When Students Fail:

Neurocognitive Mechanisms Underlying Test Anxiety

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm Biologie

der Georg-August University School of Science (GAUSS)

vorgelegt von

Frieder Leon Schillinger

aus Filderstadt

Göttingen, 2018

Betreuungsausschuss

Prof. Dr. Roland Grabner, Arbeitsbereich für Begabungsforschung, Institut für Psychologie, Karl-Franzens-Universität Graz

Prof. Dr. Bert De Smedt, Parenting and Special Education Research Unit, Faculty of Psychology and Educational Sciences, Katholieke Universiteit Leuven

Mitglieder der Prüfungskommission

Referent: Prof. Dr. Roland Grabner, Arbeitsbereich für Begabungsforschung, Institut für Psychologie, Karl-Franzens-Universität Graz

Korreferent: Prof. Dr. Bert De Smedt, Parenting and Special Education Research Unit, Faculty of Psychology and Educational Sciences, Katholieke Universiteit Leuven

Weitere Mitglieder der Prüfungskommission:

Dr. Nivedita Mani, Nachwuchsgruppe Spracherwerb, Georg-Elias-Müller-Institut für Psychologie, Georg-August-Universität Göttingen

Prof. Dr. Margarete Boos, Abteilung für Sozial- und Kommunikationspsychologie, Georg-Elias-Müller-Institut für Psychologie, Georg-August-Universität Göttingen

Prof. Dr. Uwe Mattler, Abteilung für Experimentelle Psychologie, Georg-Elias-Müller-Institut für Psychologie, Georg-August-Universität Göttingen

Prof. Dr. Lars Penke, Abteilung für Biologische Persönlichkeitspsychologie, Georg-Elias-Müller-Institut für Psychologie, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 22.03.2018

Abstract

Test anxiety can hinder students from achieving their full potential in evaluative situations, such as tests or examinations. Converging evidence suggests that performance-related worries impair the working memory of these students. However, the mechanisms by which worries affect the working memory of test-anxious students remain poorly understood. The present work aimed to fill this gap by comparing the electroencephalography (EEG) of lower and higher test-anxious students performing a cognitive task in both a low and a high pressure condition. Two studies addressed the response monitoring in test-anxious students by analyzing the error-related negativity (ERN) – an event-related potential occurring shortly after an erroneous response. Results revealed that the ERN was enhanced by performance pressure in higher but not in lower test-anxious students. The third study assessed the working memory costs of performance pressure by analyzing frontal midline theta (FMΘ) – a signal known to reflect working memory processes. Results showed that FMΘ was increased by performance pressure, especially in students with higher test anxiety. Across studies, self-reported worries during the task were unrelated to EEG measures and test-anxious students did not show performance impairments. Taken together, results demonstrate that test-anxious students exhibit increased neural activity under performance pressure, suggesting that they need compensatory effort to maintain performance in evaluative situations.

Acknowledgements

Preliminary Note

In the following text, I will refer to the authors in plural (i.e., using the pronoun "we" instead of "I"). This acknowledges that my colleagues at the Section of Educational Neuroscience, headed by Professor Roland Grabner, as well as Professor Bert De Smedt of the University of Leuven have significantly contributed to the present work. However, I hereby state that the present studies are based on my own research ideas and have been both contrived and conducted leaded by me.

Three empirical investigations of the neurocognitive mechanisms underlying test anxiety are presented. The first study has been published in a peer-reviewed journal and the following text appears unchanged. However, some of the figures of the original publications have been adjusted to be consistent with the present format. The two other studies are manuscripts in preparation and the text was shortened for this thesis in order to avoid redundancy.

Original publication

Schillinger, F. L., De Smedt, B., & Grabner, R. H. (2016). When errors count: an EEG study on numerical error monitoring under performance pressure. ZDM - Mathematics Education, 48(3), 351–363. http://doi.org/10.1007/s11858-015-0746-8 (Erratum published)

Other publications by the author cited in the thesis

Schillinger, F. L., Vogel, S. E., Diedrich, J., Grabner, R. H. (2018). Math anxiety, intelligence, and performance in mathematics: Insights from the German adaptation of the Abbreviated Math Anxiety Scale (AMAS-G). Learning and Individual Differences, 61, 109-119.

Schillinger, F. L. (2016). Event-related potentials (ERPs) reflecting feedback and error processing in the context of education. Zeitschrift für Psychologie, 224(4), 286–289. http://doi.org/10.1027/2151-2604/a000264

Table of Contents

List of Figures

Supplementary Figures

List of Tables

Supplementary Tables

List of Abbreviations

AC	Accuracy
ACC	Anterior cingulate cortex
AIC	Akaike information criterion
ANOVA	Analysis of variance
AVG	Average
CRN	Correct response negativity
DLPFC	Dorsolateral prefrontal cortex
DSM	Diagnostic and statistical manual of mental disorders
EEG	Electroencephalography
ERD	Event-related desynchronization
ERN	Error-related negativity
ERP	Event-related potential
ERS	Event-related synchronization
FAS	Faces anxiety scale
fMRI	Functional magnetic resonance imagining
FMϴ	Frontal midline theta
GCSE	General certificate of secondary education
GPA	Grade point average
HTA	High test anxiety
ICA	Independent component analysis
LTA	Low test anxiety
M	Mean
PAF	German test anxiety inventory (Prüfungsangstfragebogen)
RT	Response time
SAT	Scholastic assessment test
SD	Standard deviation
SE	Standard error
STAI	State-trait-anxiety inventory

1. General Introduction

We live in a test-conscious, test-giving culture in which the lives of people are in part determined by their test performance.

Seymour B. Sarason (1959, p. 26)

Stated over five decades ago, this sentence holds true nowadays more than ever. Without doubt, tests and examinations play an important role in shaping the career of individuals in contemporary societies. For instance, consider the educational path towards becoming a psychologist in Germany. To enroll for a bachelor degree in psychology, most German universities require that students have achieved a certain Grade Point Average (GPA) in high school. In recent years, this GPA standard has been fairly competitive, so that only a fraction of applicants have been admitted to study psychology (see Formazin, Schroeders, Köller, Wilhelm, & Westmeyer, 2011). Or consider the usage of standardized tests for college admission in the United States. Here, the achievement in standardized tests, such as the Scholastic Assessment Test (SAT) or the American College Test (ACT), is one of the most important factors for a successful college application (see Atkinson & Geiser, 2009). These are just two examples in which the evaluation of performance has a direct and far-reaching impact on people's lives.

The increasing importance of performance evaluation has been linked to a rise of anxiety in students (see Pekrun & Götz, 2006; Zeidner, 1998). For instance, anxiety has been shown to be the most frequent emotion reported by university students regarding performance evaluation (Pekrun, 1992). Similarly, fear of failure has been suggested to have increased in school children (see McDonald, 2001). Such anxiety towards evaluation is commonly referred to as test anxiety in the literature (see Zeidner, 1998).

A study by Folin, Denis, and Smillie (1914) is considered the first empirical investigation on test anxiety (see Spielberger & Vagg, 1995). Folin and colleagues reported that the level of glycosuria (i.e., the excretion of glucose into the urine) was elevated in about one fifth of medical students after taking an important examination. The authors suggest that the rise of glycosuria in these students is due to increased levels of anxiety during the examination. However, systematic research on test anxiety started did not start until the 1950s with the work by Mandler and S. B. Sarason at Yale University. Their pioneering work included formulating the first specific theory on test anxiety, introducing a questionnaire to assess test anxiety, and demonstrating that test anxiety is related to lower performance in standardized tests (Mandler & Sarason S. B., 1952; S. B. Sarason & Mandler, 1952). Since then, the number of studies on

test anxiety have been constantly increasing with particularly booming years in the 1980s (see Stöber & Pekrun, 2004; Zeidner, 1998).

Results demonstrate that test anxiety is a prevalent problem affecting students across educational levels with female students being especially at risk (see Hembree, 1988). Strikingly, test anxiety is not just an "unpleasant feeling" but is inversely related to the well-being, self-esteem, and physical health of students (see Hembree, 1990; Zeidner, 1998). Moreover, students with high test anxiety show in average lower achievements in school and in standardized tests than their lower test-anxious classmates (for meta-analyses, see Ackerman & Heggestad, 1997; Hembree, 1988). Converging evidence suggest that this link between test anxiety and test performance is due in part to the detrimental effect of performance-related worries on the working memory of test-anxious students (see Hembree, 1988; Moran, 2016; Wine, 1971). However, the specific mechanisms by which worries affect the working memory of test-anxious students remain an area of active research (see Berggren & Derakshan, 2013; Eysenck, Derakshan, Santos, & Calvo, 2007; Maloney, Sattizahn, & Beilock, 2014). Neuroimaging techniques, such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), promise to significantly contribute to the unraveling of these mechanisms. Given the great importance of performing well in evaluative situations to succeed in modern societies, there is an urgent need to improve our understanding of the mechanisms underlying test anxiety by using both behavioral and neuroscientific research techniques.

1.1. Definition and Phenomenology of Test Anxiety

Test anxiety has not been listed as a mental disorder in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013) or the International Statistical Classification of Diseases and Related Health Problems (ICD-10; World Health Organization, 1992).[1] Therefore, there is no generally accepted definition of test anxiety in the literature and descriptions of the nature and phenomenology of test anxiety vary between authors. Spielberger (1972) has suggested differentiating between anxiety as a relatively stable personality trait (i.e., trait anxiety) and anxiety as transient state (i.e., state anxiety). Based on this distinction, test anxiety can be conceptualized as a situation-specific form of trait anxiety (Spielberger, Anton, & Bedell, 1976). Accordingly, Zeidner (1998) defines test anxiety as a "set of phenomenological, physiological, and behavioral responses that accompany concern about possible negative consequences or failure on an exam or similar evaluative situation" (p. 17).

---

[1] Note that in the recently revised DSM-5 it is now possible to specify for a social anxiety disorder that it is restricted to "speaking or performing in public" (p. 203).

Evaluative situations are most common in school and universities in which intellectual abilities are assessed by class tests, written and oral exams, or standardized tests. However, the anxiety over the evaluation of sensorimotor skills can also be reasonably subsumed under the concept of test anxiety. For instance, imagine an audition of a student applying for a music school or conservatory. Furthermore, performance is often evaluated in the performing arts and in sports. Typically, in situations in which sensorimotor skills are evaluated an audience or a jury is present. DeCaro and colleagues (2011) have therefore suggested differentiating between two type of performance pressure, namely, monitoring pressure and outcome pressure. While monitoring pressure refers to the evaluation of a performance by an observer or an audience, outcome pressure arises when incentives for optimal performance are large or poor performance is associated with negative consequences. In most evaluative situations, aspects of both types of performance pressure are present. However, the salience of monitoring pressure is typically higher when sensorimotor skills are evaluated (e.g., audition) than when intellectual skills are evaluated (e.g., written exam) (see DeCaro et al., 2011).

The phenomenon that students often show lower performance in high-stake tests than expected given their skill level has been referred to as *choking under pressure* in the literature (Baumeister, 1984; Beilock & Carr, 2001; Beilock, Kulp, Holt, & Carr, 2004; Running, Ligon, & Miskioglu, 1997). Students with high levels of test anxiety have been suggested to be especially prone to choke under pressure (Beilock et al., 2004; Ramirez & Beilock, 2011). As such, both concepts are closely related: while *choking under pressure* describes performance impairments due to performance pressure, test anxiety refers to individual differences in responding to performance pressure.

Test anxiety has been discussed as being related to other forms of anxiety, especially to general anxiety and to math anxiety. General anxiety has been conceptualized as a trait anxiety that is not specific to a situation but affects individuals in various aspects of their life (Spielberger, 1972). In contrast, test anxiety is thought to be specific to evaluative situations. Not surprisingly, these two forms of anxiety have been shown to be correlated (for a meta-analysis, see Hembree, 1988). Indeed, individuals with high general anxiety are also likely to be anxious towards evaluation. Although empirically related the two forms of anxiety are addressed as distinct in the literature (Hembree, 1988; Zeidner, 1998). Math anxiety, in turn, refers to the "feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations" (Richardson & Suinn, 1972, p. 551). Two seminal meta-analyses by Hembree revealed that test anxiety (Hembree, 1988) and math anxiety (Hembree, 1990) share some key commonalities. For instance, they are both related to lower academic performance—in the case of test anxiety in tests in general, in the case of math anxiety in tests on mathematics (Hembree, 1990). However, measures of test anxiety and math anxiety have been reported to

correlate only with small-to-medium effect sizes with each other while different measures of the same construct are highly correlated (Dew, Galassi, & Galassi, 1983; Hembree, 1990; Schillinger, Vogel, Diedrich, & Grabner, 2018). Therefore, test anxiety and math anxiety are widely considered as distinct yet related forms of anxiety in the literature.

Regarding the symptoms of test anxiety, Liebert and Morris (1967) have suggested to differentiate between a cognitive component, labeled as worry, and an affective component, labeled as emotionality. Worry can be described as "any cognitive expression of concern about one's own performance" (Liebert & Morris, 1967, p. 975), including ruminations, negative or interfering thoughts, and catastrophizing. Emotionality, in contrast, refers to affective or autonomic reactions towards evaluative situations, such as accelerated heartbeat, sweating, muscle tension, nausea, or dizziness (Hembree, 1988; Liebert & Morris, 1967). In addition, some authors have emphasized the role of behavioral and motivational symptoms associated with test anxiety, most prominently the tendency to avoid the test situation (Pekrun et al., 2004; Zeidner, 1998).

Taken together, test anxiety is conceptualized in the present work as a situation-specific trait anxiety over the evaluation of intellectual or sensorimotor skills having a cognitive and an affective facet.

1.2. Assessing Test Anxiety

The first questionnaires to assess test anxiety, including the Test Anxiety Questionnaire (TAQ; Mandler & S. B. Sarason, 1952), the Test Anxiety Scale for Children (TASC; S. B. Sarason, Davidson, Lighthall, Waite, & Ruebush, 1960), and the Test Anxiety Scale (TAS; I. G. Sarason, 1958; I. G. Sarason & Ganzer, 1962), conceptualized test anxiety as a unidimensional construct. This view was challenged by factor analytic findings showing that the items of these questionnaires in fact loaded on multiple factors (Gorsuch, 1966; Sassenrath, 1964). In turn, Liebert and Morris (1967) suggested differentiating in the assessment of test anxiety between worry and emotionality as discussed in the previous section. Based on this suggestion, Spielberger (1980) introduced the 20-item Test Anxiety Inventory (TAI) with a subscale for each of the two facets of test anxiety. Items are rated on 4-Point Likert scale with labels ranging from 1 (almost never) to 4 (almost always). To this day, the TAI is one of the most popular questionnaires to assess test anxiety (see Szafranski, Barrera, & Norton, 2012).

The TAI was adapted into German (TAI-G) by Hodapp and colleagues (Hodapp, 1991; Hodapp, Laux, & Spielberger, 1982) including several revisions. Firstly, while in the original TAI items refer to situations before, during, and after an examination, the German version focuses on individuals' responses in the test situation. Secondly, the worry component was

further divided into worry about the test and task-irrelevant, interfering thoughts (see also I. G. Sarason, 1984). Finally, a new subscale assessing the lack of confidence was introduced (e.g. "I am convinced that I am going to perform well"). The four subscales of the questionnaire (i.e., worry, emotionality, interference, lack of confidence) could be validated by means of confirmative factor analyses, and all subscales show a good-to-excellent internal consistency (Hodapp & Benson, 1997; Keith, Hodapp, Schermelleh-Engel, & Moosbrugger, 2003). Most recently, the number of items of the TAI-G was reduced to five items for each subscale (in total 20) and the questionnaire was renamed into German Test Anxiety Inventory (Prüfungsangstfragebogen, PAF) (Hodapp, Rohrmann, & Ringeisen, 2011). The PAF was shown to be reliable and valid psychometric tool and was normed on a large sample of both school students and university students (Hodapp et al., 2011). In conclusion, the PAF can be considered as the state-of-the-art questionnaire to assess test anxiety in German-speaking students.

1.3. Prevalence of Test Anxiety

Surprisingly, few data are available on the prevalence of test anxiety. One reason for this might be the different definitions of test anxiety within the literature as well as the various instruments that have been used to assess test anxiety. In addition, there has been a lack of adequate norms to classify which levels of test anxiety can be considered as heightened (see McDonald, 2001). In the following, we will first summarize studies on the prevalence in school children, followed by studies on the prevalence in college and universities students. Finally, we will briefly discuss differences in test anxiety between groups of students.

Hill and Wigfield (1984) have speculated that 20 to 25% of the elementary and secondary school students in the United States would have heightened levels of test anxiety. Recent studies on test anxiety in secondary school students in the United Kingdom (UK) suggest that the prevalence might be slightly lower, with a reported mean prevalence rate of 15.1% (for a meta-analyis, see D. Putwain & Daly, 2014). Since most of these studies had rather small sample sizes, D. Putwain and Daly (2014) also collected data from a more comprehensive sample (2435 secondary school students in 11 schools). Results revealed that 16.4% of the surveyed students reported themselves to be high test-anxious, with twice as many female students (22.5%) as male students (10.3%) scoring above the cut-off. Regarding the prevalence of test anxiety in Germany, Zech (1979) has reported that as many as 51% of the surveyed secondary school students exhibit heightened levels of test anxiety. However, more recent studies are lacking and the prevalence of test anxiety in German students might be more similar to that of the students in the UK.

In a study by Chapell and colleagues (2005), 13.6 % of the university students were classified as having high test anxiety. While there was no difference between Master's and PhD students, undergraduate students showed significantly higher level of test anxiety than graduate students. This estimate of the prevalence of test anxiety in higher education is in line with a large annual survey conducted by the German student services (Middendorf et al., 2016). In 2016, 13% of the surveyed students reported to have a need for counseling regarding high levels of test anxiety. Again, the proportion of test anxiety was significantly higher in female students (15%) than in male students (10%). The lower prevalence rate of test anxiety in universities students than in school students also correspond to meta-analytic findings regarding the change of the average test anxiety across students in the course of education (Hembree, 1988). More specifically, test anxiety was reported to occur as early as in the first grade and to increase in subsequent grades until reaching a plateau in grades eight to nine. Through grades 10-12 as well as in college the average level of test anxiety was then shown to slightly decrease (but cf. D. W. Putwain, 2007).

Female students were shown to report higher levels of test anxiety than male students in both secondary school (D. Putwain & Daly, 2014) and higher education (Middendorf et al., 2016). Hembree (1988) could demonstrate that this gender gap is significant across all levels of education. Similarly, studies suggest that test anxiety tends to be higher in African-American and Hispanic students (Hembree, 1988). However, more recent studies did not find a significant difference between Caucasian and African-American students in elementary school (Beidel, Turner, & Trager, 1994; Turner, Beidel, Hughes, & Turner, 1993) and current data regarding test anxiety in Hispanics as well as in minorities in Germany are missing.

In summary, test anxiety affects students across all educational levels, with female students being especially at risk. While evidence suggest that about one fifth of school students exhibit significant levels of test anxiety, the prevalence rate of test anxiety in higher education has been estimated to be around 10-15%.

1.4. Test Anxiety and Test Performance

From the beginning of empirical investigations into test anxiety, researchers have been interested in the question of how test anxiety is related to performance in tests and examinations. In a seminal study, S. B. Sarason and Mandler (1952) addressed this question by analyzing academic achievements as a function of test anxiety. The group of students with the highest test anxiety showed significantly lower scores in standardized tests including the scholastic aptitude test (SAT) and the mathematics aptitude test (MAT) as compared to those students with the lowest test anxiety. Across students, test anxiety was negatively correlated with the scores of an intelligence test administered in a group setting.

Similar findings have been reported regarding the link between anxiety and performance in school children. Hill and Sarason (1966) compared the performance in standardized tests of the 10% most test-anxious fifth and sixth graders with those of the 10% least test-anxious students. Results revealed that while the former group was lagging almost a year behind in reading and mathematics the latter groups was well ahead compared to national norms. In another study, D. W. Putwain (2008) analyzed how individual differences in test anxiety relate to the General Certificate of Secondary Education (GCSE), a test taken at the end of compulsory secondary schooling in the UK. There was a small but significant inverse relationship between test anxiety and grades in the GCSE. The difference between high test-anxious and low test-anxious students equated to almost one grade in the GCSE (D. Putwain & Daly, 2014).

The meta-analysis by Hembree (1988) provides a fine-grained analysis of the relationship between test anxiety and different facets of academic performance including intelligence tests, aptitude/achievement tests, and grades (see also Ackerman & Heggestad, 1997; Seipp, 1991). Test anxiety turned out to be consistently related to lower intelligence scores throughout education (i.e., from grade 1 to postsecondary). Similarly, high test-anxious students performed lower in aptitude/achievement tests, but this relationship did not become significant until grade 4. Comparing high test-anxious to low test-anxious students across standardized tests (i.e., intelligence and aptitude/achievement tests) revealed a mean effect size of -0.48. Also, course grades and the GPA were negatively correlated with test anxiety although effect sizes appear to be slightly smaller. This was confirmed by Chapell and colleagues (2005) who reported a significant but small inverse relationship between test anxiety and the GPA in both undergraduate ($r = -.15$) and graduate students ($r = -.09$). Finally, Hembree (1988) has reported that, across the different measures, test performance was more closely related to worry than to emotionality within test anxiety (see also Cassady & Johnson, 2002; Seipp, 1991)

Taken together, there is a wealth of data showing that test anxiety and especially the worry component is related to poor academic performance throughout education.

