One manifest indicator was used to model anxiety (average score across the three adjectives) and cortisol per assessment point. The intercept factors of anxiety and cortisol were modeled by fixing the factor loadings of the manifest indicators to 1. To model the latent slopes, we assumed a linear decrease from before the exam (t2) until after the announcement of the grades (t4). Considering that the interval between t2 and t3 was 60 min, while the interval between t3 and t4 was 30 min, the factor loadings had numerical values of $\lambda_{t2} = 0$, $\lambda_{t3} = 2$, and $\lambda_{t4} = 3$. Therefore, the mean score of the two intercept factors represented the pre-exam levels of anxiety and cortisol on the exam day. Correlations among the respective slope and intercept factors were allowed.

To investigate the associations of SE, stress responses, and academic performance (Hypotheses 2), we added SE, threat appraisals and grade to the model. First, we calculated correlations between the growth curve coefficients of anxiety and cortisol, and the remaining study variables (model 2). Subsequently, we specified a sequential indirect effects model (Little, 2013) based on the assumptions of the transactional model of test anxiety (model 3): The latter model included direct effects from SE on threat appraisals, anxiety and cortisol on the control day, on the intercept and slope factors of anxiety and cortisol on the exam day, and on exam performance. Furthermore, direct effects from threat appraisals on all stress measures and exam performance, and direct effects from the stress measures on exam performance were specified. In order to test the indirect effects of SE on anxiety, cortisol and performance on the exam day mediated by threat appraisals and the stress responses on the control day, we calculated bootstrapped confidence intervals (boot = 1000)[2]. For model 1, we used uncentered predictor variables to obtain parameter estimates for the intercept and the slope factors. For models 2 and 3, all predictor variables were grandmean-centered. Gender and cultural background were included as covariates in all models.

Data were analyzed using Mplus Version 7.0 (Muthén & Muthén, 2013) with the robust maximum likelihood (MLR) estimator, which is recommended for small and medium sample

[2] As bootstrapped confidence intervals are not available with MLR estimation, we used the confidence intervals from an analogous model with ML estimation.

sizes (Wang & Wang, 2012). Missing values were taken into account with the full information maximum likelihood (FIML) algorithm. Model fit was estimated using primary fit indices as recommended by Hu and Bentler (1999): chi-square test of model fit ($\chi^2$), root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardized root mean square residuals (SRMR). For the CFI, a value close to 1 exemplifies an excellent model fit, a value >.95/.90 a good/acceptable model fit. For the SRMR and RMSEA, a value close to 0 denotes a perfect model fit, whereas values ≤ .06/.08 are good/acceptable (Hu & Bentler, 1999).

## 3 Results

### *3.1 Changes in Cortisol Concentration and State Anxiety*

Descriptive statistics, range, skewness, and kurtosis of the study variables are presented in Table 1. Skewness values below two and kurtosis values below seven signify a normal distribution (West, Finch, & Curran, 1995), implying that all variables showed a normal distribution. To determine the change patterns in anxiety and cortisol concentrations, we first computed repeated measurement ANCOVAs with the between subject factors “gender” (men/women) and “cultural background” (Non-German/German students) and the within-subject factor “time of measurement”. Age, SE and threat appraisals were included as covariates. For cortisol, medical information was included as additional between-subject factors (smoker: yes/no; the use of oral contraceptives: yes/no), and additional covariates (average cigarette consumption per day; wake-up time on the control day).

- insert Table 1 about here -

Comparing pre-exam anxiety levels on the control and the exam day, none of the interactions between time, gender, cultural background, and any of the covariates were significant (*F*s < 1.01, *p*s > .31). However, a significant main effect of time was found, $F(2, 83) = 8.59$, $p < .01$, $\eta_p^2 = .09$, signifying that anxiety levels were greater on the exam day, as compared to the control day (see Figure 1). Likewise, for cortisol concentrations, a significant main effect of time, $F(1, 54) = 19.17$, $p < .01$, $\eta_p^2 = .26$, indicated that pre-exam cortisol concentration was greater on the exam day, as compared to the control day (see Figure 1). Moreover, all two- and three-way interactions between gender, cultural background, and time were at least marginally significant (*F*s > 3.33, *p*s < .07). In contrast, none of the remaining interactions involving the covariates (*F*s < 1.70, *p*s > .31) nor any other main effects were significant (*F*s < 2.06, *p*s > .16).