1.5. Accounting for the Anxiety-Performance Link

Given that test anxiety is inversely related to test performance, the question arises how this link can be accounted for. In general, two types of models, assuming different directions of causality, have been proposed in the literature: interference models and deficit models (Hembree, 1988; Sommer & Arendasy, 2014; Tobias, 1985). According to the interference models, test anxiety exercises a direct and detrimental influence on test performance. Thus, test anxiety is thought of as being the cause of the lower test performance of test-anxious

students. Alternatively, deficit models assume that students who struggle with tests in the first place are more prone to worry about their performance and thus to become test-anxious. Here, poor test performance is thought to be the actual cause of test anxiety. In the following, we will discuss both models in more detail and summarize the evidence for either direction of causality. We conclude by synthesize a working model for the present empirical investigations.

1.5.1. Interference Models

In one of the first theories on test anxiety, Mandler and S. B. Sarason (1952) postulated that a test situation evokes two kinds of learned drives in students: task drives and anxiety drives. Task drives can be described as the need to complete a task as well as possible in an evaluative situation. These drives are reduced by responses that contribute to completing the task at hand, such as solving a problem. Anxiety drives, in contrast, are thought to be associated with two classes of responses. Task-centered responses are functionally equivalent to the responses triggered by task drives and will result in an increased effort to complete the task. Self-centered responses, instead, will be manifested in "[…] feelings of inadequacy, helplessness, heightened somatic reaction, anticipations of punishment or loss of status and esteem, and implicit attempts at leaving the test situation" (Mandler & S. B. Sarason, 1952, p. 166). According to Mandler and Sarason (1952), test-anxious students are characterized by a high anxiety drive as well as the tendency toward self-centered responses to anxiety. As a consequence, students with high test anxiety experience in evaluative situations what Mandler and S. B. Sarason (1952) phrased *anxiety-produced interference.*

In a seminal review, Wine (1971) summarized the evidence for the notation that test anxiety is interfering with task performance. Based on the findings that especially the worry component seems to predict task performance (Liebert & Morris, 1967), the review concluded with giving a more attentional account of interference:

> *An attentional interpretation states simply that the reason "worry" debilitates task performance is that it is attentionally demanding and distracts attention from the task.* (Wine, 1971, p. 100)

As such, Wine (1971) translated the behavioristic theories by S.B. Sarason into a cognitive-attentional theory of interference (see Zeidner, 1998), which has been also referred to as distraction theory in the literature (e.g. Beilock & Carr, 2001; Markman, Maddox, & Worthy, 2006; Yu, 2015).

The idea that worries are the key mechanism by which anxiety affects cognitive performance has been recurring in the more recent processing efficiency theory (Calvo & Eysenck, 1992). This theory provides a general framework to explain how anxiety is related to

cognition. An important innovation to the distraction theory by Wine (1971) – reminiscent of the two responses to the anxiety drive postulated by Mandler and S. B. Sarason (1952) – is that worry is thought to have two reverse effects on test performance. Firstly, "worrisome thoughts consume the limited attentional resources of working memory, which are therefore less available for concurrent task processing." (Eysenck et al., 2007, p. 337). This detrimental effect of worries is conceptualized to primarily affect the phonological loop and the executive functions of the working memory (see Baddeley, 2003). Secondly, being worried about failing is activating compensatory efforts to maintain task performance. Such compensatory efforts are thought to comprise both an increase in effort (e.g., "trying harder" to complete a task) and the usage of auxiliary strategies (e.g., articulatory rehearsal when reading). If such resources are available, task performance (i.e., the performance effectiveness) will remain unaffected. However, given that more cognitive resources have been used to maintain task performance, the processing efficiency is reduced. In general, anxiety is predicted to have a greater effect on processing efficiency than on performance effectiveness. According to Eysenck and colleagues (2007), this prediction can be tested by analyzing both response times and accuracies in a task. While reduced accuracy in a task would indicate a lower performance effectiveness, prolonged response times are thought to reflect reduced processing efficiency.

In the attentional control theory, a recent update of the processing efficiency theory, Eysenck and colleagues (2007) further specify how anxiety is thought to exercise its influence on the executive functions. The executive functions are further differentiated into (1) inhibition (i.e., to prevent interference from task-irrelevant stimuli), (2) shifting (i.e., to dynamically adapt to task demands), and (3) updating (i.e., to monitor working memory representations) (see Miyake et al., 2000). The attentional control theory states that the shifting and the inhibition function require more attentional control than the updating function and are therefore primarily affected by anxiety. However, in test-anxious individuals, the updating function has also been reported to be impaired when tested in stressful conditions (Calvo, Ramos, & Estevez, 1992; Darke, 1988).

Both the distraction theory (Wine, 1971) and the attentional control theory (Calvo & Eysenck, 1992; Eysenck et al., 2007) assume that the detrimental effect of worries on the working memory is central for explaining the link between test anxiety and cognitive performance. There is a wealth of evidence in support of this notion, including that (1) high test-anxious students show lower task performance than low test-anxious students when instructions are ego-involving but not when instructions are neutral, (2) test-anxious students report more worries during a test and that worries can impair students' working memory, (3) test performance of test-anxious students improves after interventions aiming to reduce performance-related worries.

A bulk of studies have addressed the role of ego-involving or ego-threatening instructions on test performance. Such instructions are usually implemented by informing participants that the task at hand is diagnostic about their intelligence or cognitive abilities. According to Wine (1971), results suggest that high test-anxious students perform more poorly if instructions are ego-involving compared to neutral task instructions. The opposing pattern of results holds true for low test-anxious students. Directly comparing the two groups of students shows that only with ego-involving instruction high test-anxious students are outperformed by low test-anxious students. In a recent meta-analysis, Moran (2016) has summarized studies addressing the effect of ego-involving instruction on simple span measures. In the digit span task, for instance, participants are asked to memorize and recall a serially presented sequence of digits. The digit span is the longest sequence of digits a participant can reproduce. Across studies, ego-involving instructions had a large aggregated effect ($g = -.62$) on simple span measures, i.e., fewer items could be recalled as compared to neutral task instructions. Finally, evidence of the detrimental effects of ego-involving instructions comes from studies on *choking under pressure* (Beilock, 2008; Beilock & Carr, 2001, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004; Beilock, Rydell, & McConnell, 2007; Byrne, Silasi-Mansat, & Worthy, 2015; Chib, De Martino, Shimojo, & O'Doherty, 2012; DeCaro et al., 2011; Lee & Grafton, 2014; Markman et al., 2006; Mattarella-Micke, Mateo, Kozak, Foster, & Beilock, 2011; Mobbs et al., 2009; Ramirez & Beilock, 2011; Yu, 2015). In one study, Ramirez and Beilock (2011) asked participants to perform two blocks of complex arithmetic problems. After completing the first block, participants were randomly assigned to one of two conditions. In the control condition, they were instructed to continue solving the problems in the second block just as in the first block. In the pressure condition, instead, participants were exposed to an ego-involving test scenario in which they were filmed during the task and received a monetary reward if they could improve their performance in the second block. Results revealed that participants in the control condition showed a better performance in the second block as compared to the first block. Participants in the pressure condition, however, showed a marked drop in performance in the second block.

Students with test anxiety have been repeatedly shown to report more worries during a test as compared to lower test-anxious students (for reviews, see Wine, 1971; Zeidner, 1998) In an elegant study, Deffenbacher (1978) compared low and high test-anxious students in solving anagrams either in a low stress or in a high stress condition. After completing the task, students were asked to rate their emotions during the test and to estimate how much time they had effectively spent on the task. High test-anxious students in the stress condition solved fewer anagrams and reported more worries than both high test-anxious students in the low stress condition and low test-anxious students in the high stress condition. Strikingly, high test-anxious students under stress reported spending only 60% of their time working on the task

as compared to 77.65% reported by test-anxious students without stress. Another line of research has addressed the causal role of worries by experimentally inducing worries while participants perform random generation tasks. In such tasks, participants have to produce a random sequence of items (e.g., by pressing different buttons) and hence try to avoid repeating or stereotyped patterns. These tasks have been shown to have relatively high working memory demands (see Moran, 2016). Hayes and colleagues (2008) compared low and high test-anxious students performing a random generation task in two conditions. While participants were asked in the worry condition to think about a personal topic that had been bothering them, they should think about a positive topic in the positive thought condition. Comparing the randomness of the produced sequences between groups and conditions revealed that there was no difference between groups in the positive thought condition. However, high test-anxious students showed a lower performance in the worry condition as compared to the low test-anxious students. Across studies, Moran (2016) reported a significant effect of engaging in worries on random generation tasks with a medium effect size ($g = -.38$).

Finally, evidence for the interference of test anxiety on test performance comes from studies on the treatment of test anxiety. Hembree (1988) concluded in his meta-analysis that both behavioral and cognitive-behavioral intervention are effective in reducing test anxiety. Each type of treatment was accompanied by an improvement in performance in both standardized tests and the GPA. Instead, improving study skills alone had no significant effects, neither on the level of test anxiety nor on test performance. In a more recent study by Ramirez and Beilock (2011), a short intervention of expressive writing, which is thought to alleviate worries and ruminations, was tested on the *choking under pressure effect* on complex arithmetic as discussed above. Before starting with the second block of complex arithmetic, participants were asked to either write about their feelings towards the upcoming test (expressive writing) or about an unemotional event (unrelated writing). While students in the unrelated writing condition showed the expected drop in performance, this *choking under pressure effect* was neutralized in the expressive writing condition. A similar effect of expressive writing was also reported in ninth graders facing their final high-school exam in biology. Students were assigned to an expressive writing condition or to think about a topic that would not be covered on the exam. In the analysis, participants were further divided into low test-anxious and high test-anxious students. Strikingly, while test-anxious students in the expressive writing group improved in their final exam grade as compared to the control condition, no significant effect was found for low-test anxious students. This suggest that the intervention was effective in reducing worries in test-anxious students and that without this burden students could improve their performance in the final exam.

1.5.2. Deficit Models

While interference models of test anxiety have been the predominating view in the literature, deficit models have also received some support (see Zeidner, 1998). Researchers have suggested two types of deficits that can account for the lower performance of test-anxious students (Tobias, 1985). Firstly, students with high test anxiety might have lower study skills as compared to lower test-anxious students. According to this explanation, test-anxious students exhibit deficits in the initial acquisition and storage of the learning material. Secondly, test-anxious students might have inferior test-taking skills. Test-taking skills refer to strategies in a test, such as monitoring the time during the test or returning to difficult items at the end of the test (see Kirkland & Hollandsworth, 1980). In both scenarios, students with lower skills are thought to realize that they are doing poorly before or during a test and start worrying about their performance. As such, test anxiety is seen as an epiphenomenon of poor study or test-taking skills without having a direct effect on task performance. Indeed, there is some evidence that test-anxious students have lower study and test-taking skills, and that these skills are related to poor test performance.

Studies suggest that high test-anxious students have on average lower study skills than students with low test anxiety (for reviwes, see Tobias, 1985; Zeidner, 1998). In one study, Naveh-Benjamin and colleagues (1981) compared the study skills of low and high test-anxious undergraduate students in a psychology course. Results revealed that test-anxious students had specific problems acquiring the learning material throughout the semester. More specifically, they performed poorly in take-home examinations, reported problems learning the material in the course, and had problems picking out the important points in a reading assignment. Not surprisingly, these students received a worse course grade than their low test-anxious classmates. More recently, Cassady (2004) investigated the interplay between test anxiety and study skills across the learning–testing cycle. In the preparation phase, test-anxious students reported lower study skills and prepared less efficient notes that they were allowed to use in the actual test. The reported study skills were negatively correlated with students' test anxiety. However, while test anxiety was inversely related to course exam performance, there was no significant relation between study skills and performance.

Similarly, studies have linked test anxiety to deficits in test-taking skills (for a review, see Tobias, 1985). In one study, Bruch and colleagues (1983) addressed the question of whether test performance is predicted by test-taking skills and test anxiety over and above scholastic abilities (as assessed by the SAT). Students were tested in a multiple-choice test and a short math test while reporting their test-taking skills and their state anxiety. For both tests, test-taking skills explained significant variance in test performance over and above the

ability level. Instead, the level of anxiety during the test appeared to be only related to the performance in the math test and explained significant less variance than the test-taking skills.

1.5.3. Present Working Model

The present literature review revealed that both interference models and deficit models have received empirical support in the literature. This suggests that the relationship between test anxiety and test performance is to some degree bidirectional. According to a reciprocal model, students with low study or test-taking skills are more prone to feel anxious in the preparation phase as well as during a test. Being anxious in the test, in turn, will further harm the test performance of these students (for a detailed discussion, see Zeidner, 1998). This can result in a vicious cycle in which students become increasingly test-anxious while academic performance is declining (see also Carey, Hill, Devine, & Szücs, 2016). However, test anxiety has been reported to be only moderately correlated with study skills (Cassady, 2004) and test-taking skills (Bruch, 1981). Moreover, studies could dissociate test anxiety and skill deficits by comparing low and high test-anxious students with both low and high skills (Birenbaum & Pinku, 1997; Naveh-Benjamin, 1991). In light of this evidence, deficits in study and test-tasking skills might be best conceptualized as an additional factor in a more complex interplay between test performance and test anxiety (see Tobias, 1985).

The overwhelming majority of the reviewed studies suggest that being anxious in a test situation has a direct and detrimental effect on cognitive performance and that performance-related worries mediate this relationship. The aim of the present empirical investigations was therefore to better understand the effect of worries on the cognition of test-anxious students. For this, the effect of test anxiety on task performance was conceptualized based on the theories by Wine (1971) and Eysenck (Calvo & Eysenck, 1992; Eysenck et al., 2007). This distraction model is depicted in Figure 1. In this model, test-anxiety is conceptualized as a personality trait that interacts with the evaluative character of a situation in inducing performance-related worries in students. Theses worries, in turn, are thought to impair the working memory of students including the executive functions and the phonological loop (see Baddeley, 2003). As a consequence, working memory resources have to be split between the task at hand and the distracting worries. If the combined demands of task-related and worry-related processes exceed the available working memory resources, performance impairments will result.
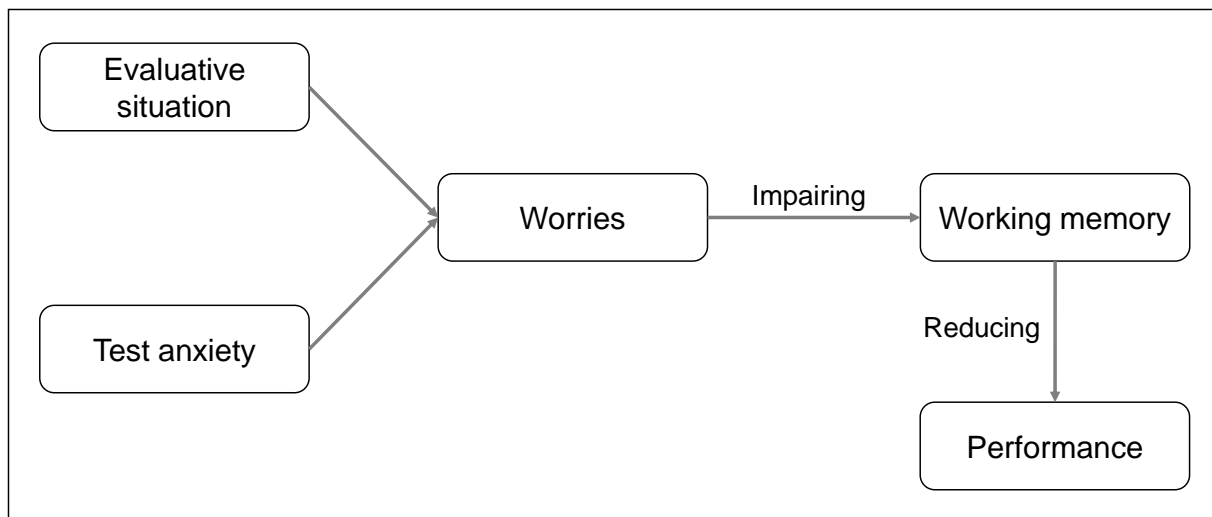
Figure 1. Distraction model of the effect of test anxiety on cognitive performance (Calvo & Eysenck, 1992; Eysenck et al., 2007; Wine, 1971).

1.6. Using Neuroimaging Techniques

Converging evidence suggest that performance-related worries impair the working memory of test-anxious students. However, the specific mechanisms by which worries affect the working memory of test-anxious students remain an area of active research (for recent reviews, see Berggren & Derakshan, 2013; Maloney, Sattizahn, & Beilock, 2014). Eysenck and colleagues have suggested that neuroimaging techniques, such as fMRI and EEG, are promising to unravel theses mechanisms (Derakshan & Eysenck, 2009; Eysenck & Derakshan, 2011; Eysenck et al., 2007). In particular, neuroimaging techniques can provide information about the working memory costs of anxiety that are not detectable by measuring overt behavior alone. For instance, an EEG study by Righi and colleagues (2009) addressed the effect of general anxiety on attentional control using a Sustained Attention to Response Task (SART). In the task, a series of digits between 1 and 9 was presented in quick succession. Participants had to response to all digits (Go trials) except for the 3 in which case a response had to be inhibited (NoGo trials). Results revealed that neither trait nor state anxiety were significantly related to task performance. However, students with higher anxiety exhibited an increased amplitude of the N2 in the NoGo trials as compared to students with lower anxiety. The N2 is an event-related potential (ERP) peaking between 200 and 350 ms and has been linked to attentional control (for a review, see Folstein & Van Petten, 2008). Thus, on the behavioral level, attentional control appeared to be unaffected by individual differences in anxiety. On a neural level, however, the increased amplitude of the N2 suggest that anxious students used compensatory effort for inhibiting a response in the NoGo trials. Similarly, T. L. Ansari and Derakshan (2011) have reported that high anxious students show an increased

frontal Contingent Negative Variation activity (CNV) during an antisaccade task, while there was no difference in the task performance between low and high anxious students. Furthermore, fMRI studies have linked anxiety to an increased activity in the dorsolateral prefrontal cortex (DLPFC) - a candidate region for implementing attentional control in the human brain (Basten, Stelzel, & Fiebach, 2011; Fales et al., 2008; but cf. Bishop, 2009).

In the studies reviewed above, high anxious individuals performed at about the same level as low anxious individuals but exhibited significantly increased neural activity. Consequently, Eysenck and colleagues have suggested that the anxiety-related increase in neural activity reflects compensatory effort by anxious individuals to maintain task performance (Derakshan & Eysenck, 2009; Eysenck & Derakshan, 2011; Eysenck et al., 2007). Such compensatory efforts are thought to comprise of both the increase in attentional control in the first place (e.g. "trying harder" to inhibit a response) and the usage of auxiliary strategies (e.g. articulatory rehearsal when reading).

1.7. Aims and Outline of Studies

Evidence suggest that increased neural activity in anxious individuals during a cognitive task reflects compensatory effort to maintain task performance (for reviews, see Derakshan & Eysenck, 2009; Eysenck & Derakshan, 2011; Eysenck et al., 2007). These studies have addressed the neural activity during cognitive tasks in participants varying in their level of general anxiety (T. L. Ansari & Derakshan, 2011; Basten et al., 2011; Bishop, 2009; Fales et al., 2008; Righi et al., 2009; Savostyanov et al., 2009). However, studies on the neural activity in test-anxious students are largely lacking. The present work aimed to fill this gap by directly addressing the neurocognitive mechanisms underlying test anxiety. For this, we related different measures in the human EEG to individual differences in test anxiety. Compared to other neuroscientific techniques (e.g., fMRI), the temporal resolution of EEG is superior (up to 1 ms) allowing us to gain fine-grained and precise temporal information about cognitive processes (see Schillinger, 2016). Based on the distraction model depicted in Figure 1, test-anxiety was conceptualized as a personality trait that interacts with the evaluative character of a situation in inducing performance-related worries. Therefore, in all three studies students with lower and higher test anxiety were tested performing a cognitive task in both a high pressure condition and a low pressure control condition. Table 1 provides an overview of the present studies including the study design and the EEG measures used.

Table 1. Overview of the present studies with study design (including the factor evaluation and test anxiety) and the different measures in the electroencephalography (EEG).

Study	Study Design		EEG
	Evaluation	Test Anxiety	
Response Monitoring Under Performance Pressure	Within	Continuous	ERN
Response Monitoring in High Test-anxious Students	Within	Extreme Groups	ERN
Working Memory Costs of Performance Pressure	Between	Continuous	FMΘ

Note. ERN = Error-related negativity, FMΘ = Frontal midline theta.

1.7.1. Studies on the Error-related Negativity (ERN)

A crucial skill for success in a test is to monitor ongoing responses and to dynamically adapt to errors (see Hirsh & Inzlicht, 2010). More specifically, students in a test situation have to evaluate their response to a given problem in a limited period of time. Was the given response correct, or did they commit an error? If an error has been committed, students need to increase attentional control in order to maintain task performance (Eysenck et al., 2007; Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004). However, little is known about how worries affect response monitoring in a test situation. Two present studies addressed this question by analyzing a negative deflection in the human EEG emerging shortly after an error has been committed. This error negativity (Ne; Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991) or error-related negativity (ERN; Gehring, Goss, & Coles, 1993) peaks around 50 to 100 ms relative to response onset and is most pronounced at frontocentral recording sites. The ERN is typically studied in reaction time tasks in which participants have to respond to a target stimulus by pressing one of two buttons (two-alternative forced choice), such as the Flanker, Go/NoGo, or Stroop paradigm (for a review, see Gehring, Liu, Orr, & Carp, 2012).

Different lines of research suggest that the amplitude of the ERN is modulated by worries. Firstly, the ERN has been shown to be enhanced in patients with mental disorders linked to worries and ruminations, including general anxiety disorder and obsessive-compulsory disorder (for a review, see Weinberg et al. 2011). In the same vein, the ERN has been demonstrated to be enhanced in individuals with higher general anxiety (Hajcak, McDonald, & Simons, 2003), negative affect (Hajcak, McDonald, & Simons, 2004; Luu, Collins, & Tucker, 2000), math anxiety (Suárez-Pellicioni, Núñez-Peña, & Colomé, 2013), perfectionism (Perrone-McGovern et al., 2017), and helplessness (Pfabigan et al., 2013). Secondly, worry has been shown to be the dimension of anxiety which is most closely related to the reported increase in the ERN (for a meta-analysis, see Moser, Moran, Schroder, Donnellan, & Yeung, 2013). For instance, Moser and colleagues (2012) reported that the ERN

amplitude was negatively correlated to worry but not to arousal in female undergraduate students. Finally, a recent study has shown that a short intervention of expressive writing, which is thought to alleviate worries and ruminations (see Ramirez & Beilock, 2011), is effective in reducing the ERN in individuals with chronic worry (Schroder, Moran, & Moser, 2018).

Taken together, studies have consistently demonstrated that anxiety and, in particular, worry is related to an enhanced ERN. At the same time, performance in tasks used to measure the ERN (e.g., Stroop) appears to be uncompromised in anxious individuals (for a meta-analysis, see Moser et al., 2013). Moser and colleagues (2013) have therefore suggested that the increased ERN in anxious individuals reflects compensatory efforts to maintain task performance. According to the *compensatory error-monitoring hypothesis*, anxious individuals focus their attention on internal worries thereby reducing resources dedicated to performing the actual task. As a result, anxious individuals have to increase their attentional control to maintain task performance, including monitoring ongoing responses and to dynamically adapt to errors. This compensatory effort is thought to give rise to more neural activity after committing an error as reflected in an increased ERN.

The present two studies aimed to better understand the effect of worries on the response monitoring in test-anxious students. In the first study, 18 female university students were tested in performing a numerical Stroop task in both a high pressure condition modeling a real-life test situation and a low pressure control condition. Test anxiety was assessed with the German Test Anxiety Inventory (Prüfungsangstfragebogen, PAF) (Hodapp et al., 2011) and correlated with both task performance and EEG indices of response monitoring. While behavioral performance data provided mixed evidence, EEG indices suggest changed response monitoring in the high pressure condition as well as in relatively test-anxious participants. The second ERN study aimed to extend these findings by assessing test anxiety more systematically and by directly relating worries to the amplitude of the ERN. For this, two groups of 25 students were selected from a larger pool of subjects aiming to maximize the difference between groups in test anxiety. The two extreme groups were tested as in Study 1 in a high pressure and a low pressure condition performing a numerical Stroop task. In addition, participants reported their worries regarding their performance throughout the experiment using a Faces Anxiety Scale (FAS).

1.7.2. Study on Frontal Midline Theta (FMΘ)

Another way to elucidate the mechanisms by which test anxiety affects the working memory of test-anxious students is by measuring the neural oscillations during a cognitive task. Mental activity has been shown to be accompanied by neural oscillations in different frequency bands (see, Klimesch, 1999). Most noticeably, frontocentral oscillations in the theta band (4-8 Hz), referred to as frontal midline theta (FMΘ), have been linked to working memory processes (for reviews, see Hsieh & Ranganath, 2014; Inanaga, 1998). In particular, FMΘ has been demonstrated to increase with the memory load in classical working memory tasks, such as the Sternberg task (e.g., Jensen & Tesche, 2002) and the N-back task (e.g., Gevins, Smith, McEvoy, & Yu, 1997). Interestingly, ERPs following uncertain events such as errors (error-related negativity; ERN), punishment (feedback-related negativity; FRN), and conflict-inducing stimuli (N2) have been recently shown to originate from a transient oscillatory synchronization of FMΘ (for reviews, see Cavanagh & Frank, 2014; Cavanagh & Shackman, 2015). These ERPs, in turn, have been consistently reported to increase with the level of anxiety (for a meta-analysis, see Cavanagh & Shackman, 2015). Based on these findings, Cavanagh and Shackman (2015) have proposed that the increased FMΘ in anxious individuals reflects the need for attentional control regarding uncertain or threatening events.

In the present study, we aimed to directly assess the effect of worries on the working memory of test-anxious students. According to the distraction model (see Figure 1), worries additionally tax the working memory of test-anxious students in an evaluative situation. Given that FMΘ is sensitive to working memory demands, we reasoned that FMΘ is increased in test-anxious students under performance pressure. To test this prediction, students with varying levels of test anxiety performed two blocks of a working memory demanding arithmetic task while reporting their worries (see Beilock et al., 2004). After completing the first block, participants were randomly assigned to one of two pressure conditions. While the control group was instructed to work through the second block of problems just as before, the pressure group was exposed to a high-pressure test scenario. Previous studies using this paradigm have reported that participants in the pressure group show a significant performance drop in the second block as compared to the control group (Beilock & Carr, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004; Ramirez & Beilock, 2011). In the present study, we related this choking effect to both individual differences in test anxiety and the modulation of FMΘ. To test whether test-anxious students show higher working costs under performance pressure due to worries, a mediation approach was used.

2. Study 1 – Response Monitoring Under Performance Pressure

2.1. Introduction

2.1.1. Choking Under Pressure in Mathematics

Acquiring mathematical knowledge and procedures through formal education plays an ever-increasing role in modern society (De Smedt et al. 2011; Grabner and Ansari 2010). Despite the importance of mathematical education, mathematics is often linked to worries and feelings of anxiety. According to the Programme for International Student Assessment (PISA), an alarming percentage of 61% of the interviewed students reported that they are worried to get poor grades in mathematics and 59% stated that they are afraid that it will be difficult for them in mathematics classes (OECD, 2013).

Worries related to mathematics have been suggested to be a potential cause of poor mathematics achievement. Especially in high-stake tests, when students try to perform at their best, such worries might hinder students from achieving their full potential (Ashcraft, 2002; Beilock, 2008). Across domains, students often show lower performance in high-stake tests than expected, given their still level (Beilock et al., 2004). In the literature, the term *choking under pressure* has been used to describe this phenomenon (Baumeister, 1984). According to the distraction account, a prominent theory to explain choking under pressure of academic skills, worries are the initial condition of a cognitive mechanism that ultimately leads students to fail (DeCaro et al., 2011; Eysenck et al., 2007; Wine, 1971). In detail, the theory claims that being in a test situation is inducing worries and task-irrelevant thoughts about the test, one's performance, and potential consequences of doing poorly. These ruminations are thought to coopt working memory, and when the combined demands of task-related and extraneous processing are exceeding the individual working memory capacity, this will result in performance impairments.

Acquiring mathematical concepts and applying mathematical procedures impose a considerably high working memory load on the learner (Raghubar, Barnes, & Hecht, 2010). Thus, following the distraction account, mathematics should be especially vulnerable to choking under pressure. This is in line with empirical evidence provided by Beilock and colleagues demonstrating that choking under pressure affects mathematical problem solving (for a review, see Beilock 2008). In an initial study, they compared the accuracy with which participants solved arithmetic problems in a high pressure condition to a low pressure control condition (Beilock et al., 2004). While participants were instructed to solve problems as good as possible in the control condition, they were subjected to performance pressure in the high pressure condition. For this, participants were told that they would receive a monetary reward for good performance as part of a team effort (outcome pressure). In addition, participants

were videotaped during the experiment and were led to believe that the recording would be used for educational purposes (monitoring pressure) (see DeCaro et al. 2011 for a detailed discussion of pressure situations). Results showed that the problem solving rate was impaired by performance pressure, but only for arithmetic problems with high working memory demands (see also, Beilock and DeCaro 2007; Ramirez and Beilock 2011).

2.1.2. The Role of Test Anxiety

While choking under pressure is a general phenomenon, students with test anxiety have been suggested to be particularly affected by it (Beilock et al., 2004). Test anxiety is a specific fear of failure before or during a test which is accompanied by increased arousal, tension, and bodily reactions—on a affective level—and worries, irrelevant thoughts, and catastrophizing—on a cognitive level (for a review, see Zeidner 2007). According to the distraction account, the tendency of worrying about a test is making test-anxious students more prone to fail in evaluative situations (Eysenck et al., 2007; Wine, 1971). In a study by Calvo and colleagues (1992), high test-anxious individuals exhibited inferior working memory capacity as compared to lower test-anxious participants, but only under evaluative stress condition. Moreover, in a longitudinal study, Ramirez and Beilock (2011) showed that test anxiety is linearly related to academic achievement in ninth-grade students. However, this relationship was alleviated by an expressive writing intervention which specifically aimed to reduce performance-related worries. Thus, choking under pressure depends on both situational performance pressure and individual test anxiety.

2.1.3. Response Monitoring in Test Situations

To succeed in an academic test, it is important to monitor ongoing responses and to dynamically adapt to errors (Hirsh & Inzlicht, 2010). More specifically, in a test situation, students have to evaluate their response to a given problem in a limited period of time. Was the given response correct, or did they commit an error? If an error has been committed, students need to take measures in order to uphold task performance. One way to adjust to an error is to increase cognitive control in order to avoid committing another error (Eysenck et al., 2007; Ridderinkhof et al., 2004). Previous research has highlighted the role of prefrontal brain structures for the implementation of cognitive control, including performance monitoring and behavioral adjustment to errors (Carter, 1998; Ridderinkhof et al., 2004).

So far, only few studies have addressed the question of how responses are monitored in the domain of mathematics. In one functional magnetic resonance imaging (fMRI) study, Ansari and colleagues (2011) compared the brain activation of highly mathematical competent

participants to relatively lower mathematical competent participants during arithmetic problem solving. Crucially, in the analysis, the authors directly compared correctly solved arithmetic problems to incorrectly solved arithmetic problems. Results indicated higher brain activity in prefrontal areas when an arithmetic error was committed. In the right dorsolateral prefrontal cortex, this effect was modulated by the mathematical competence of participants. Highly mathematical competent participants exhibited stronger activation in this area compared to relatively lower mathematical competent students. These results suggest that individuals with high mathematical competence exhibit improved performance monitoring during a mathematical task and implement greater cognitive control following the commission of an arithmetic error.

Furthermore, in mathematics, response monitoring seems to be especially crucial since mathematical problems are often associated with a single correct solution. For instance, there is only one appropriate solution to an arithmetic problem. Therefore, errors might be more salient in mathematics as compared to other school-related domains. Also, evaluating responses in mathematics is particularly important in order to avoid consequential errors.

2.1.4. The Error-related Negativity (ERN)

When looking at the brain response, as measured through EEG, a response in a choice task is typically followed by a negative deflection in the EEG signal, which has been shown to be more pronounced for erroneous responses (error-related negativity, ERN) than for correct responses (correct response negativity, CRN) (Falkenstein et al., 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). The ERN has been demonstrated in various tasks, including Flanker (e.g., Falkenstein et al. 1991), Go/NoGo (e.g., Kim et al. 2007), color-naming Stroop (e.g., Hirsh and Inzlicht 2010), and—most recently—numerical Stroop task (Suárez-Pellicioni et al., 2013). It peaks around 50 to 100 ms relative to response onset and is most pronounced at centro-parietal recording sites. Most studies have suggested the anterior cingulate cortex (ACC) as a candidate brain region for generating the ERN (for a review, see Gehring et al. 2012).

The ERN has been proposed to reflect activity related to cognitive processes underlying response monitoring (Falkenstein et al., 1991; Gehring, Goss, Coles, et al., 1993). More specifically, in a choice reaction time, an error is often committed before the processing of the target stimulus is completed. As the response is carried out, e.g., by pressing a button, the processing of the stimulus continues. Then, the difference of the correct response, based on the ongoing stimulus processing, is compared to the actual response, which has been initiated. When the two representations do not match, an error signal arises, which is thought to be

reflected in the ERN (Coles, Scheffers, & Holroyd, 2001). However, the precise functional significance of the ERN is a subject of ongoing debate (for a review, see Gehring et al. 2012).

While the ERN is thought to reflect activity specific to error processing, the difference potential between CRN and ERN (ΔERN) has been suggested to reflect activity more broadly related to response monitoring (Riesel, Weinberg, Endrass, Meyer, & Hajcak, 2013). Thus, it is common to analyze in addition to the ERN the difference potential ΔERN. The test-retest reliability of both measures have been shown to be excellent over a period of two weeks (Olvet & Hajcak, 2009a) and moderately high over a period of up to 2.5 years (Weinberg & Hajcak, 2011). Furthermore, when measured with different tasks (viz., Flanker, Stroop, Go/NoGo) the ERN has been demonstrated to exhibit a moderately high and the ΔERN a slightly higher convergent validity (Riesel et al., 2013). Taken together, the ERN and the ΔERN can be considered as stable, trait-like electrophysiological measures (see Weinberg and Hajcak 2011).

Interestingly, converging evidence suggest that the ERN is sensitive to affective and motivational factors, including feelings of anxiety. Specifically, individual differences in trait anxiety (e.g. Hajcak et al. 2003) and negative affect (e.g. Hajcak et al. 2004) have been shown to be directly related to an increased ERN amplitude. In a recent study, Suárez-Pellicioni and collegues (2013) showed that math-anxious students exhibit an enhanced ERN in a numerical Stroop task, but not in a classical, color-naming Stroop. In addition, patients with a diagnosis of generalized anxiety disorder as well as obsessive-compulsive disorder are characterized by an increased ERN amplitude (for a review, see Weinberg et al. 2011). Among different dimensions of anxiety, apprehension or worry has been shown to be most closely associated with the ERN (Moser et al., 2013). Thus, in the present study, we expected the ERN/ΔERN to linearly increase with individual level of test anxiety.

In addition to individual differences, the ERN has been demonstrated to be modulated by situational factors inducing performance pressure (Ganushchak & Schiller, 2008; Hajcak, Moser, Yeung, & Simons, 2005; Kim, Iwaki, Uno, & Fujita, 2005). In an elegant study, Hajcak and colleagues (2005) investigated how the ERN is modulated by monetary incentive, on the one side, and by social evaluation, on the other side, using a Flanker task. In the first experiment, monetary incentive was varied on a trial-by-trial basis using a cue ahead of the target stimulus. Participants could either earn "5" or "100" points for responding correctly to the target stimulus. (Participants were instructed that points were transferred into money after completion of the experiment.) Results showed that the amplitude of the ERN following errors in trials with high value was significantly higher than in trials with low value. In the second experiment, a control condition in which participants were instructed to respond as quickly and as accurately as possible was compared to an evaluative condition. In the evaluative condition, participants were monitored throughout the experiment by a research assistant and were told

that their performance would be compared to the result of other students. All participants were tested in both conditions within a single EEG session starting with either of the conditions in a counterbalanced order. Similarly to the results of the first experiment, the ERN was significantly increased in the evaluative condition relative to the control condition. Thus, in the present study, combining monetary incentives and social evaluation, we expected the ERN/ΔERN to be increased in the high pressure condition as compare to the low pressure condition.

2.1.5. Numerical Stroop Paradigm

In this study, ERPs were measured using a numerical Stroop paradigm in which participants are required to compare two digits either according to their numerical magnitude or according to their physical size while ignoring the respective other dimension (Besner & Coltheart, 1979; Henik & Tzelgov, 1982; Kaufmann et al., 2005). The physical task (i.e., selecting the physically larger digit) is considered to provide a measure of the automatic processing of numerical magnitude (Bugden and Ansari 2011). More specifically, when the physically larger digit has to be selected, the numerical information of the pair of digits is irrelevant to the task. Nevertheless, participants typically show slower response times when stimuli are incongruent (i.e., the physically larger digit is numerically smaller) than when stimuli are congruent (i.e., the physically larger digit is also numerically larger) or neutral (both digits are the same number but differ in physical seize) (e.g., Kaufmann et al. 2005). Thus, the numerical information is processed automatically either facilitating (congruent trials) or inhibiting (incongruent trials) the response to the physical size. The numerical task (i.e., selecting the numerically larger digit), instead, is a number comparison task with conflicting physical size information. Here, the numerical magnitude information needs to be processed explicitly in order to select the numerically larger digit. Similar to the physical task, the irrelevant dimension, physical size, influences stimulus processing and congruency effects occur. Combining physical and numerical task, the numerical Stroop paradigm (as used in the present study) provides an elegant way to investigate automatic as well as explicit magnitude comparison.

Importantly, the ability to compare numbers has been suggested to be an important predictor of general mathematical achievement (Holloway & Ansari, 2009; Vogel, Remark, & Ansari, 2015). In one study, for instance, children's magnitude comparison ability was predicting mathematics achievement assessed one year later. This effect hold true when controlling for age, intellectual ability, and speed of number identification (De Smedt, Verschaffel, & Ghesquière, 2009). Furthermore, children with atypical development of numerical abilities, such as developmental dyscalculia, have been shown to exhibit impaired number comparison (Landerl, Bevan, & Butterworth, 2004; Rousselle & Noël, 2007). Taken

together, the numerical Stroop paradigm taps into magnitude comparison processes which are considered as an essential step in the development of mathematical competencies.

In addition, in a speeded version, the numerical Stroop paradigm is well suited for measuring the ERN (respectively ΔERN) which requires a large number of repetitions in order to obtain a sufficient number of error trials (see also Suárez-Pellicioni et al. 2013).

Finally, the interference between physical and numerical information in comparing two numbers requires attentional control to respond accurately (Posner & Petersen, 1990). Interestingly, attentional control has been suggested to be impaired by anxiety, including individual differences in anxiety as well as anxiety-inducing contextual factors (for a review, see Eysenck et al. 2007).

Regarding the behavioral performance, we expected that the performance in the numerical Stroop is impaired in the high pressure condition relative to the low pressure condition. Furthermore, we predicted that the task performance is linearly related to the individual level of test anxiety. Since the numerical task used in the present study is relatively easy compared to the complex arithmetic problems used by behavioral studies on choking under pressure in mathematics (e.g., Beilock et al. 2004) we expected response times to be a more sensitive indicator of task impairments than accuracy rates.

2.1.6. Hypotheses

In summary, we hypothesized that the ERN, as an indicator of error processing, and/or the ΔERN, as an indicator of general response monitoring, are increased by situational performance pressure. Moreover, we expected electrophysiological indices to be linearly related to the individual level of test anxiety. At the behavioral level, we expected, firstly, that response times are higher in the high pressure condition relative to the low pressure condition. Secondly, we predicted response times to linearly increase with the individual level of test anxiety.

2.2. Methods

2.2.1. Participants

Twenty-one female psychology students participated in the present study.[2] All participants were right-handed and had normal or corrected-to-normal vision. Two data sets had to be excluded from further analyses due to a technical malfunction during the EEG recording and one data set because of near-perfect task performance (inclusion criterion: minimum of five errors per pressure condition, see Olvet & Hajcak, 2009a, 2009b). All participants met the inclusion criteria of a maximum of 20 % errors per pressure condition. Thus, the remaining sample comprised 18 participants, aged between 18 and 32 years ($M =$ 21.06, $SD =$ 3.45 years). Participants gave written informed consent and received course credit for their participation. The study was approved by the ethics committee of the psychology department at the University of Göttingen.
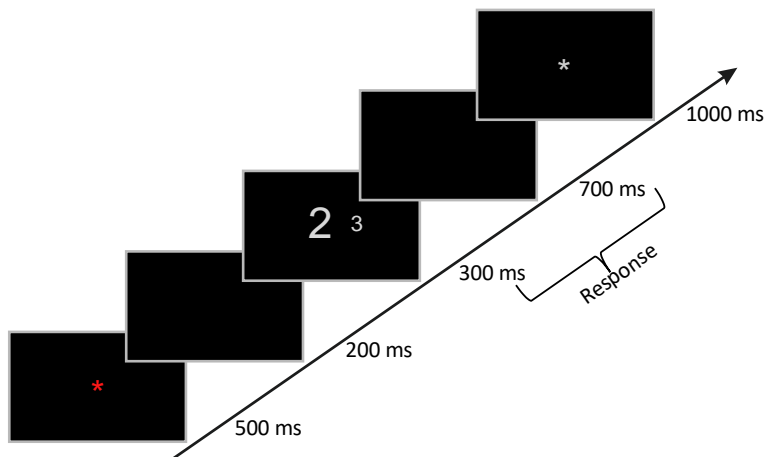
Figure 2. Schematic display of an EEG trial with incongruent number pair.