Following up on these significant interaction effects, we examined levels and change trajectories of anxiety and cortisol on the exam day by means of growth curve modeling controlling for the levels of anxiety and cortisol on the control day. Gender and cultural background were included as covariates (model 1). The model reflected an acceptable fit to the data (see Table 3). The growth curve coefficients for anxiety revealed a significant mean of the intercept ($M_{\text{intercept anxiety}} = 2.97$, $p < .01$) with negligible variability ($Var_{\text{intercept anxiety}} = 0.30$, $p = .06$). The significant negative mean of the anxiety slope indicated a linear decrease in state anxiety ($M_{\text{slope anxiety}} = -0.39$, $p < .01$) with significant interindividual differences in the strength of the decrease ($Var_{\text{slope anxiety}} = 0.06$, $p < .01$). The growth curve coefficients for cortisol yielded a significant mean and variance of the intercept ($M_{\text{intercept cortisol}} = 34.12$, $p < .01$; $Var_{\text{intercept cortisol}} = 269.35$, $p < .01$). The significant negative mean ($M_{\text{slope cortisol}} = -5.93$, $p < .01$) and the significant variance of the cortisol slope ($Var_{\text{slope cortisol}} = 25.61$, $p < .01$) pointed toward a linear decrease in the cortisol concentrations with significant interindividual differences in the strength of the linear decrease. Thus, participants varied in their cortisol concentration on the exam day, and in the strength of their linear decrease in anxiety and cortisol from before until after the examination, but they showed similar pre-exam anxiety levels on the exam day.

An inspection of the associations of the slope and intercept factors revealed that greater anxiety levels on the exam day were associated with a steeper anxiety decrease ($r = -.77$, $p < .01$). Likewise, a greater cortisol concentration on the exam day was associated with a steeper concentration decrease ($r$ = -.69, $p$ < .01). The growth curve coefficients of anxiety and cortisol were unrelated. In terms of autoregressive effects, greater anxiety levels on the control day were associated with greater pre-exam anxiety levels on the exam day ($\beta$ = .70, *SE* = .11, $p$ < .01). Likewise, a greater cortisol concentration on the control day was related to greater pre-exam concentrations on the exam day ($\beta$ = .53, *SE* = .11, $p$ < .01). With regard to the demographic variables, a few significant effects were observed. Compared to Germans, non-Germans reported higher anxiety levels on the control day ($\beta$ = .27, *SE* = .10, $p$ < .01). Compared to women, men reported lower anxiety levels on the control day ($\beta$ = .21, *SE* = .09, $p$ < .05), and a steeper decrease in anxiety on the exam day ($\beta$ = -.40, *SE* = .12, $p$ < .01).

In sum, the hypothesized patterns could be confirmed for anxiety and the cortisol concentrations. Compared to a control day, the level of either stress response was higher on the exam day and showed a linear decrease from before until after the exam. Moreover, levels and decline of anxiety were unrelated to concentrations and decrease of cortisol.

### *3.2 Associations between Self-Efficacy, Threat, Anxiety, Cortisol, and Performance*

Concerning hypothesis 2, we extended the growth curve model and added threat appraisals, self-efficacy, and the grade. First, we calculated correlations between the growth curve coefficients of anxiety and cortisol and the remaining study variables (model 2). Subsequently, we specified a sequential indirect effects model (model 3). Gender and cultural background remained included as covariates (model 2). The extended growth model with correlations between all study variables reflected only an acceptable fit (see Table 3). Screening the intercorrelations, the predicted patterns could largely be confirmed (see Table 2). The intercept and slope factor of either stress measure were negatively associated while no relations between anxiety and cortisol emerged. SE correlated negatively with threat appraisals, anxiety on the control day, and both pre-exam anxiety and cortisol on the exam day. Threat appraisals were positively associated with anxiety levels on the control and the exam day, but not with the cortisol concentrations. A better grade in the exam was related to lower anxiety levels on the control day, a steeper decline of anxiety on the exam day, and higher SE.

The sequential indirect effects model (model 3) showed a good fit to the data (see Table 3), which turned out to be significantly better than the fit of model 2, as indicated by a $\chi^2$-difference test ($\Delta\chi^2 = 24.90$, $\Delta df = 7$, $p < .001$) (Yuan & Bentler, 2004). Thus, consistent with hypothesis 2, the data rendered support for the sequential indirect effects model. However, the hypothesized association patterns could only be confirmed in part (see Figure 2).

- insert Table 2, Table 3, and Figure 2 about here -

The relations between the growth curve coefficients and the assessments of anxiety and cortisol on the control day, which had already been depicted in model 1, remained essentially the same. With regard to the additional associations, the following patterns could be found: SE showed negative relations with threat appraisals ($\beta = -.33$, $SE = .09$, $p < .01$), and with anxiety levels on the control day ($\beta = -.30$, $SE = .08$, $p < .01$). Surprisingly, SE was related neither to the cortisol concentrations on the control day nor to the growth curve coefficients of anxiety and cortisol on the exam day. Greater threat appraisals were associated with greater anxiety on the control day ($\beta = .48$, $SE = .07$, $p < .01$) and the exam day ($\beta = .33$, $SE = .07$, $p < .01$) and with a flatter anxiety decline on the exam day ($\beta = -.36$, $SE = .07$, $p < .01$). Finally, a better grade in the exam was associated with a steeper anxiety decrease from before until after the exam ($\beta = -.52$, $SE = .07$, $p < .05$). The grade was unrelated to SE, threat appraisals, anxiety levels, and the cortisol concentrations. Tests of the indirect effects yielded substantial effects of SE through anxiety levels on the control day on anxiety levels on the exam day ($b = -.15$ [-.05, -.24], $SE = .05$, $p < .01$), and through threat appraisals and anxiety levels on the control day on anxiety levels on the exam day ($b = -.08$ [-.01, -.15], $SE = .04$, $p < .05$).