2.2.2. Apparatus and Materials

2.2.2.1. Numerical Stroop task

In the numerical Stroop task, a pair of Arabic digits was presented to the participants in the middle of the computer screen using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA, USA), and subjects had to indicate the larger of the two numbers. The digit pairs were created by combining the cardinal numbers without the boundary numbers one and

---

[2] Only female participants were recruited since gender differences have been reported for both mathematics (Else-Quest et al. 2010) and test anxiety (Hembree 1988). To avoid possible confutations with stereotype threat effects (Nguyen and Ryan 2008) all experiments were conducted by a team of a female and a male examiner.

nine while keeping a numerical distance of one. This results in six pairs of digits (2–3, 3–4, 4–5, 5–6, 6–7, 7–8), which were presented in both lateral configurations (e.g., 2–3, 3–2). The single digits of the number pairs were displayed in different font sizes (small: 40, medium: 60, large: 80) creating an interference between the numerical and the physical size of the digits. In the congruent condition, the numerically larger number was also displayed in larger font size (e.g., 2 **3**), while in the incongruent condition numerical magnitude and font size did not match (e.g., **2** 3). Participants' task consisted of either selecting the numerically larger (numerical task) or the physically larger number (physical task). Thus, in the numerical task, participants had to respond to the numerical magnitude while ignoring the physical size of the digits. In the physical task, instead, participants had to respond to the physical size of the digits ignoring the numerical magnitude. The neutral stimulus condition, in which one dimension should not influence stimulus processing, differed between the two task conditions. For the numerical task, both digits of a pair were displayed in medium font size (e.g., 2 3), while in the physical task condition the same digit was displayed in different font sizes (e.g., **2** 2). Task condition was alternated in blocks and participants were instructed at the beginning of each block to which dimension of the stimuli they were required to respond.

The temporal sequence of an EEG trial is depicted in Figure 2. Each trial started with the presentation of a red fixation asterisk for 500 ms, followed by a 200 ms blank screen. Then, the target stimulus (i.e., digit pair) was presented for 300 ms, and response recording started with stimulus onset. Subsequently, the screen turned black for an interval of 700 ms in which response recording continued. Trials were interleaved by a 1000 ms inter-trial interval displaying a grey asterisk. In order to minimize eye artefact, participants were asked to avoid eye blinks unless the asterisk turned grey. Participants selected the left digit by pressing the left button of a button box with the index finger and the right digit by pressing the right button with the middle finger.

2.2.2.2. German Test Anxiety Inventory (PAF)

Test anxiety was assessed using the German Test Anxiety Inventory (Prüfungsangstfragebogen, PAF) (Hodapp et al., 2011), which is originally based on the test anxiety inventory by Spielberger (1980). The questionnaire uses a 4-point Likert scale and consists of 20 items asking for the feelings and thoughts experienced in test situations (e.g., During tests, I find myself thinking about the consequences of failing). The resulting overall score ranges between a minimum of 20, indicating very low test anxiety, and a maximum of 80, indicating very high test anxiety.

The PAF is designed to assess test anxiety as a situation- specific trait anxiety, which is stable over time. This assumption has been confirmed by previous studies using latent-trait analysis (Keith et al., 2003). Thus, in the present study, it is assumed that participants' test anxiety score were stable between the time of administrating the test anxiety questionnaire and the previous two EEG sessions.

2.2.2.3. EEG data acquisition

EEG was acquired through a BioSemi Active Two system (BioSemi, Amsterdam, The Netherlands) from 64 scalp electrodes placed according to the extended 10–20 system. An electrooculogram (EOG) was recorded from three additional electrodes; two placed horizontally at the outer canthi of both eyes, and one placed above the nasion between the inner canthi of the eyes. EEG and EOG signals were sampled at 512 Hz and filtered between DC and 128 Hz.

2.2.3. Procedure

All participants were tested in two separate EEG sessions: a high pressure condition and a low pressure control condition. Half of the participants were randomly assigned to start with the high pressure condition, the other half to start with the low pressure condition. Between the two EEG sessions was an interval of 2 weeks. In each condition, participants performed a numerical Stroop task. In a third session, which was scheduled on a separate day after completing both EEG sessions, participants filled out the test anxiety inventory and answered demographic questions.

In the high pressure condition, a test score was computed for each participant based on speed and accuracy in the numerical Stroop task. The student with the highest score was rewarded with a 30€ gift card (outcome pressure). In addition, participants were monitored by the experimenters while doing the task through a web cam in the EEG cabin (monitoring pressure). Both measures were explained to the participants at the beginning of the study. The camera was mounted before recording started in front of the participants and the experimenter reiterated that their performance would be evaluated. To demonstrate the functionality of the monitoring, the video recorded by the webcam was transferred onto participants' monitor for about a minute before the numerical Stroop task started. After completion of the series of all experiments, participants were debriefed, and the student with the best test score received the gift card.[3]

---

[3] The high pressure scenario used in the present study differed in some points from the scenario reported by Beilock and colleagues (e.g. Beilock et al. 2004). Since the participating psychology

In the low pressure condition, participants were instructed to respond as quickly and as accurately as possible, as common in psychological testing. No camera was recording the participants during the session and participants were not rewarded for their performance.

Except for the pressure manipulation, the experimental procedure of both EEG sessions was identical. Upon arrival, participants were seated in an EEG cabin, and EEG electrodes were placed. The recording started with a 3 min sequence in which participants were asked to deliberately execute eye movements. This data were used later on in the analysis to remove eye movement artifact from the data (see data analysis). Participants received task instructions on screen, and then completed a block of 14 practice trials for each task condition (numerical, physical) in which feedback was provided (viz., "correct", "wrong", "too slow"). Then, the experiment started and stimuli were presented in blocks alternating between task conditions. At the beginning of each block, participants were instructed whether they had to select the numerically or the physically larger number. To keep the error rate in the range of 5–25 %, participants received manipulated feedback at the end of each block (see Gehring, Liu, Orr, & Carp, 2012). If less than 5 % errors were committed in a block, participants were asked to respond quicker; and if more than 25 % errors were committed, participants were asked to slow down. In between that range, participants were instructed to continue as before. This "feedback" was provided at the beginning after 36 trials (short block) and then after 72 trials (long block). In total, 936 number pairs were presented in six short and ten long blocks. To avoid participants losing their concentration a short break was implemented after half of the trials. Each EEG test session took about 1.5 h, including task instructions as well as mounting and de-mounting of electrodes.

2.2.4. Data Analysis

2.2.4.1. Linear mixed models

A linear mixed model approach was employed to test whether performance pressure and test anxiety modulated task performance as well as electrophysiological indices of response monitoring. Linear mixed models predict an outcome variable based on a combination of discrete or continuous variables while allowing to specify random terms for repeated measurements (for a review, see, Kliegl et al. 2011). In the present study, this allowed to predict the outcome variables (i.e., response time, accuracy, ΔERN) based on the discrete variable pressure (repeated measures) and the continuous variable anxiety. Linear mixed models with random intercepts were set up using the lme4 package (Bates, Maechler, Bolker,

---

students in the present study were likely to be familiar with psychological experiments it seemed implausible to us to involve a cover story with confederates.

& Walker, 2014) in the R environment (R Core Team, 2015). To test whether the factor pressure and anxiety as well as the interaction of both significantly affected the outcome variables, we compared a model containing the factor of interest against a restricted model using likelihood ratio tests. In addition to the $\chi 2$ statistics of the model comparison, the Akaike information criterion (AIC) was computed for each model. Parameters' estimate and standard error (SE) for the model with the highest goodness of fit are reported.

2.2.4.2. Behavioral Data

The overall score of test anxiety was calculated for each participant by adding up all single item scores. For the linear mixed model analysis, raw values were z-transformed to improve the interpretability of the estimates (see Schielzeth 2010).

Response time and accuracy were aggregated individually for each condition, i.e., pressure (low, high), task (numerical, physical), and congruency (congruent, neutral, incongruent). Missed responses were treated as errors for calculating the accuracy rate and were excluded from the mean response time and the ERP analyses. To test whether performance pressure affected task performance, a 2 (pressure) × 2 (task) × 3 (congruency) repeated measurement ANOVA was conducted with response time and accuracy as dependent variable. To further analyze the interaction of performance pressure with individual test anxiety, a linear mixed model was setup for each dependent variable based on the result of the ANOVA. This restricted model was then compared against a model including a term for test anxiety and a model including the interaction of performance pressure and test anxiety.

2.2.4.3. Electrophysiological Data

EEG data were analyzed using EEGLAB 13.2.1 (Delorme & Makeig, 2004) and ERPLAB 4.0 (Lopez-Calderon & Luck, 2014) in the MATLAB environment (Mathworks, Inc., Natick, MA, USA, release 2013a). The signal was rereferenced to the mean (common average reference) and bandpass filtered with cutoffs of .1 and 30 Hz (Olvet & Hajcak, 2009b). Eye movement artifacts were automatically corrected using regression coefficients which were estimated based on data of the eye movement sequence (Schlögl et al., 2007). Response-locked epochs from 400 ms before response onset to 600 ms after response onset were extracted with a 100 ms baseline [−200 −100]. To detect remaining artifacts, a moving window peak-to-peak threshold of 100 µV with a window size of 200 ms and a step size of 100 ms was applied to channel FCz. The valid trials were then averaged for error responses (obtaining the ERN) and correct responses (obtaining the CRN) for each participant and pressure condition. To quantify the ERN and CRN, the mean amplitude between 0 and 100 ms at electrode FCz

was extracted (see Olvet and Hajcak 2009b). Finally, the ΔERN was computed by distracting the mean CRN from the mean ERN for each participant and pressure condition (Riesel et al., 2013).

To test whether the mean amplitude of ERN and of the ΔERN were modulated by performance pressure as well as individual test anxiety, we compared nested linear mixed model containing the factor of interest against restricted models.

(a)
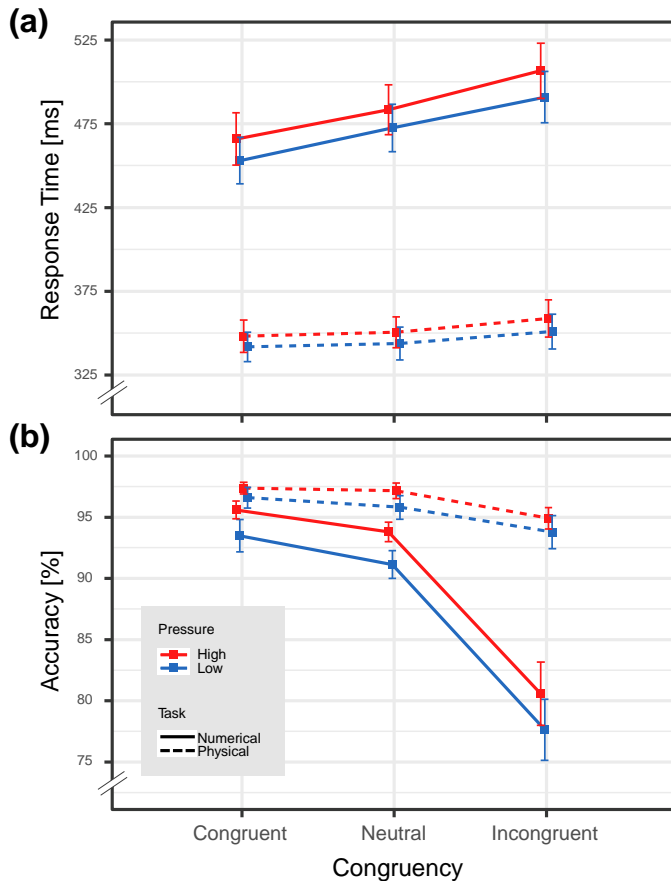
Figure 3. (a) Mean response time and (b) accuracy as a function of congruency (congruent, neutral, incongruent), task (numerical, physical) and pressure (low, high). Error bars indicate +/- 1 standard error.

(b)

2.3. Results

2.3.1. Behavioral Data

Test anxiety scores varied among participants between a minimum of 36 and a maximum of 61 (*M* = 47.5, *SD* = 7.18). Individual test anxiety scores are depicted in Figure 5. Compared to the norm sample (Hodapp et al., 2011), 15 participants showed average test anxiety and 3 participants (ID: 9, 13, 14) showed heighted levels of test anxiety (cut-off raw value: 56). None of the participants had a test score indicating strongly heightened test anxiety (cut-off raw value: 67).

Figure 3 shows the mean response time and accuracy as a function of pressure, task and congruency. Both, task condition, $F(1, 17) = 151.50$, $p < .001$, and stimulus congruency, $F(2, 34) = 57.98$, $p < .001$, significantly affected response times, as expected based on previous literature. Furthermore, there was a significant interaction between task and congruency, $F(2, 34) = 45.71$, $p < .001$. Confirming our hypothesis, response times were significantly increased when performance pressure was applied, $F(1,17) = 4.791$, $p < .001$. The effect of pressure did not interact with task condition nor stimulus congruency (all p values ≥ .252).

Similar to response times, stimulus congruency, $F(2, 34) = 41.46$, p < .001, and task condition, $F(1, 17) = 96.12$, $p < .001$, had a significant effect on accuracy as well as the interaction between these factors, $F(2, 34) = 38.23$, $p < .001$. Surprisingly, performance pressure significantly improved accuracy, $F(1, 17) = 5.23$, $p = .035$. No other interaction reached statistical significance (all p values ≥. 125). Hence, the restricted linear mixed model contained an interaction between stimulus congruency and task condition as well as a main effect for pressure (Congruency * Stroop + Pressure).[4] The model comparison revealed that test anxiety [$\chi^2(1) = .84$, p = .359, AIC: 1943.0 vs. 1944.2] did not significantly affected response times. However, the interaction between test anxiety and performance pressure turned out significantly [$\chi^2(1) = 7.22$, $p = .007$, AIC: 1939.0]. Compared to the low pressure condition, response times were decreased by 5.37 ms ± 1.98 (standard errors) for a gain of one standard score in test anxiety. Regarding the accuracy, the percentage of correct responses was neither significantly modulated by test anxiety [$\chi^2(1) = 2.48$, $p = .115$, AIC: 1301.6 vs. 1301.1] nor by the interaction of test anxiety and performance pressure [$\chi^2(1) = .18$, $p = .670$, AIC: 1302.9]. (For scatter plots of both output variables as a function of test anxiety, see Figure S 1) Thus, the hypothesis according to which response times linearly increase with the individual level of test anxiety could not be confirmed. Performance pressure did not interact with stimulus congruency or task condition in affecting task performance. Therefore, trials were averaged across congruency and task conditions in the EEG analysis to increase the number of errors and thus the signal to noise ratio.

---

[4] In the notation of the linear mixed models used in the present manuscript, "+" indicates a main effect and "*" indicates an interaction term.
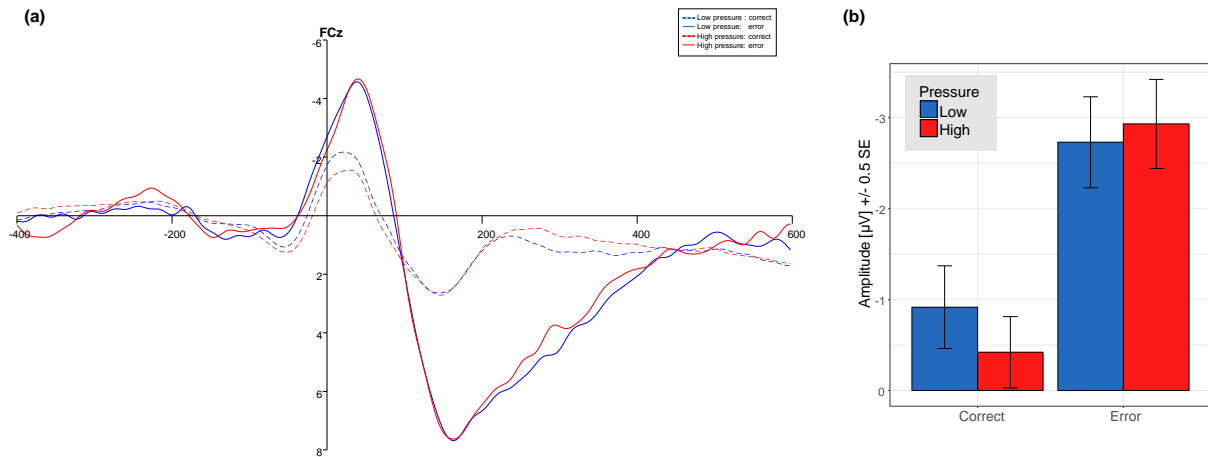
Figure 4. (A) Average waveforms of correct and erroneous response for low pressure and high pressure condition. (B) Mean amplitude in an interval of 100 ms after response onset for response type and pressure condition.

2.3.2. Electrophysiological Data

Figure 4a shows the average waveforms of correct and erroneous responses in the low pressure and the high pressure condition. The mean amplitude of the ERN and the CRN (0–100 ms relative to response onset) are depicted in Figure 4b, indicating a subtle difference in the ERN and a more pronounced difference in the CRN between pressure conditions. The following linear mixed model analysis was based on the mean amplitude of the ERN and the difference between ERN minus CRN (ΔERN). Results of model comparison predicting the ERN and ΔERN are listed in Table 2. The ERN was neither significantly modulated by performance pressure, individual test anxiety nor the interaction of both factors. However, performance pressure significantly affected the ΔERN, increasing it by about .70 µV ± .28 (standard errors). Furthermore, test anxiety significantly modulated the amplitude difference between CRN and ERN. In detail, the ΔERN was increased by .83 µV ± .34 for a gain of one standard score in test anxiety. Including an interaction term between performance pressure and test anxiety did not further improve the goodness of fit of the model. Figure 5 depicts the ΔERN as a function of test anxiety for both pressure conditions. Taken together, electrophysiological data indicate that response monitoring is modulated by performance pressure and test anxiety independently.

Table 2. Result summary for nested mixed linear models predicting the output variable (OV) ERN and ΔERN. Models were compared against each other using maximum likelihood tests providing χ² values, degrees of freedom between models df χ², and corresponding p values. Akaike information criterion (AIC) are listed as an indicator of goodness of fit.

OV	Predictor	AIC	χ²	df χ²	p
ERN	Intercept	345.22			
	Pressure	347.10	0.12	1	.729
	Pressure + Test anxiety	349.09	0.00	1	.944
	Pressure * Test anxiety	351.00	0.09	1	.759
ΔERN	Intercept	137.43			
	Pressure	133.96	5.47	1	.019
	Pressure + Test anxiety	130.96	5.00	1	.025
	Pressure * Test anxiety	132.86	0.10	1	.753

Note. ERN = Error-related negativity.
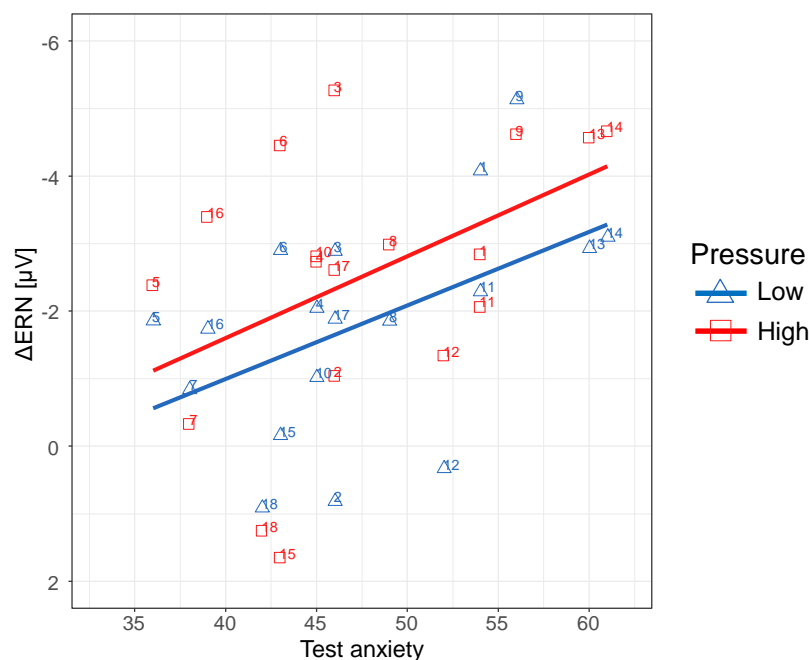
Figure 5. Scatter plot of ΔERN as a function of performance pressure and test anxiety (raw values). Numbers indicate individual subjects. Linear regression lines were fitted for each pressure condition for illustrative purposes.