With regard to the demographic variables, non-Germans, compared to Germans, reported lower SE levels ($\beta = -.24$, $SE = .11$, $p < .05$) and higher anxiety levels ($\beta = .27$, $SE = .10$, $p < .01$)

on the control day. Compared to women, men reported lower anxiety levels ($\beta$ = .21, *SE* = .09, $p < .05$) and lower threat appraisals ($\beta$ = .33, *SE* = .09, $p < .01$) on the control day, and a steeper decrease in anxiety on the exam day ($\beta$ = -.40, *SE* = .12, $p < .01$).

## 4 Discussion

The current study investigated the relations between changes in state anxiety and cortisol, SE, threat appraisals, and performance in the context of an oral exam. Our results add to the understanding of exam-related stress responses since our study is the first to investigate changes in anxiety and cortisol simultaneously by means of latent growth modeling in the context of a real-life exam (Dickerson & Kemeny, 2004; Zeidner, 1998). Over the course of an oral exam, the levels of both stress responses were greater on the exam day, as compared to a control day, and showed a linear decline from before the exam until after announcement of the grades. These patterns are in line with the transactional model of test anxiety which assumes that students expand their knowledge on the requirements and their performance over the course of an exam, which should be indicated by gradual decreases in anxiety and endocrinological stress responses (Carver & Scheier, 1994; Lowe et al., 2008; Spielberger & Vagg, 1995; Zeidner, 1998).

Moreover, our findings extend prior study results since we investigated the relations between changes in anxiety and cortisol, and SE, as well as threat appraisals and performance (Bandura, 1992, 1997; Lowe et al., 2008; Zeidner, 1998, 2007). SE was negatively related to threat appraisals and anxiety on the control day. Higher levels of threat were associated with greater pre-exam anxiety and a steeper anxiety decrease over the course of the exam, which in turn, was related to better performance in the exam. In terms of indirect effects, we found the pathways from SE through threat appraisals and anxiety levels on the control day on anxiety levels on the exam day to be significant. Conceptually, these patterns support the assumption that state test anxiety may develop if students with poor self-directed competence beliefs feel threatened in social-evaluative performance settings (cf. Putwain & Symes, 2014): Students with high SE beliefs should experience lower threat appraisals and lower pre-exam anxiety levels, and a smaller anxiety decrease during the exam. Thus, a better exam performance may not be associated with lower anxiety levels per se but vary with the degree to which anxiety decreases over the course of the exam (cf. Putwain et al., 2013). In case students experience a sharp decline in anxiety from before until after an exam, working memory capacity may be released which is beneficial for exam-related information processing, and may thus go along with better exam performance (Oberauer et al., 2016).

The non-significant associations between anxiety and cortisol, and between SE, cortisol and performance, may be reflected in light of theoretical and methodological considerations. In line with Campbell and Ehlert (2012), who stated that only 25% of existing studies identify significant correlations between cortisol concentrations and emotional stress variables, we found no evidence that the changes in both stress responses were related. An explanation for the independence of anxiety and cortisol may be derived from studies on appraisal emotion theories and hormonal regulation mechanisms, which suggest that either stress response may be associated with different cognitive processes (Bandura, 1997; de Kloet, Joels, & Holsboer, 2005; Fonseca et al., 2014; Pekrun et al., 2007; Putwain et al., 2013; Wolf, 2006). In terms of information processing, a high pre-exam cortisol concentration signifies high levels of alertness and arousal, which helps the students to deal with unexpected questions throughout the exam. After the exam, requirements and the contents are fully clear, which is usually accompanied by a drop in attention, a rise in immediate control appraisals, and an endocrinological feedback loop that leads to a decrease in the cortisol concentration. Cortisol should thus be linked to immediate control expectancies, which may benefit from high SE beliefs but largely depend on situational variables (Schönfeld et al., 2017). Lower levels of anxiety, on the other hand, can be conceptualized as a function of positive achievement outcome expectancies, which are primarily triggered by high SE (Pekrun et al., 2007; Putwain et al., 2013).

Alternatively, some researchers interpret non-significant relationship between performance and anxiety levels/the cortisol concentrations in light of the Yerkes-Dodson Law (Diamond, 2005; Hanoch & Vitouch, 2004). With regard to absolute levels of either stress measure an inverted U-shaped relation with performance may be expected: High levels of exam-related anxiety may hinder performance, while low to moderate levels are considered functionally activating as they go along with awareness for adequate preparation and deeper information processing which is conducive to performance (Fonseca et al., 2014). Pre-exam levels of anxiety or cortisol which are slightly above average may thus be unrelated to performance, as in the current study. An inverse relation may be observed if the cortisol concentration is low to moderate (Schilling, Kölsch, Larra, Zech, Blumenthal et al., 2013).