2.4. Discussion

The aim of the present study was to shed light on how responses are monitored in high-stake tests. Moreover, we were interested to see whether altered response monitoring is related to choking under pressure. The study extends previous research on response monitoring under pressure in two ways. Firstly, we conjointly analyzed the influence of situational performance pressure and individual test anxiety on task performance as well as electrophysiological indices of response monitoring. Previous studies, instead, have either

focused on the question of how individual differences in anxiety or how situational factors affect response monitoring. Thus, to the best of our knowledge, the present study is the first to combine both individual and situational factors in investigating response monitoring. Secondly, the present study was designed in a more applied and educational setting and hence has a better ecological validity compared to previous studies. More specifically, (a) in the present study, a real-life test situation was modeled by combining outcome and monitoring pressure making a connection to the literature on choking under pressure (see DeCaro et al. 2011). Similarly, (b) we specifically addressed the question of how test anxiety affects response monitoring and behavioral performance measurements. Compared to general trait anxiety or obsessive-compulsive disorder, as investigated by the majority of previous studies addressing the relationship between the ERN and anxiety (see Moser et al. 2013), test anxiety is highly relevant to the phenomenon of choking under pressure. (c) Finally, unlike previous studies, we used a numerical task in the present study (cf. Suárez-Pellicioni et al. 2013) which is related to mathematical achievement and competencies. Taken together, the present results are highly relevant to educational settings in general and mathematics education in particular.

Regarding the ERP analyses, we hypothesized that the ERN and/or the ΔERN are increased by both performance pressure and test anxiety. At the behavioral level, we expected, firstly, that response times are higher in the high pressure condition relative to the low pressure condition. Secondly, we predicted response times to linearly increase with the individual level of test anxiety.

Electrophysiological data indicated that response monitoring is changed by performance pressure. In general, this finding confirms previous studies according to which the ERN is modulated by performance pressure (Ganushchak & Schiller, 2008; Hajcak, Moser, et al., 2005; Kim et al., 2005). Unlike previous studies, we only found a significant effect on the difference potential between CRN and ERN (ΔERN), but not on the ERN alone. While the ERN is thought to reflect activity specific to error processing, the ΔERN has been suggested to reflect activity more broadly related to response monitoring (Riesel et al., 2013). However, we are cautions in drawing any firm conclusions from the absence of a significant effect regarding the ERN. As depicted in Figure 4a, the absolute amplitude of the ERN differed between pressure conditions in the predicted direction. A possible explanation of why this effect did not turn out to be significant is that a different experimental procedure was used in the present study in contrast to the study by Hajack and colleagues (2005), which has reported a significant effect of social evaluation on both ΔERN and ERN. More specifically, while Hajack et al. (2005) manipulated social evaluation stress within a single EEG session, we manipulated performance pressure between two separated EEG sessions. The test-retest reliability of the ERN over a period of 2 weeks has been reported to be excellent (Olvet &

Hajcak, 2009a). Nevertheless, recording data in two separated sessions is likely to increase the variability in the data. Thus, even though the sample sizes of both studies were identical, the difference of ERN amplitude between pressure conditions might have failed to reach statistical significance in the present study. The ΔERN, however, was significantly increased by performance pressure. As depicted in Figure 4a and b, this effect resulted from both a subtle difference in the ERN and a more pronounced difference in the CRN between pressure conditions. The exact interplay of the ERN and the CRN is a subject of ongoing research. Both components have been suggested to be mutually dependent and to reflect the same cognitive control mechanism (see Simons 2010). For instance, the CRN has been shown to be less negative on trials preceding errors relative to trials preceding correct responses (Hajcak, Nieuwenhuis, Ridderinkhof, & Simons, 2005). Accordingly, we interpreted the modulation of the ΔERN in the present study as indicating a general change of response monitoring. Thus, the present results provide further evidence that response monitoring is changed in high-stake tests.

The analysis of the ERPs further revealed that the ΔERN was linearly related to individual differences in test anxiety. This suggests that high test-anxious individuals differ in response monitoring compared to low test-anxious individuals. To the best of our knowledge, this provides first evidence that test-anxious individuals exhibit altered response monitoring. In general, this finding fits into the emerging picture according to which the CRN as well as the ERN are modulated by various forms of anxiety (see Moser et al. 2013). However, in the present analysis, the interaction between test anxiety and performance pressure did not further improve the goodness of fit of the linear mixed model. Thus, performance pressure and test anxiety modulated ΔERN independently. A possible explanation of why both factors did not interact with each other might be that both pressure conditions were perceived as an evaluative situation to some degree by participants. Indeed, the frequent and manipulated feedback typically used in ERN paradigms might be perceived as pressuring, especially for high test-anxious individuals. A way to overcome this limitation would be to label the low pressure condition explicitly as non-diagnostic.

Despite the difference in electrophysiological measures, behavioral data provided mixed evidence in regard of our hypotheses. As expected, response times were higher in the high pressure condition than in the low pressure condition. However, this effect was subtle having a range of 10 to 20 ms, depending on stimulus congruency and task condition. Accuracy, instead, was notably increased by 1-3 % when pressure was applied. Contrary to our expectations, response times significantly decreased with the individual level of test anxiety, but only in the high pressure condition. Accuracy, instead, was neither affected by test anxiety nor by the interaction between test anxiety and performance pressure.

Considering the effect of performance pressure, the pattern of behavioral results rather indicates that participants emphasized accuracy at the cost of response time in the high pressure condition. Thus, regarding the question of whether altered response monitoring in high-stake tests contribute to task impairments due to performance pressure, the present study remains inconclusive. A possible explanation for the lack of pressure-related task impairments might be that the numerical Stroop task used in the present study was too easy for participants. According to the distraction account, task impairments result when the combined working memory demands of task processing and extraneous processing (i.e. worries about the test) are exceeding the individual working memory capacity. However, if the demands stay within the limit of the working memory capacity, task performance remains unaffected. The working memory load imposed by the present task is arguably lower than the one imposed by the complex arithmetic tasks used by behavioral studies which have reported substantial task impairments under performance pressure (Beilock & Carr, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004). However, this presents a major challenge of investigating the ERN in regard of the phenomenon of choking under pressure in general. Paradigms which have been used to measure the ERN, including Flanker, Stroop, and Go/NoGo are similar to the numerical Stroop task regarding their working memory demands. However, complex mathematical problems, as used by behavioral studies on choking, are not well suited to measure the CRN/ERN for two reasons. Firstly, it is unclear whether the ERN can be measured in a task which is not in a two-alternative forced choice format (Gehring et al. 2012). Secondly, a number of at least 6 error trials per condition is needed to obtain a relatively stable ERN (Olvet & Hajcak, 2009b). But, at the same time, an error has to be a relatively unlikely event within the task and therefore error rates should stay moderately low (see Gehring et al. 2012). Thus, a large number of trials is needed to measure the ERN reliably. However, complex mathematical tasks usually require a certain processing time which makes it hardly possible to record a sufficiently large number of trials in a single EEG session. A possible alternative to the approach of testing response monitoring and task impairments within the same paradigm might be to measure each variable in a separate experiment. More specifically, electrophysiological indices of response monitoring could be measured in a single EEG session under social evaluation stress using a paradigm which meets the beforehand mentioned requirements for measuring the ERN (e.g., a numerical stoop task). In a second experiment, the same participants would be tested behaviorally in two pressure scenarios using complex mathematical tasks having high working memory demands (e.g. Gauss's modular arithmetic). Correlating the ERN/ΔERN measures, obtained in the EEG experiment, with the task performance under performance pressure, obtained in the behavioral experiment, might allow investigating how response monitoring is related to the phenomenon of choking under pressure.

Regarding the effect of individual differences in test anxiety, results suggest that, within the range of test anxiety scores included in the present study, higher test anxiety improves response times when academic pressure is applied. One explanation for this unexpected result might be that test-anxious students employ additional resources or compensatory effort to maintain task performance in an evaluative situation (Eysenck et al., 2007). If the task is relatively easy, as in the present study, this might even lead to improved performance of test-anxious students.

A limitation of the present study is that only a relatively small group of participants was tested. Compared to the norm sample, 15 participants showed average test anxiety and 3 showed above-average test anxiety. However, none of the participants had a test score indicating strongly heightened test anxiety. Thus, participants' test anxiety scores did not cover the full range of test anxiety, limiting the generalizability of the present results. Therefore, it would be important to test students with strongly heightened test anxiety in order to further extend our understanding of how test anxiety affects task performance as well as response monitoring. For this, participants could be first screened for test anxiety and then data could be collected from a sample covering the full range of test anxiety or of two extreme groups (low vs. high test anxiety). Given that the present study found a linear relationship between test anxiety and the ΔERN, the former approach might be a promising way.

To conclude, the present study provides further evidence that response monitoring is altered in high-stake tests using a numerical task. Furthermore, it provides first evidence that high test-anxious individuals exhibit altered response monitoring. To determine how response monitoring is related to the phenomenon of choking under pressure further studies are needed. Performance tests are commonly used in schools and universities to assess mathematical ability and thus play an important role in shaping the career of individuals. Therefore, a better understanding of the social and affective conditions for students to achieve their full potential in mathematics tests is of great importance.

3. Study 2 – Response Monitoring in High Test-anxious Students

3.1. Introduction

Results of the previous study suggest that the response monitoring is modulated by both situational pressure and individual difference in test anxiety (Schillinger, De Smedt, & Grabner, 2016). More specifically, the ΔERN (i.e., the difference score between CRN and ERN) was enhanced under performance pressure as compared to the low pressure control condition and linearly increased with individual test anxiety in both conditions. However, there was no significant interaction between performance pressure and test anxiety in modulating the ΔERN. On a behavioral level, the study provided mixed evidence with slower response times but increased accuracies in the high pressure condition as compared to the low pressure condition. Contrary to predictions, students with higher test anxiety responded significantly faster under performance pressure than students with lower test anxiety.

The lack of a significant interaction between performance pressure and test anxiety in Study 1 seems to be at variance with the conceptualization of test anxiety as a situation-specific personality trait (Spielberger et al., 1976; Zeidner, 1998). According to this concept, test anxiety is specifically triggered by evaluative situations. Therefore, the detrimental effect of test anxiety on cognitive processes should be more pronounced under performance pressure than in a neural condition. One reason for the absence of such an interaction effect in Study 1 could have been the limited range of test anxiety within the study sample. Of the 18 data sets included in the final sample, 15 participants had average test anxiety and three participants had high levels of test anxiety. None of the tested participants had very high levels of test anxiety and none had low or very low levels of test anxiety. Thus, the difference between students with lower and higher test anxiety was relatively small in the previous study. This limited variance of test anxiety scores is likely to make the difference in the response to the pressure manipulation more subtle and consequently less pronounced in the EEG indices. Furthermore, we did not control for differences in general anxiety in the previous study. General anxiety has been consistently shown to be moderately correlated with test anxiety (for a meta-analysis, see Hembree, 1988). Unlike test anxiety, general anxiety is not thought to be specific to a situation (Spielberger, 1972) and can be therefore expected to affect cognitive processes also in tasks with neutral instructions. In fact, the relatively test-anxious students in Study 1 are likely to have been also more anxious in general. This would have further diluted the effect of performance pressure on the ERN/ΔERN. Therefore, the present study aimed to compare the response monitoring of low and high test-anxious students while minimizing differences in general anxiety.

Moreover, the distraction model assumes that worries are the central mechanisms by which test anxiety impairs working memory processes (Eysenck et al., 2007; Wine, 1971). In an evaluation situation, test-anxious students are predicted to worry about their performance and the consequences of failing. In turn, such worries have been suggested to enhance the amplitude of the ERN (for a review, see Moser et al., 2013). The results of Study 1 are in line with this suggestion given that the ΔERN was enhanced by test anxiety and performance pressure, both of which are likely to increase worries. However, the link between the ΔERN and worries remained speculative given that worries were not directly assessed during the task. In the present study, we therefore aimed to directly relate worries in the test situation to the EEG indices of response monitoring.

3.1.1. Present Study

The present study aimed to extend the first study by (a) comparing the response monitoring of low and high test-anxious students while minimizing differences in general anxiety and (b) directly relating the EEG indices of response monitoring to worries reported under performance pressure. To this end, 341 university students were pre-screened for their test anxiety and general anxiety as part of a larger test battery (see Schillinger et al., 2018). The selection procedure aimed to form a high test anxiety (HTA) and a low test anxiety (LTA) group while minimizing differences in general anxiety between groups. Both groups were tested under performance pressure as well as in a low pressure control condition performing a numerical Stroop. We reasoned that the comparison of the two extreme groups would be more sensitive in revealing the predicted interaction between test anxiety and performance pressure in modulating the ERN/ΔERN. To directly relate EEG indices of response monitoring to worries, participants were asked to rate their performance-related worries at regular intervals throughout the experiment using a Faces Anxiety Scale (FAS) (Bieri, Reeve, Champion, Addicoat, & Ziegler, 1990; Punaro & Reeve, 2012; Trezise & Reeve, 2014b).

In addition, the following changes were implemented to the methods of the present study. Firstly, the results of the previous study revealed that the numerical task of the numerical Stroop (i.e., selecting the numerically larger number) was significantly more difficult than the physical task (i.e., selecting the physically larger number). To increase the number of errors and hence the signal-to-noise ratio we focused on the numerical task in the present study. Secondly, both pressure conditions were recorded within a single EEG session in the present study to reduce the variability between conditions. This was done because the previous study failed to find a significant effect on the ERN in isolation, as reported by Hajcak and colleagues (2005). While Hajcak and colleagues (2005) recorded both a low pressure and a high pressure condition within a single EEG session, the previous study recorded the conditions in two

separate sessions. This might have increased the variability in the data, for instance, by differences in the electrode mounting between sessions. Thirdly, a manipulation check was included to check whether the pressure condition was actually perceived as being more stressful than the low pressure condition. For this, we adapted a procedure reported by Beilock and colleagues (2004) for the present study. In their study, participants were randomly assigned to perform a complex arithmetic task either in a high pressure condition or in low pressure control condition. Similar to the present study, performance pressure was manipulated by instructing participants that they receive a performance-based reward and would be filmed during the task. To check the effectiveness of this pressure manipulation, Beilock and colleagues (2004) assessed the state anxiety after each condition and asked participants to rate the importance, pressure, and success regarding the task. Finally, the results of the previous study suggest that test anxiety affects both the CRN and the ERN. More specifically, the main effect of performance pressure on the ΔERN resulted from two reverse effects on the CRN and the ERN respectively. While the ERN appeared to be enhanced by performances pressure, the CRN was rather decreased. These effects were added up in computing the difference scores between CRN and ERN (i.e., ΔERN). To disentangle the effect of performance pressure and test anxiety on both the CRN and the ERN we included both components as a factor within an analysis of variance (ANOVA).

3.1.2. Hypotheses

Regarding the manipulation check, we expected to find a similar pattern of results as reported by Beilock and colleagues (2004). More specifically, participants were predicted to report more performance pressure, lower success, and higher state anxiety in the high pressure condition than in the low pressure condition. Instead, the rating of the importance of the task was expected to be similar between pressure conditions. Text anxiety was included as an additional factor in the analysis of the manipulation check to explore whether the two groups differed in their perception of the pressure conditions.

Given that Study 1 as well as previous research on the ERN (for a meta-analysis, see Moser et al., 2013) provided mixed evidence regarding performance impairments of anxious individuals, we did not formulate specific hypotheses regarding differences in response time or accuracy between conditions in the present study.

Performance pressure and test anxiety were reasoned to interact in modulating worries during the task. While the HTA group was predicted to report higher worries under performance pressure than the LTA group, no difference between groups was predicted in the low pressure condition.

On an electrophysiological level, we hypothesized that EEG indices of response monitoring are modulated by both performance pressure and test anxiety. Based on previous research (for a meta-analysis, see Moser et al., 2013), we expected to see the effects of test anxiety and performance pressure primarily on the ERN and, to a lesser degree, on the CRN. Both LTA and HTA students were expected to display an enhanced ERN under performance pressure. Moreover, we predicted that the HTA group would exhibit a higher ERN than the LTA group in the high pressure condition but not in the low pressure condition. Finally, the ERN amplitude was predicted to be linearly related to the worry reported by LTA and HTA students in both pressure conditions.

3.2. Methods

3.2.1. Participants

Participants were recruited from a pool of 341 university students who had been pre-screened for test anxiety and general anxiety as part of a larger test battery (for a detailed description of the subject pool, see Schillinger, Vogel, Diedrich, & Grabner, 2018). The selection procedure aimed to form a high test anxiety (HTA) and a low test anxiety (LTA) group while minimizing differences in general anxiety between groups. For each group, 24 participants were tested, but two data sets (both HTA group) had to be excluded due to low EEG data quality (see section EEG analysis), one data set due to technical problems during the recording (HTA group), and one data set due to a participant aborting the experiment because of nausea (LTA group). Thus, the final sample was comprised of 21 participants in the HTA group and 23 participants in the LTA group. All participants were right-handed with normal or corrected-to-normal vision and no history of psychological or neurological diseases. They gave written informed consent regarding the EEG recording but were not informed until the end of the experiment about the true purpose of the study. Fields of study of the participants included psychology ($N_{LTA} = 12$, $N_{HTA} = 10$), humanities ($N_{LTA} = 4$, $N_{HTA} = 5$), and science ($N_{LTA} = 7$, $N_{HTA} = 6$). Participants either received €12 ($N_{LTA} = 18$, $N_{HTA} = 17$) or 2 course credits ($N_{LTA} = 5$, $N_{HTA} = 4$) as compensation for participating in the study. The ethics committee of the University of Graz approved the study.

Table 3 summarizes the differences between the groups in demographics and anxiety measures. The HTA group (score range: 51 to 70) had significantly higher levels of test anxiety than the LTA group (score range: 21 to 39). However, the two groups also differed significantly regarding their general anxiety with the HTA group (score range: 32 to 60) exhibiting higher levels of general anxiety than the LTA group (score range: 26 to 48). Importantly, the effect size of the difference in test anxiety (Cohen's $d = -5.41$) was considerably larger than those of the difference in general anxiety (Cohen's $d = -1.52$). In the LTA group, 16 participants were

classified as having low or very low test anxiety and six participants as having average test anxiety (see Table 3). In the HTA group, 17 participants were classified as having high or very high test anxiety and four participants as having average test anxiety. Note that there is a small overlap of participants with average test anxiety in both groups. However, the difference between the highest sum score in the LTA group and the lowest sum score in the HTA group was still more than 10. A list of all anxiety scores and the group assignment can be found in Table S 1. The proportion of female and male participants was counterbalanced between groups and the two groups did not differ significantly in age (see Table 3).

Table 3. Differences between the low test anxiety (LTA) and high test anxiety (HTA) group in demographics and anxiety measures. Norm-based classification in five categories ranging from "very low anxiety" to "very high anxiety".

	LTA	HTA	Test statistic
Age M (SD)	21.35 (3.19)	20.10 (1.73)	$t(42) = 1.60$, $p = .117$, $d = 0.48$
Females/Males	14/9	12/9	$\chi^2(1) = <0.01$, $p = .999$, $\varphi = -0.38$
Test anxiety			
M (SD)	32.26 (4.37)	57.52 (4.95)	$t(42) = -17.99$, $p < .001$, $d = -5.43$
Classification	5/11/7/0/0	0/0/4/15/2	
General anxiety			
M (SD)	34.17 (5.65)	45.62 (9.06)	$t(42) = -5.08$, $p < .001$, $d = -1.53$
Classification	0/0/21/2/0	0/0/14/3/4	

Note. LTA = low test anxiety, HTA = high test anxiety, M = mean, SD = standard deviation. Classification of anxiety sum scores as "very low" / "low" / "average" / "high" / "very high"; d = Cohen's d, φ = Phi effect size.

3.2.2. Apparatus and Materials

3.2.2.1. Numerical Stroop

The numerical Stroop paradigm was adjusted based on the findings of Study 1 and presented with Psychopy 1.73.04 (Peirce, 2007). To increase the number of error trials, we focused on the numerical task (i.e., selecting the numerically larger number) in the present study. This task has been shown in Study 1 to be significantly more difficult than the physical task (i.e., selecting the physically larger number). In each pressure condition, participants performed 11 blocks á 36 trials of the numerical task, resulting in a total of 396 trials per condition. Three blocks of the physical task á 36 trials were included to enhance the salience of the irrelevant physical dimension (inserted after the 1st, 4th, and 9th block of the numerical task). These distractor blocks were discarded from the behavioral and electrophysiological analyses. Furthermore, a fix blank interval of 200 ms was implemented after a response was given to ensure that the recording of the ERN was not confounded by the onset of the ITI.

Finally, the duration of the ITI was varied between 500 and 1000 ms to avoid monotonous response patterns. All other parameters of the paradigm remained unchanged as compared to Study 1, including the used number pairs and the trial order. The temporal sequence of an EEG trial is depicted in Figure 6.
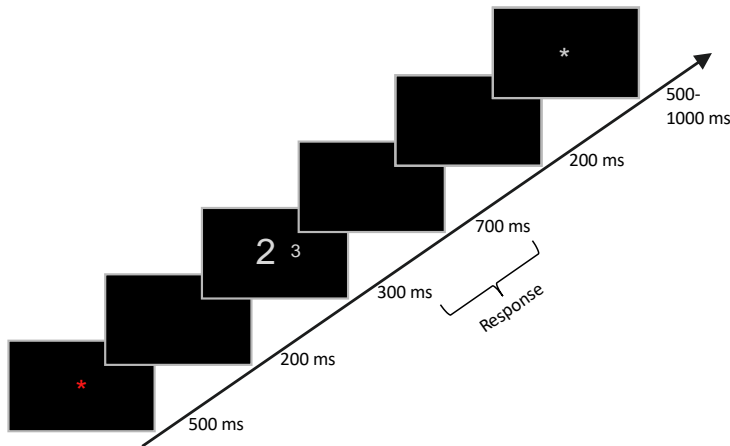
Figure 6. Schematic display of an EEG trial with incongruent number pair. The blank interval of 200 ms after stimulus presentation was only presented when a response was given.

3.2.2.2. German Test Anxiety Inventory (PAF)

Test anxiety was assessed using the German Test Anxiety Inventory (Prüfungsangstfragebogen, PAF) (Hodapp et al., 2011), which is described in more details in the method section of Study 1. In the present study, sum scores were classified as either "very low", "low", "average", "high", or "very high" based on the gender-specific norm sample of German university students (Hodapp et al., 2011).

3.2.2.3. State-Trait Anxiety Inventory (STAI)

The STAI is a well-known questionnaire used to assess anxiety including a state and a trait scale (Spielberger, 1980). While the state scale refers to the experience of anxiety in the particular moment of filling out the questionnaire, the trait scale refers to how frequent these feelings are experienced in general. Both scales consist of 20 items, which are rated on a 4-point Likert scale. The labels of the state scale range from "not at all" to "very much so" and the labels of the trait scale from "almost never" to "almost always". The German version of the STAI has been reported to be a reliable psychometric instrument as indicated by an excellent internal consistency for both the trait scale (Cronbach alpha = .90) and state scale (Cronbach alpha = .90) (Laux, Glanzmann, Schaffner, & Spielberger, 1981). A sum score was obtained

after reversing positively worded items (possible score range: 20-80). The sum score for the trait scale was classified as either "very low", "low", "average", "high", or "very high" based on the gender-specific norm sample (Laux et al., 1981).

3.2.2.4. Worry Scale

Worries were assessed using a Faces Anxiety Scale (FAS) displaying six facial expressions (see Figure 7). This scale is based on the Faces Pain Scale by Bieri and colleagues (1990) and has been previously used to assess worries in children related to literacy (Punaro & Reeve, 2012) and math performance (Punaro & Reeve, 2012; Trezise & Reeve, 2014a, 2014b). While the present study is, to our knowledge, the first to use this scale with adult participants in the context of learning, FASs have been previously reported for adults in medical settings (see Stuppy, 1998). The FAS was presented after each numerical task block (i.e., 11 times per condition) referring to participants' worries about their performance in the preceding block regarding both response time and accuracy. Participants were instructed to select the first face (neutral expression) if they were not worried at all and the sixth face (expression of discomfort) if they were very worried about their performance.
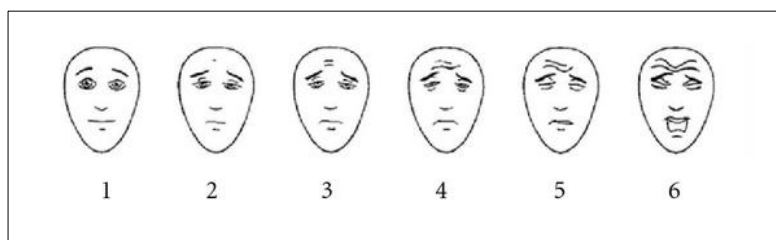
Figure 7. Faces Anxiety Scale (FAS) to assess performance-related worries. Participants were instructed to select the first face if they were not worried at all and the sixth face if they were very worried about their performance.

3.2.2.5. Manipulation Check

Participants were asked to answer a number of questions in order to check whether the high pressure condition was actually perceived as being more stressful than the low pressure control condition (see Beilock, Kulp, Holt, & Carr, 2004). For this, the state scale of the STAI was administered after each condition to assess the level of state anxiety. At the end of the experiment, participants were then asked to rate both pressure conditions regarding (a) how important they felt it was to perform at a high level, (b) how much pressure they felt to perform at a high level, (c) how highly they would rate their performance in the task. Items were rated on a 7-Point Likert scale with answers ranging from (a) "not important at all" to "very important", (b) "no pressure at all" to "very much pressure", (c) "very bad" to "very good". In addition, participants were asked to describe why it was important for them to perform at a high level in

either condition using an open answer format. The items of the manipulation check can be found in Supp. Material 1 (in German).

3.2.2.6. EEG data acquisition

EEG was acquired through a BioSemi Active Two system (BioSemi, Amsterdam, The Netherlands) from 64 scalp electrodes placed according to the extended 10–20 system. An electrooculogram (EOG) was recorded from three additional electrodes; two placed horizontally at the outer canthi of both eyes, and one placed above the nasion between the inner canthi of the eyes. EEG and EOG signals were sampled at 256 Hz and filtered between DC and 128 Hz.

3.2.3. Procedure

The study was advertised as an EEG study on individual differences in comparing numbers. Upon arrival, participants filled out a consent form and were assigned to start either with the high pressure or the low pressure condition. Both conditions were recorded within the same EEG session separated by a 2 min break. The order of pressure conditions was counterbalanced between groups as well as between female and male participants.

In the high pressure condition, participants were instructed that their performance in the numerical Stroop task would be evaluated. Firstly, outcome pressure was induced by telling participants that the computer will track their performance in the task and compute a performance score based on their response time and accuracy. Out of ten participants, the highest three performance scores would be rewarded with an additional €20, €10, and €5 respectively. Secondly, monitoring pressure was induced by filming participants during the task with the alleged intention to evaluate their behavior. For this, the experimenter mounted a webcam in front of the participant and demonstrated the functionality of the camera by transferring the recorded video onto participant's monitor while adjusting the camera.

In the low pressure condition, participants were instructed to respond as quickly and as accurately as possible, as common in psychological testing. No camera was recording the participants during the task and no performance-based reward was given.

The experimenter reminded participants before each pressure condition whether their performance would be evaluated or not. Except for the pressure manipulation, the experimental procedure of both conditions was identical: Participants were seated in an EEG cabin and EEG electrodes were placed. The recording started with a 3 min sequence in which participants were asked to deliberately execute eye movements. Then, a resting state EEG was recorded, which is not within the scope of the present study and is therefore not reported

in the following. Participants received task instructions on screen and then completed six practice trials for each task condition (i.e., numerical and physical) in which trial-based feedback was provided (viz., "correct", "wrong", "too slow"). At the end of the practice trials, participants were instructed on how to use the FAS and familiarized themselves with selecting a face by pressing the left and right button of the button box. Stimuli were presented in blocks with 11 numerical task blocks and three distractor blocks in which the physically larger number should be selected. After each numerical task block, participants rated their worries regarding their performance in the preceding block. Manipulated feedback was given with the aim to keep the error rate in the range from 5 to 25%. To find a better balance between controlling the error rate and the comparability between conditions (see Gehring et al., 2012), participants received a fixed negative feedback of being too slow after the numerical task block 2, 5, and 8. In the rest of the blocks, feedback was provided based on the error rate as described in the procedure section of Study 1. At the end of each condition, participants were given the state scale of the STAI to assess their state anxiety. After completing both conditions, participants additionally rated the importance, pressure, and success regarding each pressure condition. Each condition took about 25 min resulting in a total length of the experiment of about 1.5 h including instructions and (de)mounting of the electrodes.

Importantly, neither the participants nor the experimenter knew whether an individual was assigned to the LTA or the HTA group (double blinding). At the end of the experiment, participants were debriefed about the true purpose of the study and invited to contact the experimenter to learn about their individual level of test anxiety.

3.2.4. Data Analysis

3.2.4.1. Statistical Software

EEG data were analyzed using EEGLAB 14.1.1 (Delorme & Makeig, 2004) and ERPLAB 6.1.3 (Lopez-Calderon & Luck, 2014) in the MATLAB environment (Mathworks, Inc., Natick, MA, USA, release 2014b). Statistical analyses for both behavioral and electrophysiological data were carried out with the software SPSS 22.0 (IBM Corp. Released, 2011).

3.2.4.2. Behavioral Data

To check whether participants perceived the pressure condition as being more stressful than the control condition, 2 (pressure) x 2 (group) analyses of variance (ANOVAs) were conducted for the importance, pressure, and success rating as well as the state anxiety. Behavioral performance was analyzed by 2 (pressure) x 2 (group) x 3 (congruency) ANOVAs

with mean response time (in ms) and mean accuracy (in percentage) as dependent variables. Misses were excluded from computing the response time and treated as errors regarding the accuracy. Finally, self-reported worries were averaged across the 11 blocks of the numerical Stroop task for each pressure condition and analyzed with a 2 (pressure) x 2 (group) ANOVA.

3.2.4.3. Electrophysiological Data

The EEG signal was re-referenced to the mean (common average reference) and the baseline (channel mean) was removed. To correct for a delay of the USB button box, the time stamp of the button press was corrected for 20 ms. Eye movement artefacts were corrected by means of independent component analysis (ICA). For this, the signal was first high-pass filtered with 1 Hz and response-locked epochs from 700 ms before response onset to 700 ms after response were extracted. Then, the signal was manually controlled and epochs with non-stereotyped noise (e.g., excessive movement artefacts) were removed. The ICA was run on the remaining signal using the EEGLAB routine *runica*. For each participant and pressure condition, independent components reflecting vertical and horizontal eye movements were manually selected for removal. The EEG signal was then reloaded and preprocessed as described above but bandpass filtered with cutoffs of 0.1 and 30 Hz (see Olvet & Hajcak, 2009b). In the next step, the selected independent components were removed from the signal. To detect remaining artifacts, a moving window peak-to-peak threshold of 50 μV with a window size of 200 ms and a step size of 100 ms was applied to channel FCz. Two data sets were excluded from further analyses due to a rejection rate of more than 10% of all trials confirming the low data quality that had been noted by the experimenter during the recording. The valid trials were averaged for error responses (obtaining the ERN) and correct responses (obtaining the CRN) for each participant and pressure condition. To quantify the ERN and CRN, the peak amplitude (μV) between 0 and 100 ms at electrode FCz was determined using ERPLAB. The peak amplitude was used as the dependent variable in an ANOVA with the factors *response type* (correct vs error), *pressure* (low pressure vs. high pressure), and *group* (LTA vs. HTA). To test whether the ERN is related to worries during the test, the ERN amplitude was correlated with the mean worry rating for each pressure conditions and test anxieties group.

3.3. Results

3.3.1. Manipulation Check

Table 4 summarizes the differences in the mean rating of the manipulation check. The task was rated by both groups as relatively important given that all means were larger than five. There was a trend towards significance with participants rating the high pressure condition

as more important than the low pressure condition (for a summary of the ANOVAs, see Table 5). Importantly, the pressure rating in the high pressure condition was significantly higher than in the low pressure condition with a large effect size ($\eta_p^2 = .243$). Moreover, there was a significant trend of *group* with students with high test anxiety reporting more pressure than students with low test anxiety. Directionally, the state anxiety was higher and the success rating was lower in the high pressure than in the low pressure condition in each group, but differences failed statistical significance. Taken together, results suggest that the manipulation of performance pressure was effective in the present study.

Table 4. Means and standard deviations (in brackets) of the manipulation check ratings (R.) and the state scale of the State-Trait Anxiety Inventory (STAI) for each pressure condition and test anxiety group.

		Importance R.	Pressure R.	Success R.	State Anxiety
LTA	Low Pressure	5.26 (0.96)	3.78 (1.20)	3.70 (1.30)	42.13 (9.87)
	High Pressure	5.35 (1.03)	4.17 (1.40)	3.48 (1.44)	42.70 (10.83)
HTA	Low Pressure	5.19 (1.25)	4.10 (1.30)	3.86 (1.28)	43.90 (10.74)
	High Pressure	5.71 (0.72)	5.00 (1.10)	3.62 (1.07)	46.05 (11.22)

Note. R. = Rating, LTA = low test anxiety, HTA = high test anxiety, Importance, Pressure, and Success were rated using a 7-Point Likert Scale.

Table 5. Result summary of Analyses of Variance (ANOVAs) with the rating of importance, pressure and success as well as state anxiety as dependent variables.

	$F(1,42)$	p	η_p^2
Importance Rating			
Pressure	3.37	.074	.074
Group	0.34	.564	.008
Pressure x Group	1.72	.196	.039
Pressure Rating			
Pressure	13.52	< .001	.243
Group	2.86	.098	.064
Pressure x Group	2.12	.153	.048
Success Rating			
Pressure	2.41	.128	.054
Group	0.18	.675	.004
Pressure x Group	0.01	.944	< .001
State Anxiety			
Pressure	1.92	.173	.044
Group	0.70	.408	.016
Pressure x Group	0.65	.424	.015

Note. η_p^2 = Partial Eta Squared.

3.3.2. Behavioral Performance

As depicted in Figure 8a, the response time did not significantly differ between the pressure conditions nor between LTA and HTA students (for a summary of the ANOVA, see Table 6). Not surprisingly, the response time was significantly modulated by the congruency of the target stimulus. Bonferroni-corrected comparisons of the means revealed that the response time was lowest in the congruent condition ($M = 423.49$, $SD = 59.24$) and significantly increased in both the neutral condition ($M = 442.46$, $SD = 59.59$, $p < .001$) and the incongruent condition ($M = 458.12$, $SD = 61.34$, $p < .001$). A similar pattern of results was found regarding the mean accuracy (see Figure 8b). There was no significant effect of *pressure condition* nor of *group* on the percentage of correct responses (for a summary of the ANOVA, see Table 7). Again, the accuracy differed significantly between congruency conditions. The accuracy was highest responding to a congruent number pair ($M = 94.47$, $SD = 3.89$) and decreased significantly responding to neutral ($M = 91.71$, $SD = 5.77$, $p < .001$) and incongruent number pairs ($M = 72.22$, $SD = 14.36$, $p < .001$).
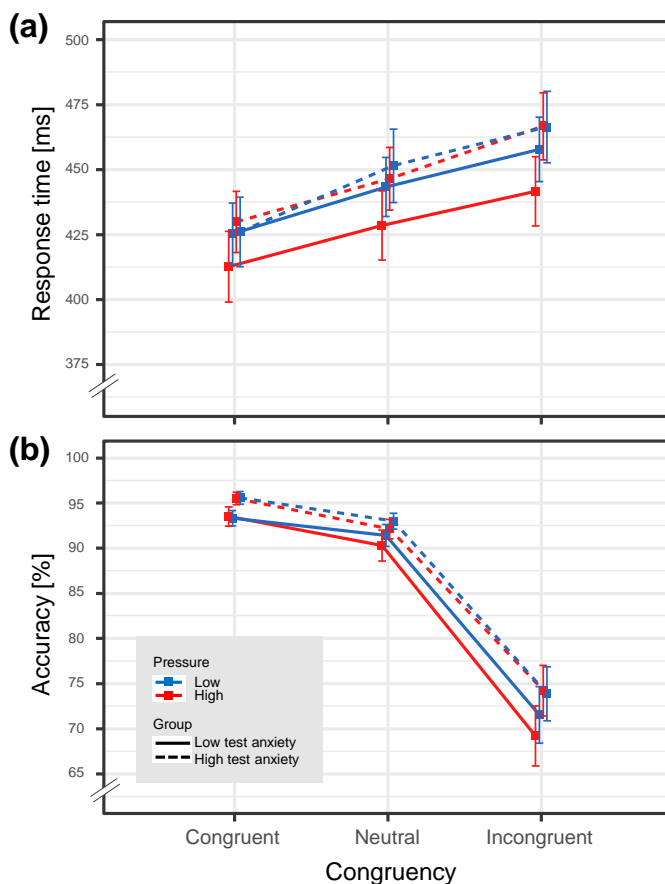
Figure 8. (a) Mean response time and (b) accuracy as a function of stimulus congruency, pressure condition and test anxiety. Error bars indicate +/- 1 standard error.

Table 6. Result summary of Analysis of Variance (ANOVA) with response time (in ms) as dependent variable.

	$F(1,42)$	p	η_p^2
Main effects			
Pressure	2.60	.114	.058
Congruency	115.44	< .001	.733
Group	0.56	.459	.013
Interactions			
Pressure x Group	2.42	.127	.055
Congruency x Group	1.47	.236	.034
Pressure x Congruency	2.18	.119	.049
Pressure x Congruency x Group	1.07	.347	.025

Note. η_p^2 = Partial Eta Squared.

Table 7. Result summary of Analysis of Variance (ANOVA) with accuracy (in % correct) as dependent variable.

	$F(1,42)$	p	η_p^2
Main effects			
Pressure	1.05	.311	.024
Congruency	135.53	< .001	.763
Group	1.46	.233	.034
Interactions			
Pressure x Group	0.51	.477	.012
Congruency x Group	0.25	.780	.006
Pressure x Congruency	0.71	.496	.017
Pressure x Congruency x Group	1.20	.307	.028

Note. η_p^2 = Partial Eta Squared.

3.3.3. Worry Rating

Means and standard deviations of the worry rating are summarized in Table 8. There was no significant main effect *Pressure* on worry, $F(1,42) = 0.78$, $p = .377$, $\eta_p^2 = .019$. Also, there was no significant difference between LTA and HTA students, $F(1,42) = 0.80$, $p = .382$, $\eta_p^2 = .018$. Contrary to our prediction, the interaction *Pressure x Group* failed statistical significance, $F(1,42) = 0.37$, $p = .544$, $\eta_p^2 = .009$.

Table 8. Means and standard deviations (in brackets) of the worry rating and correlations (Cor) with the ERN and accuracy (AC) for each pressure condition and test anxiety group.

		Low Pressure	High Pressure
	Mean (SD):	3.25 (1.08)	3.27 (1.15)
LTA	Cor (Worry ~ ERN):	$r = -.23$, $p = .299$	$r = -.02$, $p = .929$
	Cor (Worry ~ AC):	$r = -.57$, $p = .005$	$r = -.57$, $p = .004$
	Mean (SD):	3.45 (0.81)	3.57 (0.76)
HTA	Cor (Worry ~ ERN):	$r = .16$, $p = .503$	$r = .20$, $p = .382$
	Cor (Worry ~ AC):	$r = -.09$, $p = .699$	$r = -.27$, $p = .243$

Note. LTA = low test anxiety, HTA = high test anxiety

3.3.4. EEG Indices of Response Monitoring

The average waveforms of the CRN and the ERN are depicted in Figure 9. In the LTA group, the waveforms of the CRN run parallel in both pressure conditions. The peak amplitude of the ERN instead appeared to be slightly enhanced under performance pressure. In the HTA group, the waveforms of the CRN were again highly similar between pressure conditions. However, the time course of the ERN started to drift apart between pressure conditions just before the button press was registered. The amplitude of the ERN was noticeably enhanced in the pressure condition as compared to the low pressure condition.