Methodological factors may have also affected the associations between the study variables. Several researchers found the relationship between affective and hormonal indicators of stress to vary with the design of the study, the intervals of assessment, or the specific measure that was used (Campbell & Ehlert, 2012; Dickerson & Kemeny, 2004). For instance, some studies measured anticipatory cortisol concentrations levels on the exam day, while others assessed pre-exam levels a day before, yielding differences in association patterns with subjective stress measures (Preuß et al., 2010; Schoofs et al., 2008). Similarly, other studies found that the strength of associations between SE, anxiety, and physiological stress indicators depends on situational characteristics such as the type of the (academic) challenge (e.g., delivering a talk or taking an oral exam), the amount of perceived control, and the status of a present audience (experts vs. novices; e.g., Dickerson & Kemeny, 2004; Gerin, Litt, Deich, & Pickering, 1995, Sanz & Villamarín, 2001). As such, the associations between SE and cortisol may be stronger when the cortisol assessments are aggregated across different measurement points, balancing variations in situational influences (Pruessner et al., 1997; Dickerson &

Kemeny, 2004). One should bear these methodological aspects in mind when interpreting the current association patterns between cortisol and anxiety/SE.

Finally, in line with previous results, a few group differences emerged. Non-Germans, compared to Germans, reported lower SE levels and higher anxiety levels on the control day. Compared to women, men reported lower anxiety levels and lower threat appraisals on the control day, and a steeper decrease in anxiety on the exam day. The stronger stress responses of some Non-Germans may be due to their language skills. Despite their proficiency in German, subtle language problems of non-native speakers may lead to greater stress responses, especially in anticipation of oral exams (Mills et al., 2006; Tibubos et al., 2013).

## 5 Strengths and Limitations

The present study has a number of strengths. To our knowledge, it is the first study to examine the relations between changes in state anxiety and cortisol, SE, threat appraisals, and performance in the context of a real exam. This way, a social-evaluative performance setting with high ecological validity could be realized, considering that oral exams constitute the most threatening exam type in the eyes of most students (Dickerson & Kemeny, 2004; Zeidner, 1998). To enable a fine-grained analysis of change patterns, we assessed anxiety and cortisol on three occasions on the exam day. In addition, we employed an intraindividual control design, and assessed baseline levels of anxiety and cortisol on a control day (Dickerson & Kemeny, 2004; Foley & Kirschbaum, 2010). This way, the baseline levels of both stress parameters could be incorporated into subsequent data analyses. Moreover, we applied growth curve modeling to investigate the changes in anxiety and cortisol on the exam day to control for measurement error.

Besides these strengths, some methodological limitations need to be considered when interpreting our results. First, the sample size was rather small. Thus, specifically the non-significant relations between anxiety and cortisol concentrations should be interpreted with caution. However, studies which assess both affective and endocrinological stress responses by means of multiple measurements are complex and often limited to samples sizes up to 100 participants due to practical reasons (Campbell & Ehlert, 2012; Zeidner, 1998, 2007). Second, we did not control for the time the participants woke up on the exam day. Differences in cortisol concentrations could thus be due to differences in wake-up time on the exam day, as compared to the control day, rather than reflecting differences in threat appraisals, anxiety or SE. For instance, previous studies found that participants slept significantly worse in anticipation of an upcoming exam, or experienced earlier/later wake-up times which may have altered the timing of the diurnal cortisol pattern and associated differences in cortisol and anxiety in the current study (van Lenten & Doane, 2016; Baglioni, Spiegelhalder, Lombardo, & Riemann, 2010). Third, we assessed threat appraisals only once before the exam which does not allow us to determine the causality of mediation between SE, threat, and stress responses. An analysis of longitudinal mediation requires at least a half-longitudinal design (Little, 2013) which could not be realized due to organizational constraints. We therefore recommend assessing SE, threat appraisals and stress responses at least twice over time to disentangle the true mediation effects in future studies. Moreover, we recommend studies with larger samples and greater numbers of observations for anxiety and cortisol to enable a multi-level analysis of either stress response nested within days (control/exam day) which are nested within individuals. Fourth, it remains unknown to what extent the results of our study were specific to the context of the psychology module, and/or specific to exams conducted by the examiner, who conducted all oral exams. Thus, when investigating stress responses in the context of real-life exams, we are confronted with the common limitations of quasi-experimental studies. That is, results may be explained by a third variable or process and/or may thus not be transferable to other exam formats, or to other subject domains. For future studies, we recommend a randomized assignment of different examiners. Fifth, the reliability of the self-efficacy measure was just acceptable. Since the Academic Self-efficacy Scale (Jerusalem & Satow, 1999) is composed of task-specific items like dealing with learning obstacles, acquiring knowledge, or solving assignments, it is likely that students differed in their self-perceived competence to manage the different academic challenges. Task-specific judgments thus enable more precise ratings and better construct validity although they carry the risk of impaired reliability if the required skills differ intraindividually and/or in relation to different modes of exam (Pajares, 1996).