The 2 (pressure) x 2 (group) x 2 (response type) ANOVA is summarized in Table 9. Results revealed a significant main effect *Response Type* with errors eliciting a more pronounced negativity than correct responses. Pairwise comparisons confirmed that the peak amplitude of the ERN was significantly more negative than the peak amplitude of the CRN across pressure conditions and groups (highest p value = .001; see Table S 2). Furthermore, there was a significant main effect *Pressure* as well as two significant interactions *Response Type* x *Pressure* and *Pressure* x *Group*. These effects were qualified by the significant 3-way interaction *Response Type x Pressure x Group*. Pairwise comparisons revealed that the ERN was significantly enhanced in the high pressure condition as compared to the low pressure condition in the HTA group ($p < .001$). However, there was no significant difference between pressure conditions regarding the CRN ($p = .418$). In the LTA group, neither the CRN ($p = .751$) nor the ERN ($p = .623$) significantly differed between pressure conditions. Directly comparing LTA and HTA students revealed that there was no significant difference regarding the CRN ($p = .882$) or the ERN ($p = .807$) in the high pressure condition. Similarly, the groups did not differ regarding the CRN ($p = .800$) or the ERN ($p = .119$) in the

low pressure condition. Taken together, results suggest that the ERN was significantly modulated by performance pressure in the HTA group but not in the LTA group. The peak amplitude for both test anxiety groups is depicted in Figure 10 as a function of response type and performance pressure.
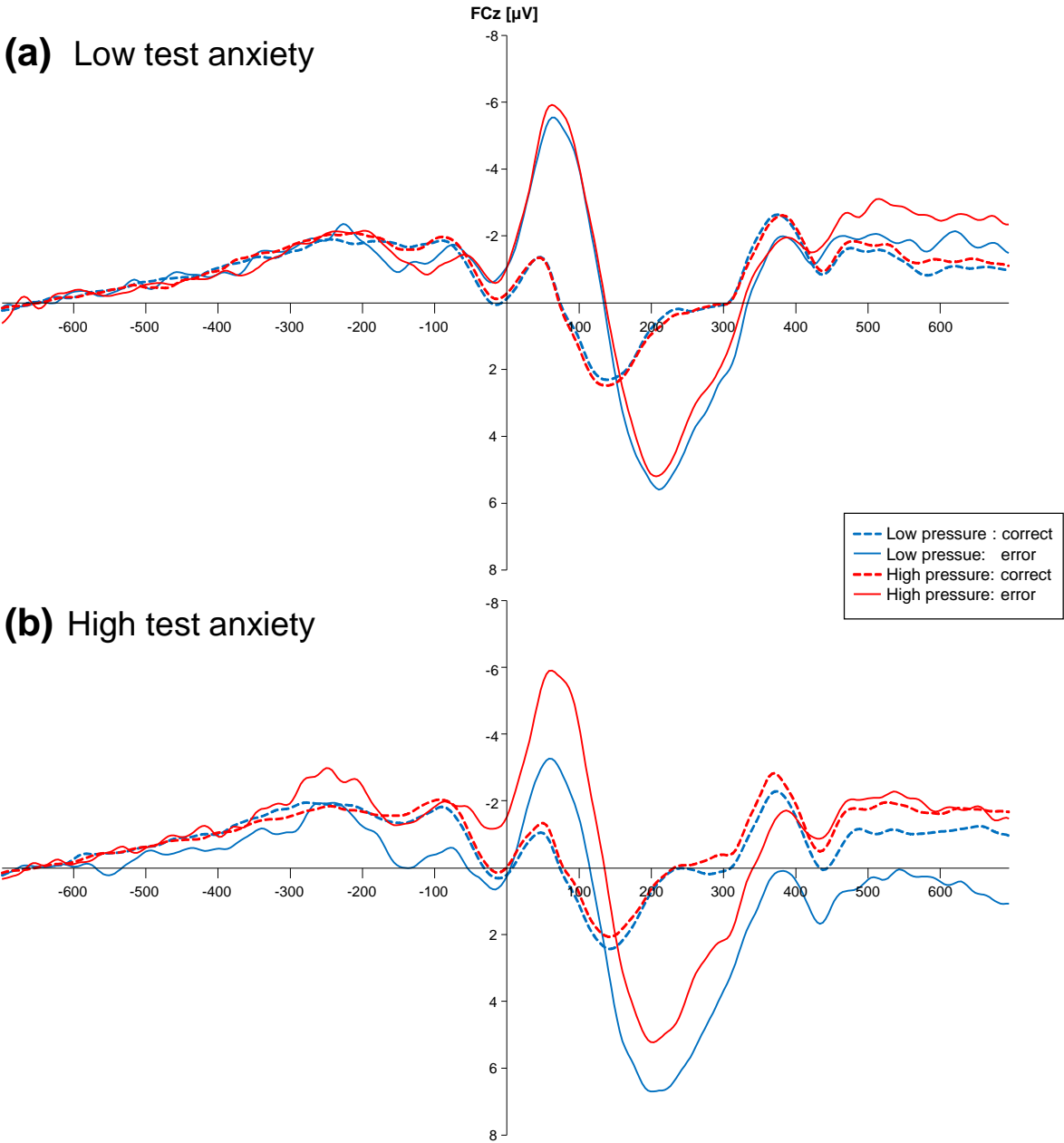
Figure 9. Average waveforms of correct and erroneous responses in each pressure condition for (a) low test anxiety (LTA) and (b) high test anxiety (HTA) students.

Table 9. Result summary of Analysis of Variance (ANOVA) with the peak amplitude (in μV) as dependent variable.

	$F(1,42)$	p	η_p^2
Main Effects			
Response Type	113.13	< .001	.729
Pressure	6.56	.014	.135
Group	0.19	.663	.005
Interactions			
Response Type x Pressure	8.82	.004	.174
Response Type x Group	1.20	.280	.028
Pressure x Group	4.84	.033	.103
Response Type x Pressure x Group	4.09	.049	.089

Note. η_p^2 = Partial Eta Squared.
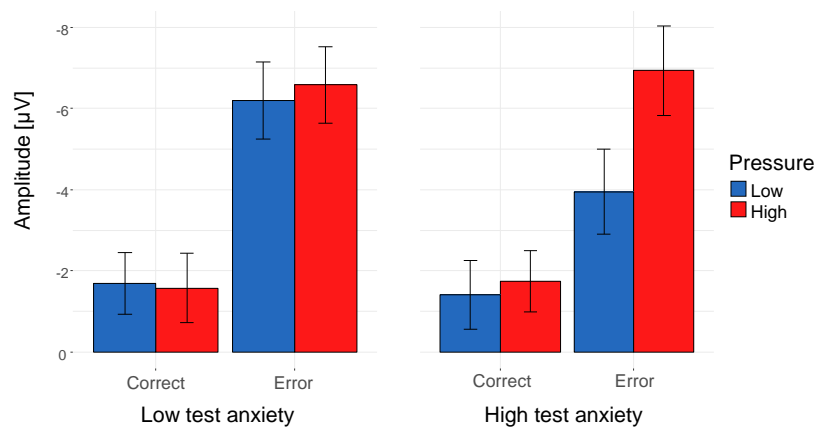
Figure 10. Peak amplitude as a function of response type, pressure condition, and test anxiety. Error bars indicate +/- 1 standard error.

3.3.5. The ERN and Worries

Since the previous analysis revealed that the mean difference in the CRN between pressure conditions and test anxiety groups was small and insignificant (see Figure 10), correlations with the worry rating were computed with the ERN instead of the ΔERN.

Contrary to our predictions, self-reported worries were not significantly correlated with the amplitude of the ERN in LTA or HTA students in either pressure condition (see Table 8). Post-hoc analyses were conducted to test whether worries were related to the number of errors in the numerical Stroop task. Correlation analyses revealed that LTA students with lower accuracy reported significantly more worries in both conditions (see Table 8). However, there was no significant relationship between the accuracy and the worry rating in either pressure condition for HTA students.

3.4. Discussion

The present study aimed to extend Study 1 by (a) comparing the response monitoring of low and high test-anxious students while minimizing differences in general anxiety and (b) directly relating the EEG indices of response monitoring to worries reported under performance pressure. To this end, 341 university students were pre-screened for their test anxiety and general anxiety as part of a larger test battery (see Schillinger et al., 2018). The selection procedure aimed to form a HTA and a LTA group while minimizing differences in general anxiety between groups. Both groups were tested in a high pressure and a low pressure condition performing a numerical Stroop. To directly relate the ERN to worries, participants were asked to rate their worries at regular intervals throughout the experiment using an FAS (Bieri et al., 1990; Punaro & Reeve, 2012; Trezise & Reeve, 2014b). Based on the results of Study 1, a number of changes in the present methods were implemented including focusing on the numerical task of the Stroop paradigm, recording both pressure conditions within a single EEG session, checking the effectiveness of the pressure manipulation, and analyzing both the ERN and the CRN using an ANOVA approach.

Data of 44 participants were included in the present analysis, 23 being assigned to the LTA group and 21 assigned to the HTA group. The two groups differed significantly in test anxiety with most participants in the LTA group being classified as either having very low or low levels of test anxiety and most participants in the HTA group having high or very high levels of test anxiety. However, we failed to completely control for differences in general anxiety between the LTA and the HTA group. More specifically, the HTA group also exhibited significantly higher general anxiety than the LTA group. Note that the effect size of the difference in general anxiety (Cohen's d = -1.52) was markedly lower than the effect size in test anxiety (Cohen's d = -5.41). Taken together, the selection procedure was successful in forming two extreme groups which are characterized primarily by their difference in test anxiety.

3.4.1. Manipulation Check

The manipulation check revealed that both pressure conditions were perceived as highly important by both LTA and HTA students as indicated by mean importance ratings being within the upper range of the 7-Point Likert scale across conditions (see Table 4). There was no significant difference between test anxiety groups but participants tended to generally rate the high pressure condition as being more important than the low pressure condition. The present importance ratings were also high as compared to the study by Beilock and colleagues (2004) based on which the manipulation check was adapted. There, participants reported slightly lower importance in both the control group (M = 4.63) and the pressure group (M =

5.03) than in the present study. That the importance ratings were high across pressure conditions is an important indicator that participants were motivated to perform the numerical Stroop task throughout the experiment. Differences between pressure conditions are therefore unlikely to result from a drop of motivation in the low pressure condition. Regarding the perceived success in the numerical Stroop, ratings were in the average range and tended to be lower under performance pressure for participants in both groups similar to the results reported by Beilock and colleagues (2004). Finally, pressure ratings were in the average or above average range with means ranging between $M = 3.78$ for LTA students in the low pressure condition and $M = 5.00$ for HTA students in the high pressure condition. Importantly, participants generally reported more performance pressure in the high pressure condition than in the low pressure condition. This difference was highly significant with a large effect size ($\eta_p^2 = .243$). Moreover, the HTA group tended to report more performance pressure than the LTA group across pressure conditions. The present pressure ratings are highly similar to those reported by Beilock and colleagues in both the low pressure ($M = 3.95$) and the high pressure condition ($M = 5.08$). A similar pattern of results was present regarding the state anxiety with the mean state anxiety ranging between $M = 42.13$ for LTA students in the low pressure condition and $M = 46.05$ for HTA students in the high pressure condition. However, the difference between the pressure conditions as well as between test anxiety groups failed statistical significance. This is in discordance with the results by Beilock and colleagues (2004) who have reported significantly higher state anxiety in the high pressure condition ($M = 42.68$) than in the low pressure condition ($M = 32.08$). However, note that the difference between the two studies is due to participants reporting higher state anxiety in the low pressure condition of the present study than in the previous study. In fact, the state anxiety in the low pressure condition was on the same level as in the high pressure condition in Beilock's study (2004). Taken together, the manipulation check confirmed that the task motivation was high throughout the experiment and that the pressure manipulation was effective.

3.4.2. Behavioral Performance

On a behavioral level, both response time and accuracy showed a significant effect of congruency with responses to incongruent trials being slower and less accurate than responses to neutral and congruent trials. However, there was no significant difference regarding the behavioral performance between test anxiety groups or pressure conditions. In contrast, we found in Study 1 that participants under performance pressure responded slower but more accurately than in the control condition. This pattern of results suggested that participants had changed their response criterion emphasizing accuracy at the cost of response times. This was clearly not the case in the present study with participants rather

responding faster and more accurately in the high pressure condition than in the low pressure condition. Furthermore, test anxiety was reported to be inversely related to the response time in Study 1. In the present study, in contrast, there was no significant difference between LTA and HTA student regarding their response time. One reason for this discrepancy between the two studies could be the different samples of students. While the sample of the first study consisted of psychology students, we tested in the present study both psychology students and students with other fields of study. As such, students in the present study are likely to have been less familiar with psychological testing. Furthermore, the majority of participants in the present study received money for their participation, while the former participants received course credits. Receiving money instead of course credits is likely to increase the motivation of participants, which is in line with the high importance ratings previously discussed. Taken together, students in the present study are likely to have uphold a high level of attention throughout the experiment, explaining the result that their performance did not differ between pressure conditions. That there are no behavioral differences between test anxiety groups is in accordance with the majority of previous studies addressing the ERN in anxious individuals (for a meta-analysis, see Moser et al., 2013).

3.4.3. Test Anxiety and Worries

Across conditions, the mean worry ratings were in the average range of the 6-Point FAS with means ranging between $M = 3.25$ for LTA students in the low pressure condition and $M = 3.57$ for HTA students in the high pressure condition (see Table 8). Contrary to our hypothesis, HTA students did not report significantly more worries in the high pressure condition than LTA students. This is surprising given that test-anxious students are characterized based on the distraction model to worry about their performance in an evaluative situation. However, it should be noted that the present pattern of results is numerically in line with this prediction. More specifically, the two test anxiety groups were more similar in the low pressure condition (mean difference = 0.20) than in the high pressure condition (mean difference = 0.30). However, the difference in the high pressure condition was still subtle and clearly failed statistical significance. A post-hoc analysis revealed that worries were significantly related to the accuracy in the numerical Stroop task in LTA, but not in the HTA students (see Table 8). While LTA students seemed to start worrying about their performance when realizing that they had made errors, this appeared not to be the case for HTA students. This difference in reporting worries might have contributed to diluting a potential difference in the worry rating between LTA and HTA students in the high pressure condition.

3.4.4. The Modulation of the ERN by Performance Pressure

On an electrophysiological level, the analysis confirmed that the ERN was more negative than the CRN across all conditions as can be expected based on the literature (for a review, see Gehring et al., 2012). The analysis further revealed a significant 3-way interaction between the response type (CRN vs. ERN), performance pressure (low vs. high), and group (LTA vs. HTA) (see Figure 10). Pairwise comparisons showed that the CRN did not differ significantly between pressure conditions in either test anxiety group. The ERN, instead, was significantly enhanced under performance pressure for HTA students but not for LTA students. Directly comparing the ERN between groups revealed that there was no significant difference in either pressure condition. Taken together, results show that the ERN was modulated by performance pressure in HTA but not in LTA students. In contrast, we reported in Study 1 that the difference score between CRN and ERN (ΔERN) but not the ERN alone was significantly modulated by performance pressure. Comparing the bar plots of Study 1 (Figure 4) and Study 2 (Figure 10) shows that the sample of Study 1 was more similar to the present LTA than to the present HTA group. More specifically, the amplitude of the CRN was lower and the ERN was higher under performance pressure in both the sample of Study 1 and the present LTA group. Differences between pressure conditions were small and each the CRN and the ERN alone did not differ significantly. However, adding up the reverse effects in computing the ΔERN explains why we found a significant effect of performance pressure on the ΔERN in the previous study. In the present study, the difference between CRN and ERN was taken into account by including the factor *Response Type* in the ANOVA, revealing that the effect of performance pressure on the ERN relative to the CRN did not reach statistical significance in the LTA group. Given that the LTA group consisted of students with mostly very low or low levels of test anxiety as compared to mostly average levels of test anxiety in the sample of Study 1, this is in support of a linear relationship between performance pressure and test anxiety in modulating the ERN. In greater detail, the subsample with the lowest test anxiety scores (i.e., the present LTA group) showed no significant modulation of the ERN or the ΔERN. The sample of Study 1 having mostly average levels of test anxiety displayed a significant modulation of the ΔERN due to the reverse effects of an increasing ERN and a decreasing CRN. Finally, the amplitude of the ERN was significantly enhanced by performance pressure in the subsample with the highest test anxiety (i.e., the present HTA group). Taken together, results suggest that the ERN is enhanced by performance pressure in higher but not in lower test-anxious students.

The present EEG analysis revealed that the ERN was modulated by performance pressure in HTA but not in LTA students. At the same time, both test anxiety groups performed on the same level in both pressure conditions as indicated by similar response times and

accuracies. This pattern of results is in line with the *compensatory error-monitoring hypothesis* suggested by Moser and colleagues (2013). According to this hypothesis, the increased ERN in anxious individuals reflects compensatory efforts to maintain task performance. More specifically, anxious individuals are thought to focus their attention on internal worries, thereby reducing resources dedicated to performing the actual task. As a result, anxious individuals have to increase their attentional control to maintain task performance, including monitoring ongoing responses, and to dynamically adapt to errors. This compensatory effort is thought to give rise to more neural activity after committing an error as indicated by an increased ERN amplitude. In the present study, HTA students showed an increased ERN under performance pressure, suggesting that test-anxious students need compensatory effort to monitor their responses and to dynamically adapt to errors in an evaluative situation.

Contrary to our hypothesis, we did not find that both test anxiety groups displayed an enhanced ERN under performance pressure. This seems to be in discordance to the study by Hajcak and colleagues (2005) who demonstrated in two experiments that the amplitude of the ERN was enhanced by evaluation as compared to a control condition. However, Hajcak and colleagues (2005) did not include individual differences in anxiety in their analysis. If we test the effect of performance pressure in the present study across the entire sample discarding the factor *group*, we find that the ERN is significantly enhanced in the high pressure condition as compared to the low pressure condition, $t(43) = 2.80$, $p = .007$ (paired-sample t-test). Thus, instead of being in discordance, the present study refines the findings by Hajcak and colleagues (2005). The previously reported effect of an enhanced ERN by evaluation (Hajcak, et al., 2005) seems to only hold true for individuals with higher levels of anxiety.

Also, we predicted that HTA students would show a higher ERN in the high pressure condition than LTA students, while no difference between groups was expected in the low pressure condition. The amplitude of the ERN was in fact higher for HTA than for LTA students in the high pressure condition, but the mean difference was small and clearly failed statistical significance (see Table S 2). In contrast, the largest yet also insignificant difference between groups was a smaller ERN for HTA than for LTA students in the low pressure condition (see Figure 10). These findings appear puzzling given that previous studies have shown that the ERN is increased in anxious individuals (for a meta-analysis, see Moser et al., 2013). However, these studies addressed the ERN mostly in individuals with more general symptoms of anxiety and worry, including general anxiety disorder (see Weinberg et al. 2011), obsessive-compulsory disorder (see Weinberg et al. 2011), general anxiety (Hajcak et al., 2003), negative affect (Hajcak et al., 2004; Luu et al., 2000), math anxiety (Suárez-Pellicioni et al., 2013), perfectionism (Perrone-McGovern et al., 2017), and helplessness (Pfabigan et al., 2013). In the present study, instead, we specifically addressed the ERN in individuals varying in situation-specific test anxiety while minimizing differences in general anxiety. It appears

therefore plausible that the difference in the ERN under performance pressure between LTA and HTA turned out subtle in the present study. This would also offer an explanation of why we found a significant main effect of test anxiety in Study 1 with a smaller sample size than in the present study. In Study 1, differences in general anxiety were not controlled for and these differences likely contributed to increase the ΔERN. However, it remains puzzling why the HTA group appears to exhibit a decreased ERN in in the low pressure condition of the present study. An attenuated ERN has been reported in patients with disorders such as severe depression, Attention deficit hyperactivity disorder (ADHD), schizophrenia, and substance abuse (for a review, see Weinberg et al. 2011). Weinberg and colleagues (2011) have suggested that the lower ERN in these individuals reflects disengagement from the task in the course of the experiment. In line with this explanation, these patients show, unlike anxious individuals, a lower performance in the paradigms used to measure the ERN as compared to healthy controls (Weinberg et al., 2011). In the present study, however, participants of both groups rated the low pressure condition as highly important. Furthermore, LTA and HTA students performed on the same level in both pressure conditions. Therefore, task disengagement of HTA students in the low pressure condition seems to be unlikely to account for the present tendency of a decreased ERN in these students. Another explanation could be that HTA students were more relieved than LTA students once they had completed the high pressure condition. This state of mind could have lowered the ERN in the subsequent low pressure condition. Note, however, that the order of the pressure conditions was counterbalanced in the present study. To test whether the order of testing had an influence on the ERN in HTA students we run an additional 2 (response type) x 2 (pressure) x 2 (order) ANOVA with the subsample of HTA students. Results revealed that the order of testing had no significant effect on the CRN or ERN (see Table S 3). Future research is therefore needed to (a) clearly show that HTA students exhibit a lower ERN amplitude than LTA students with neutral task instruction and if so (b) to provide an explanation accounting for this phenomenon.

Finally, the amplitude of the ERN did not significantly correlate with the worry rating of LTA or HTA students in either pressure condition. This is surprising given that worries have been suggested to be the component of anxiety most closely related to the reported increase of the ERN in anxious individuals (for a meta-analysis, see Moser et al., 2013). For instance, Moser and colleagues (2012) reported that the ERN amplitude was inversely related to worry but not to arousal in female undergraduate students. However, previous studies have, to the best of our knowledge, not assessed worries as a state, i.e., in the situation in which the ERN was recorded. Instead, worries were assessed as a trait, i.e., as the general tendency to worry, by using questionnaires. It remains therefore unclear whether the reported link between worries and the ERN is specific to the general tendency of anxious individuals to worry or the actual worries experienced by anxious individuals during the task. The present results suggest

rather that the worries experienced during the task are not related to the amplitude of the ERN. However, the worry rating in the present study appeared to be also related to the actual task performance, especially in LTA students. This raises the question of whether the FAS used in the present study actually assessed worries about performing well or rather participants' awareness of having made errors. Future research would therefore need to evaluate the FAS with adult participants in the context of learning and possibly improve the scale to assess performance-related worries. Taken together, the present study remains inconclusive regarding the role of worries in modulating EEG indices of response monitoring.