## 6 Conclusions and future Directions

The current study is helpful in understanding the interplay of changes in anxiety and cortisol, SE, threat appraisals, and performance over the course of an oral exam. Anxiety and cortisol concentrations changed with similar patterns from before until after the exam although levels and change trajectories in cortisol concentrations were unrelated to the other study variables. SE was negatively related to threat appraisals and anxiety while higher threat levels were associated with greater intensity and a steeper decrease of anxiety over time, which in turn, was related to better exam performance. Controlling for measurement error by means of latent growth curve modeling, the findings provide support for the assumptions of the transactional model of test anxiety with regard to changes in anxiety and cortisol on the exam day, and with regard to sequential direct and indirect effects of SE on threat appraisals on affective responses. In terms of future studies, we encourage researchers to further examine whether conceptual and/or methodological factors account for the non-significant associations between cortisol and anxiety, or between cortisol and SE. Moreover, we recommend the examine whether the interplay between SE, affect, and performance varies for other negative (e.g., boredom) and/or positive emotions (e.g., enjoyment) and/or the format of the exam (written vs. oral), thereby linking theoretical considerations with empirical findings (Pekrun et al., 2007).

References

Baglioni, C., Spiegelhalder, K., Lombardo, C., & Riemann, D. (2010). Sleep and emotions: A focus on insomnia. *Sleep medicine reviews, 14,* 227-38. 10.1016/j.smrv.2009.10.007.

Bandura, A., (1992). Self-efficacy mechanism in psychobiological functioning. In Schwarzer, R. (Ed.), *Self-Efficacy: Thought Control of Action* (pp. 355–394). Washington D.C.: Hemisphere.

Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York, NY: Freeman.

Bandura, A., Taylor, C.B., Williams, S.L., Mefford, I.N., & Barchas, J.D. (1985). Catecholamine secretion as a function of perceived coping self-efficacy. *Journal of Consulting and Clinical Psychology 53,* 406–414.

Bermeitinger, C., Hellweg, C., Andree, C., Roick, J., & Ringeisen, T. (2018). Goal (dis)engagement, emotions and cognitions in an exam situation: A longitudinal study. *Applied Cognitive Psychology, 32,* 55–65. doi: https://doi.org/10.1002/ acp.3379

Blascovich, J. (2008). Challenge and threat. In A. J. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 431-445). New York: Psychology Press.

Bollen, K. A., & Curran, P. J. (2006). *Latent Curve Models: A Structural equation Perspective.* New York: Wiley.

Bonaccio, S. & Reeve, C. L. (2010). The nature and relative importance of students' perceptions of the sources of test anxiety. *Learning and Individual Differences, 20,* 617–625. doi: 10.1016/j.lindif.2010.09.007.

Bossong, B. (1999). *Stress und Handlungskontrolle: die volitionale Kompetenz bei der Bewertung und Bewältigung aversiver Erfahrungen* [Stress and action control: the role of volitional competence in coping with aversive experiences]. Göttingen: Hogrefe.

Buchanan, T. W., al'Absi, M., & Lovallo, W. R. (1999). Cortisol fluctuates with increases and decreases in negative affect. *Psychoneuroendocrinology, 24*, 227-241. doi: 10.1016/S0306-4530(98)00078-X

Campbell, J., & Ehlert, U. (2012). Acute psychosocial stress: Does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology*, *37*, 1111-1134. doi: 10.1016/j.psyneuen.2011.12.010.

Carver, C. S. & Scheier, M. F. (1994). Situational coping and coping dispositions in a stressful transaction. *Journal of Personality and Social Psychology*, *66*, 184–195. doi: 10.1037/0022-3514.66.1.184.

de Kloet, E. R., Joels, M., & Holsboer, F. (2005). Stress and the brain: From adaptation to disease. *Nature Reviews Neuroscience, 6*, 463-475. doi: 10.1038/nrn1683.

Diamond, D.M. (2005). Cognitive, endocrine and mechanistic perspectives on non-linear relationships between arousal and brain function. *Nonlinearity in Biology, Toxicology, and Medicine, 3*, 1–7.

Dickerson, S.S., & Kemeny, M.E. (2004). Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research. *Psychological Bulletin, 130*, 355-391. doi: 10.1037/0033-2909.130.3.355.

Foley, P., & Kirschbaum, C. (2010). Human hypothalamus-pituitary-adrenal axis responses to acute psychosocial stress in laboratory settings. *Neuroscience and Biobehavioral Reviews, 35*, 91–96. doi:10.1016/j.neubiorev.2010.01.010

Fonseca, R., Blascovich, J., & Garcia-Marques, T. (2014). Challenge and threat motivation: Effects on superficial and elaborative information processing. *Frontiers in Psychology, 5*, 1170-1180. doi: org/10.3389/fpsyg.2014.01170

Glaser, R., Pearl, D. K., Kiecolt-Glaser, J. K., & Malarkey, W. B. (1994). Plasma cortisol levels and reactivation of latent Epstein–Barr virus in response to examination stress. *Psychoneuroendocrinology, 19*, 765–772. doi: 10.1016/0306-4530( 94)90023-X.

Hanoch, Y., & Vitouch, O. (2004). When less is more. Information, emotional arousal and the ecological reframing of the Yerkes-Dodson law. *Theory & Psychology*, *14*, 427–452. doi: 10.1177/0959354304044918.

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology, 34*, 163-171. doi: 10.1016/j.psyneuen.2008.10.026.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structural analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55. doi: 10.1080/10705519909540118.

Jerusalem, M., & Satow, L. (1999). Schulbezogene Selbstwirksamkeitserwartung [academic self-efficacy]. In R. Schwarzer, & M. Jerusalem (Eds.), *Skalen zur Erfassung von Lehrer- und Schülermerkmalen* (pp. 15–16). Berlin, Germany: Freie Universität.

Lacey, K., Zaharia, M. D., Griffiths, J., Ravindran, A. V., Merali, Z., & Anisman, H. (2000). A prospective study of neuroendocrine and immune alterations associated with the stress of an oral academic examination among graduate students. *Psychoneuroendocrinology, 25*, 339–356. doi: 10.1016/S0306-4530(99)00059-1

Lindahl, M., Theorell, T., & Lindblad, F. (2005). Test performance and self-esteem in relation to experienced stress in Swedish sixth and ninth graders—saliva cortisol levels and psychological reactions to demands. *Acta Paediatrica, 94*, 489–495. doi: 10.1111/j.1651-2227.2005.tb01922.x

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York: Guilford Press.

Lowe, P. A., Lee, S. W., Witteborg, K. M., Prichard, K. W., Luhr, M. E., Cullinan, C. M., & Janik, M. (2008). The test anxiety inventory for children and adolescents (TAICA). *Journal of Psychoeducational Assessment*, *26*, 215–230. doi: 10.1177/0734282907303760.

Margolis, H., & McCabe, P. (2006). Improving self-efficacy and motivation: What to do, what to say. *Intervention in School and Clinic*, *41*, 218–227. doi: 10.1177/10534512060410040401.

Mills, N., Pajares, F., & Herron, C. (2006). A reevaluation of the role of anxiety: self-efficacy, anxiety, and their relation to reading and listening proficiency. *Foreign Language Annals, 39*, 276–295. doi: 10.1111/j.1944-9720.2006.tb02266.x.

Minkley, N., Westerholt, D. M., & Kirchner, W. H. (2014). Academic self-concept of ability and cortisol reactivity. *Anxiety, Stress, & Coping*, *27*, 303-316. doi: 10.1080/10615806.2013.848273.

Muthén, L. K., & Muthén, B. O. (2013). *Mplus user's guide*. Seventh edition. Los Angeles, CA: Muthén & Muthén.

Nierop, A., Wirtz, P.H., Bratsikas, A., Zimmermann, R., Ehlert, U., (2008). Stress-buffering effects of psychosocial resources on physiological and psychological stress response in pregnant women. *Biological Psychology, 78*, 261–268.

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What Limits Working Memory Capacity? *Psychological Bulletin, 142*, 758-799. doi: 10.1037/bul0000046

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 543–578. doi: 10.3102/00346543066004543.

Pekrun, R., Frenzel, A., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions. In P. A. Schutz & R. Pekrun (Eds.), *Emotions in education* (pp. 13–36). San Diego, CA: Academic Press.

Preuß, D., Schoofs, D., Schlotz, W., & Wolf, O. T. (2010). The stressed student: Influence of written examinations and oral presentations on salivary cortisol concentrations in university students. *Stress, 13,* 221-229. doi: 10.3109/10253890903277579.

Pruessner, J. C., Gaab, J., Hellhammer, D. H., Lintz, D., Schommer, N., & Kirschbaum, C. (1997). Increasing correlations between personality traits and cortisol stress responses obtained by data aggregation. *Psychoneuroendocrinology, 22*, 615–625. doi: 10.1016/S0306-4530(97)00072-3

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology, 28,* (7), 916–931.

Putwain, D. W., Sander, P. & Larkin, D. (2013). Academic self-efficacy in study-related skills and behaviours: relations with learning-related emotions and academic success. *The British Journal of Educational Psychology*, *83*, 633–650. doi: 10.1111/j.2044-8279.2012.02084.x.

Putwain, D. & Symes, W. (2014). The perceived value of maths and academic self-efficacy in the appraisal of fear appeals used prior to a high-stakes test as threatening or challenging. *Social Psychology of Education, 17,* 229–248. doi: 10.1007/s11218-014-9249-7.

Raffety, B. D., Smith, R. E., & Ptacek, J. T. (1997). Facilitating and debilitating trait anxiety, situational anxiety, and coping with an anticipated stressor: A process analysis. *Journal of Personality and Social Psychology, 72,* 892-906. doi: 10.1037/0022-3514.72.4.892.

Rakoczy, K., Buff, A., & Lipowsky, F. (2005). Befragungsinstrumente [survey instruments]. In E. Klieme, C. Pauli, & K. Reusser (Eds.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur Videostudie "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis".* Frankfurt am Main, Germany: GFPF.

Ringeisen, T. & Buchwald, P. (2010). Test anxiety and positive and negative emotional states during an examination. *Cognition, Brain, Behavior*, *14* (4), 431–447.

Roick, J. & Ringeisen, T. (2017). Self-efficacy, test anxiety, and academic success: a longitudinal validation. *International Journal of Educational Research, 83,* 84-93. doi: 10.1016/j.ijer.2016.12.006

Sanz, A. & Villamarín, F. (2001). The role of perceived control in physiological reactivity: self-efficacy and incentive value as regulators of cardiovascular adjustment. *Biological Psychology, 56,* 219-246. doi: 10.1016/S0301-0511(01)00095-3.

Schilling, T. M., Kölsch, M., Larra, M. F., Zech, C. M., Blumenthal, T. D., Frings, C., & Schächinger, H. (2013). For whom the bell (curve) tolls: Cortisol rapidly affects memory retrieval by an inverted U-shaped dose–response relationship. *Psychoneuroendocrinology, 38*, 1565–1572. doi: 10.1016/j.psyneuen.2013.01.001.

Schnell, K., Ringeisen, T., Raufelder, D., & Rohrmann, S. (2015). The impact of adolescents' self-efficacy and self-regulated goal attainment processes on school performance – do gender and test anxiety matter? *Learning and Individual Differences, 38*, 90-98. doi: 10.1016/j.lindif.2014.12.008.

Schommer, N. C., Kudielka, B. M., Hellhammer, D. H., & Kirschbaum, C. (1999). No evidence for a close relationship between personality traits and circadian cortisol rhythm or a single cortisol stress response. *Psychological Reports, 84*(3), 840–842.

Schönfeld, P., Preusser, F., & Margraf, J. (2017). Costs and benefits of self-efficacy: Differences of the stress response and clinical implications. *Neuroscience and Biobehavioral Reviews, 75,* 40-52. doi:10.1016/j.neubiorev.2017.01.031

Schoofs, D., Hartmann, R., & Wolf, O. T. (2008). Neuroendocrine stress responses to an oral academic examination: No strong influence of sex, repeated participation and personality traits. *Stress, 11,* 52–61. doi: 10.1080/10253890701453943.

Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition and Emotion, 7,* 233-269. doi: 10.1080/02699939308409189.

Spangler, G., Pekrun, R., Kramer, K., & Hofmann, H. (2002). Students' Emotions, Physiological Reactions, and Coping in Academic Exams. *Anxiety, Stress, and Coping, 15,* 413-432. doi: 10.1080/1061580021000056555.

Spielberger, C. D., & Vagg, P. R. (1995). Test anxiety: A transactional process model. In: C. D. Spielberger & P. R. Vagg (Eds.), *Test anxiety. Theory, assessment, and treatment* (pp. 3–14). Washington, DC: Taylor & Francis.

Tibubos, A., Rohrmann, S., Hodapp, V. & Ringeisen, T. (2013). Prüfungsängstlichkeit im Kulturvergleich [test anxiety in cross-cultural research]. In P. Genkova, T. Ringeisen & F. T. L. Leong (Eds.), *Handbuch Stress und Kultur.* (S. 239-254). Wiesbaden: Springer.

Tomaka, J., Blascovich, J., Kibler, J., & Ernst, J. M. (1997). Cognitive and physiological antecedents of threat and challenge appraisal. *Journal of Personality and Social Psychology, 73*(1), 63-72.

van Eck, M. M., Nicolson, N. A., Berkhof, H., & Sulon, J. (1996). Individual differences in cortisol responses to a laboratory speech task and their relationship to responses to stressful daily events. *Biological Psychology, 43,* 69–84. doi: 10.1016/0301-0511(95)05159-7.

van Lenten, S. & Doane, L. (2016). Examining multiple sleep behaviors and diurnal salivary cortisol and alpha-amylase: Within- and between-person associations. *Psychoneuroendocrinology, 68,* 100-110. doi: 10.1016/j.psyneuen.2016.02.017.

Vedhara, K., Hyde, J., Gilchrist, I. D., Tytherleigh, M., & Plummer, S. (2000). Acute stress, memory, attention and cortisol. *Psychoneuroendocrinology, 25,* 535–549. doi: 10.1016/S0306-4530(00)00008-1.

Verschoor, E., & Markus, C. R. (2011). Affective and neuroendocrine stress reactivity to an academic examination: Influence of the 5-HTTLPR genotype and trait neuroticism. *Biological Psychology, 87*, 439–449. doi: 10.1016/j.biopsycho.2011.06.001

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using mplus.* Chichester: Wiley.

Weekes, N., Lewis, R., Patel, F., Garrison-Jakel, J., Berger, D. E., & Lupien, S. J. (2006). Examination stress as an ecological inducer of cortisol and psychological responses to stress in undergraduate students. *Stress, 9*, 199–206. doi:10.1080/10253890601029751.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R.H. Hoyle (Eds.). *Structural equation modeling: Concepts, issues and applications* (pp. 56–75). Newbury Park, CA: Sage.

Westermann, J., Demir, A., & Herbst, V. (2004). Determination of cortisol in saliva and serum by a luminescence-enhanced enzyme immunoassay. *Clinical Laboratory, 50*(1-2), 11-24.

Wiedenfeld, S.A., O'Leary, A., Bandura, A., Brown, S., Levine, S., & Raska, K. (1990). Impact of perceived self-efficacy in coping with stressors on components of the immune system. *Journal of Personality and Social Psychology, 59*, 1082–1094.

Wolf, O. T. (2006). Effects of stress hormones on the structure and function of the human brain. *Expert Review Endocrinology Metabolism, 1*, 623–632. doi: org/10.1586/17446651.1.5.623

Yuan, K.-H., & Bentler, P. M. (2004). On chi-square difference and z-tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement, 64*, 737–757. doi: 10.1177/0013164404264853

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York, NY: Plenum Press.

Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in Education* (pp. 165–184). San Diego, CA: Academic Press.

Table 1

*Descriptive statistics of the study variables*

| Variable | Cronbach's Alpha | Range | M (SD) | Skewness (SE) | K |
|---|---|---|---|---|---|
| Self-efficacy t1 | .68 | 2-3.8 | 2.83 (.38) | -.09 (.27) | .3 |
| Threat appraisals t1 | .82 | 1-4 | 2.47 (.68) | -.04 (.25) | -. |
| Anxiety t1 | .80 | 1-5 | 2.66 (.98) | .31 (.27) | -. |
| Anxiety t2 | .79 | 1-5 | 3.05 (.88) | -.03 (.27) | -. |
| Anxiety t3 | .80 | 1-4.67 | 2.34 (.92) | .39 (.27) | -. |
| Anxiety t4 | .82 | 1-4 | 1.64 (.78) | 1.25 (.27) | 1. |
| Cortisol t1 | - | .45-18.40 | 5.91 (3.78) | 1.07 (.28) | 1. |
| Cortisol t2 | - | 1.97-36.10 | 11.65 (8.09) | 1.39 (.28) | 1. |
| Cortisol t3 | - | 1.52-36.10 | 9.28 (6.28) | 1.64 (.29) | 4. |
| Cortisol t4 | - | 1.06-25.60 | 6.52 (4.16) | 1.92 (.28) | 5. |
| Grade t4 | - | 1-5 | 2.67 (.91) | .22 (.27) | -. |

**Note**. cortisol concentration is provided in nmol/L; t1 = a week before the exam (control day); t2 = 30 min before exam (exam day); t3 = right after exam (exam day); t4 =after announcement of grade (exam day).

Table 2

*Bivariate correlations between the intercept and the slope factors of anxiety and cortisol and the other study variables*

| | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. |
|---|---|---|---|---|---|---|---|---|
| -efficacy | -.39** | -.53** | -.52** | -.01 | -.09 | -.24* | .01 | .28* |
| eat Appraisals t1 | | .63** | .68** | -.27 | .06 | .09 | .06 | -.0 |
| iety t1 | | | .78** | -.05 | -.07 | .10 | .01 | -.31 |
| rcept anxiety (t2 to t4) | | . | | -.41* | -.06 | .11 | .03 | -.1 |
| e anxiety (t2 to t4) | | | | | -.01 | .06 | .01 | -.33 |
| isol t1 | | | | | | .48** | -.42** | .02 |
| rcept Cortisol (t2 to t4) | | | | | | | -.77** | -.0 |
| e Cortisol (t2 to t4) | | | | | | | | -.0 |
| de t4 | | | | | | | | |

**Note.** *$p < .05$, **$p < .01$.

Table 3

*Model fit indices for the different growth models*

| Model | *Df* | $x^2$ | p | CFI | SRMR | RMSEA | 90% CI |
|---|---|---|---|---|---|---|---|
| Model 1 | 13 | 25.637 | .019 | .960 | .048 | .080 | (.041-.121) |
| Model 2 | 28 | 47.443 | .012 | .955 | .074 | .080 | (.041-.122) |
| Model 3 | 21 | 21.552 | .426 | .999 | .038 | .017 | (.000-.091) |

**Note.** Model 1 = baseline growth model; Model 2 = growth model allowing correlations between all study variables; Model 3 = growth model with indirect sequential effects from self-efficacy, via threat appraisals, anxiety, and cortisol on grade.