3.4.5. Conclusion

The manipulation check demonstrated that the pressure manipulation was effective in the present study. On an electrophysiological level, the analysis revealed that the ERN was modulated by performance pressure in HTA students but not in LTA students. While the amplitude of the ERN was significantly enhanced in the pressure condition as compared to the control condition in HTA students, the amplitude did not significantly differ between pressure conditions in LTA students. At the same time, both test anxiety groups performed at the same level in the numerical Stoop task in both pressure conditions. Worries as assessed by an FAS turned out not to be significantly related to the amplitude of the ERN in either group. Instead, worries appeared to be directly related to the task performance for LTA but not for HTA students. Taken together, results are in line with the *compensatory error-monitoring hypothesis* (Moser et al., 2013), suggesting that test-anxious students need compensatory effort to monitor their responses and to dynamically adapt to errors in an evaluative situation.

4. Study 3 – Working Memory Costs of Performance Pressure

4.1. Introduction

4.1.1. Choking Under Pressure and Working Memory Demands

The previous two studies have demonstrated that performance pressure affects the ERN in test-anxious students, suggesting that these students need compensatory effort to monitor their responses and to dynamically adapt to errors in an evaluation situation. However, task performance was not (Study 2) or not unambiguously (Study 1) impaired in test-anxious students as commonly reported by studies addressing the ERN in anxious individuals (see Gehring et al., 2012; Moser et al., 2013). One reason for the lack of performance impairments might be the relatively low working memory demands of the experimental paradigms typically used to measure the ERN, such as Flanker or Stroop tasks (see Gehring et al., 2012). For instance, the numerical Stroop paradigm used in the present research requires participants to compare two numbers in a short amount of time. While it is challenging to respond to the pairs of numbers given the time constrain, the comparison between the two numbers itself put only minimal working memory demands on participants. Consequently, it is not surprising that test-anxious students can compensate for the postulated effect of worries on their working memory (Calvo & Eysenck, 1992; Eysenck et al., 2007; Wine, 1971). However, such rather low working memory demands stand in contrast to the demands of tasks students have to face in real-world tests or examinations. Here, working memory demands can be expected to be higher especially in the domain of mathematics (see Raghubar et al., 2010). When working memory demands increase, there is less space to compensate for distracting worries and consequently performance impairments are more likely to result.

In a seminal study, Beilock and colleagues (2004) addressed the phenomenon of *choking under pressure* (Baumeister, 1984) in mathematical problem solving. For this, they used so-called Gauss modular arithmetic as stimulus material (Gauss, 1801, as cited in Neumann, 2005). The objective of modular arithmetic is to judge the truth value of problem statements such as "$62 \equiv 37 \pmod 5$". The problem is solved by subtracting the second number from the first number (i.e., 62 - 37) and then dividing the resulting difference by the last number (i.e., $25 \div 5$). If the dividend is a whole number (here, 5), the problem is true. If there is a remainder, the statement it false. The modular arithmetic problems either had a low or a high demand. While low demand problems consisted of two operands smaller than 10 (e.g., "$9 \equiv 2 \pmod 4$"), high demand problems consisted of two 2-digit numbers and required a carry-over operation (e.g., "$62 \equiv 37 \pmod 5$"). An advantage of modular arithmetic as an experimental task is that is based on common arithmetic operations (i.e., subtraction and division) and is therefore highly ecologically valid for real-world tests assessing arithmetic

skills. At the same time, the notation of the problems as well as the specific sequence to solve them is unfamiliar to most students, allowing to control to some degree for previous task experience (see also Ramirez & Beilock, 2011). In the study by Beilock and colleagues (2004), participants were randomly assigned to either a control group or a pressure group prior to performing three blocks of modular arithmetic, with half of the problems having a low demand and the other half having a high demand. The first block served as a pretest of arithmetic performance and the second block was included to provide more practice. Crucially, the two groups received different instructions preceding the third block of problems. While the control group was informed that they would be performing another set of problems, the pressure group was exposed to a high pressure test scenario. For this, participants were instructed that they would receive an additional $5 if they could improve their performance in the last block as part of a team effort. More specifically, participants were told that they were paired with another participant who had already managed to improve her/his performance and that both would receive the reward if they could also improve in the upcoming set of problems. Finally, participants were instructed that they would be filmed during the task so that math teachers and scientists could examine their performance. Taken together, these measures were reasoned to exert performance pressure on participants similar to an actual high-stake test. This was confirmed by a manipulation check administered to both groups after completing the last block of arithmetic problems. Participants assigned to the pressure group reported significantly more performance pressure as well as higher state anxiety than those assigned to the control group. The accuracy in the first block was then compared as a pretest with the third block as a posttest of arithmetic performance. Results revealed that both groups could significantly improve in solving low demand problems in the last block. However, while the control group could also improve in solving high demand problems, the pressure group showed a marked drop in performance for these problems. This *choking under pressure effect* could be replicated by other studies by Beilock and colleagues (Beilock & Carr, 2005; Beilock & DeCaro, 2007; Beilock et al., 2007; Ramirez & Beilock, 2011) as well as by other groups (Boere, Fellinger, Huizinga, Wong, & Bijleveld, 2016, but cf. Benny & Banks, 2015). Results are in line with distraction theories (Calvo & Eysenck, 1992; Eysenck et al., 2007; Wine, 1971), suggesting that the detrimental effect of performance-related worries can be compensated for when task demands are low but will start to impact performance once the combined demands of task-related and worry-related processes exceed the available working memory resources.

4.1.2. Individual Differences in Test Anxiety

The reviewed studies above have not related individual differences in test anxiety to the *choking under pressure effect* in mathematics. This is surprising given that distraction

theories predict that test-anxious students are especially prone to worry in an evaluative situation and hence to choke under pressure (Calvo & Eysenck, 1992; Eysenck et al., 2007; Wine, 1971). First evidence in support of this notion comes from a study by Ramirez and Beilock (2011). First, they showed that a short intervention of expressive writing, which is thought to alleviate worries and ruminations, can neutralize the *choking under pressure* in two laboratory experiments. The effectiveness of the intervention was then tested in helping students to cope with performance pressure in an actual examination. For this, ninth graders were asked just before their final high-school exam in biology to either write 10 min about their feelings towards the upcoming exam (expressive writing group) or to think about a topic that would not be covered in the exam (control group). The analysis focused on the relationship between students' test anxiety and their final exam scores. While test anxiety was inversely related in the control group, this relationship was insignificant for students who had been assigned to the expressive writing condition. This suggest that the intervention was effective in reducing worries in test-anxious students and that without this burden students could improve their performance in the final exam. In reverse, it suggests that test-anxious students in the control condition performed lower than their actual skill level, hence, that they choked under pressure. This finding corresponds to the well documented inverse relationship between test anxiety and test performance (for reviews, see Ackerman & Heggestad, 1997; Chapell et al., 2005; Hembree, 1988; Seipp, 1991) and to converging evidence that this relationship is due in large parts to the detrimental effect of test anxiety on cognitive performance (for reviews, see Eysenck et al., 2007; Hembree, 1988; Mowbray, 2012; Wine, 1971). However, to the best of our knowledge, no laboratory study has yet directly related test anxiety to the phenomenon of *choking under pressure*. The first aim of the present study was therefore to test whether the degree of *choking under pressure* depends on individual differences in test anxiety.

4.1.3. Frontal Midline Theta (FMϴ)

The second aim of the study was to assess the working memory costs of performance pressure by measuring oscillatory brain responses during the task. In the EEG, mental activity is known to be associated with specific neural oscillations in different frequency bands. Most prominently, event-related synchronization (ERS) in theta frequency band (4-8 Hz) has been linked to working memory processes (for reviews, see Klimesch, 1999; Roux & Uhlhaas, 2014; Sauseng, Griesmayr, Freunberger, & Klimesch, 2010). Such working-memory related differences in the theta frequency band have been shown to be most pronounced at fronto-central recording sites, referred to as frontal midline theta (FMϴ) (for reviews, see Hsieh & Ranganath, 2014; Inanaga, 1998). In particular, FMϴ has been demonstrated to increase with the memory load in classical working memory tasks, such as the Sternberg task (e.g., Jensen

& Tesche, 2002) and the N-back task (e.g., Gevins, Smith, McEvoy, & Yu, 1997). Interestingly, ERPs following uncertain events such as errors (error-related negativity; ERN), punishment (feedback-related negativity; FRN), and conflict-inducing stimuli (N2) have been recently shown to originate from a transient oscillatory synchronization of FMϴ (for reviews, see Cavanagh & Frank, 2014; Cavanagh & Shackman, 2015). These ERPs, in turn, have been consistently reported to increase with the level of anxiety (for a meta-analysis, see Cavanagh & Shackman, 2015). Based on these findings, Cavanagh and Shackman (2015) have proposed the *adaptive control hypothesis* according to which the increased FMϴ in anxious individuals reflect the need for attentional control regarding uncertain or threatening events. However, only few studies have directly related symptoms of anxiety or situational stress to FMϴ in the time-frequency domain (Osinsky, Karl, & Hewig, 2017). Unlike ERPs, time-frequency analyses comprise both phase-locked and non-phase-locked neural activity (see Kalcher & Pfurtscheller, 1995). In one study, Mueller and colleagues (2015) could demonstrate that the state anxiety during a reinforcement learning task was linearly related to FMϴ in a group of depressive patients but not in a control group. Similarly, FMϴ was shown to be increased in individuals with higher neuroticism following negative feedback (Mueller et al., 2014). Other studies have reported a correlation between measures of trait anxiety and an ERS in the theta frequency band although at more posterior (Balconi & Crivelli, 2010) and lateral (Neo, Thurlow, & McNaughton, 2011) recording sites. In a recent study by Osinsky and colleagues (2017), participants were assigned to either a control group or a threat anticipation group prior to performing a simple two-choice task. The control group was told that they had to write an anonymous pro-and-con list about a topic of which they would be informed after the EEG session. In contrast, the threat anticipation group was informed that they have to give a speech after the EEG session, which would be videotaped and evaluated by a jury. Results revealed that FMϴ was higher during the task in the threat anticipation than in the control group for female but not for male students. In contrast, two studies have reported decreased FMϴ in anxious individuals under stress during an N-back task (Gärtner, Rohde-Liebenau, Grimm, & Bajbouj, 2014) and a mental arithmetic task (Gärtner, Grimm, & Bajbouj, 2015). In these studies, however, stress was induced by presenting an aversive, disturbing video clip to participants prior to the task which arguably involved different mechanisms as the threat of being evaluated.

4.1.4. Present Study

The aim of the present study was (a) to test whether test anxiety is related to the degree of *choking under pressure*, and (b) to directly assess the working memory costs of performance pressure by measuring oscillatory brain responses during the task. To this end, we adapted the paradigm reported by Beilock and colleagues (2004) to meet the requirements of an EEG study. Participants were randomly assigned to either a control group or a pressure group prior to performing two blocks of modular arithmetic, with half of the problems having a low demand and the other half having a high demand. Skipping the second block of arithmetic problems allowed us to increase the number of trials in each block to 60. Increasing the number of trials was important for improving the signal-to-noise ratio for the EEG analysis. Note that previous studies by Beilock and colleagues have also compared two instead of three blocks (Beilock & Carr, 2005; Ramirez & Beilock, 2011) with similar results as reported in the first study (Beilock et al., 2004). In the present study, the first block of problems was used as a pretest of arithmetic performance. Preceding the second block of problems, the two groups received different instructions. While the control group was informed that they would be performing another set of problems, the pressure group was exposed to a high pressure test scenario. This scenario was based on the study by Beilock and colleagues (2004) including a performance-based reward of €10 and filming participants during the second block. However, we suspected that participants might not believe the cover story regarding the team effort as used by the original study given that they could be expected to be familiar with psychological testing. Instead, participants were informed that of the previous participants all but one could manage to improve their performance in order to induce pressure by way of a social comparison. The effectiveness of the manipulation was checked by the same procedure reported by Beilock and colleagues (2004). More specifically, participants were asked at the end of block 2 to rate their state anxiety as well as importance, pressure, and success regarding the second block. The degree of choking (i.e., the difference in accuracy between blocks) was then correlated with the individual level of test anxiety in both groups.

Since previous studies did not assess the worries of participants during the experiment, the explanation that performance-related worries impaired the working memory of students assigned to the pressure groups remained to some degree speculative. To overcome this limitation, participants were asked in the present study to rate their worries at regular intervals throughout the experiment using a Faces Anxiety Scale (FAS) (Bieri et al., 1990; Punaro & Reeve, 2012; Trezise & Reeve, 2014b).

Working memory costs of performance pressure were assessed by analyzing the event-related (de-)synchronization (ERS/ERD) (Pfurtscheller & Aranibar, 1977; Pfurtscheller & Lopes da Silva, 1999) in the theta frequency band (4-8 Hz) (Hsieh & Ranganath, 2014;

Inanaga, 1998; Klimesch, 1999; Roux & Uhlhaas, 2014; Sauseng et al., 2010). In the first step of the analysis, the ERS/ERD during block 2 was compared between the two groups for each task demand using cluster-based permutation tests (Maris & Oostenveld, 2007). We reasoned that the pressure group should exhibit a stronger ERS in the theta frequency band than the control group at frontal midline recording sites (i.e., Fz, FCz, and Cz). FMΘ was then defined at those frontal midline electrodes that were revealed by the cluster-based permutation tests to differ significantly between groups for both demands. Then, the differences in FMΘ between blocks was correlated with participants' test anxiety scores. Finally, mediation analyses were used to test whether the hypothesized increase of FMΘ in test-anxious students in the pressure group could be explained by an increase in worries.

4.1.5. Hypotheses

Firstly, we expected to replicate the *choking under pressure effect* reported by previous studies (Beilock & Carr, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004, 2007; Boere et al., 2016; Ramirez & Beilock, 2011). Regarding the manipulation check, participants assigned to the pressure group were predicted to report more pressure and higher state anxiety than those assigned to the control group. However, the rating of importance and of success was expected to be similar between groups. Regarding the behavioral performance, both groups were expected to show a higher accuracy for low demand problems in block 2 as compared to block 1. In contrast, the accuracy in solving high demand problems was predicted to increase in the control group but to decrease for the pressure group in block 2. Secondly, we hypothesized that the degree of choking in the high pressure condition is linearly related to the individual level of test anxiety.

Participants in the pressure group were hypothesized to report more worries than participants assigned to the control group. The increase in worries between blocks was predicted to correlate with individual level of test anxiety.

Using cluster-based permutation tests, we predicted that the pressure group would exhibit a stronger ERS in the theta frequency band than the control group at frontal midline recording sites. FMΘ was then defined at frontal midline electrodes which turned out to differ significantly between groups for both task demands. The difference in FMΘ between blocks was predicted to correlate with individual test anxiety in the pressure group but not in the control group. Finally, we hypothesized that the increase of FMΘ in test-anxious students in the pressure group would be mediated by an increase in worries.

4.2. Methods

4.2.1. Participants

Sixty-four university students participated in the present study and were randomly assigned to a pressure group or a control group. Two data sets (one in each group) had to be excluded due to excessive non-stereotyped artifacts in the EEG, one data set (pressure group) due to an accuracy rate lower than chance rate (50%) for the high demand problems in block 1, and one data set due to a participant refusing to fill out the online questionnaires (pressure group). Thus, the remaining sample comprised 31 participants in the control group and 29 participants in the pressure group. As can be seen in Table 10, the two groups did not differ significantly regarding their demographics including age and gender distribution. Moreover, the two groups were highly similar in their level of both test anxiety and general anxiety.

All participants were right-handed with normal or corrected-to-normal vision and no history of psychological or neurological diseases. They gave written informed consent regarding the arithmetic task and the EEG recording, but were not informed until the end of experiment about the true purpose of the study. Furthermore, psychology and mathematics students were excluded from participating in the present study (see also Beilock et al., 2004). The former group was excluded because they could be expected to be less likely to believe that their performance would be actually evaluated. The latter group was excluded because the arithmetic task might have been too easy for them. The remaining fields of study included science ($N_{Control}$ = 11, $N_{Pressure}$ = 11), humanities ($N_{Control}$ = 14, $N_{Pressure}$ = 15), and law and economics ($N_{Control}$ = 6, $N_{Pressure}$ = 3). Participants received €12 as compensation for their participation. In addition, all participants assigned to the pressure group received an extra of €10 irrespective of whether they could improve their performance in the second block. The study was approved by the ethics committee of the University of Graz.

Table 10. Differences between the control and the pressure group regarding demographics and anxiety measures.

	Control Group	Pressure Group	Test statistic
Age *M* (SD)	22.84 (3.26)	23.48 (3.27)	$t(58)$ = -0.76, p = .448, d = -0.20
Females/Males	20/11	20/9	$\chi^2(1)$ = 0.01, p = .927, φ = 0.47
Test anxiety *M* (SD)	41.03 (11.31)	39.48 (8.62)	$t(58)$ = 0.59, p = .555, d = 0.15
General anxiety *M* (SD)	36.71 (11.06)	37.28 (9.12)	$t(58)$ = -0.22, p = .830, d = -0.06

Note. M = mean, SD = standard deviation, d = Cohen's d, φ = Phi effect size.

(a) Low demand problem

9 ≡ 2 (mod 4)		
1. Step:	*9 - 2 = 7*	
2. Step:	*7 ÷ 4 = 1*	
3. Step:	*Remainder: 3*	→ False

(b) High demand problem

62 ≡ 37 (mod 5)		
1. Step:	*62 - 37 = 25*	
2. Step:	*25 ÷ 5 = 5*	
3. Step:	*Remainder: 0*	→ True

Figure 11. Modular arithmetic problem with (a) low demand and (b) high demand with corresponding step-by-step solution.

4.2.2. Apparatus and Materials

4.2.2.1. Modular Arithmetic Task

The objective of modular arithmetic is to judge the truth value of problem statements such as "62 ≡ 37 (mod5)". The problem is solved by subtracting the second number from the first number (i.e., 62 - 37) and then dividing the resulting difference by the last number (i.e., 25 ÷ 5). If the dividend is a whole number (here, 5), the problem is true. If there is a remainder, the statement it false. Low demand problems consisted of two operands smaller than 10 and the first step (i.e., subtracting the second operand from the first operand) did not require to carry over. High demand problems consisted of two 2-digit numbers and the first step required to carry over. An example problem of each demand with step-by-step solution is depicted in Figure 11. Problems were presented in two blocks of 60 problems separated by a 2 min break. As can be seen in Figure 12, demand and truth (true vs. false) of the problems was balanced within each block. Problems were presented in a pseudorandomized order and each problem was used only once in the experiment. Participants pressed the left button of a number pad to indicate that a problem was false and the right button to indicate that a problem was true using the index fingers of their left and right hand respectively.
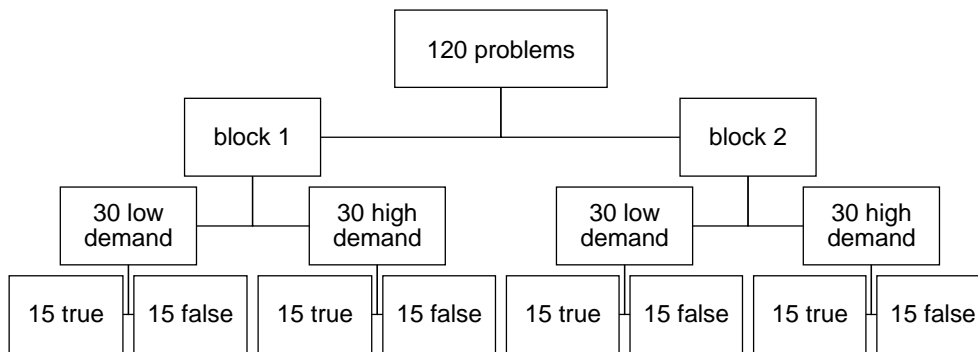
Figure 12. Overview of modular arithmetic problems presented in two blocks with demand and truth value (true or false) balanced within each block.

4.2.2.2. German Test Anxiety Inventory (PAF)

Test anxiety was assessed using the German Test Anxiety Inventory (Prüfungsangstfragebogen, PAF; Hodapp et al., 2011), which is described in more detail in the method section of Study 1. In the present study, the PAF was administered via an online form a day after the EEG recording.

4.2.2.3. State-Trait Anxiety Inventory (STAI)

State and trait anxiety were assessed using the State-Trait Anxiety Inventory (STAI; Spielberger, 1980), which is described in more detail in the method section of Study 2. In the present study, the trait scale of the STAI was administered via an online form a day after the EEG recording.

4.2.2.4. Worry scale

Performance-related worries were assed using a Faces Anxiety Scale (FAS), which is described in more detail in the method section of Study 2. The FAS was presented after every sixth arithmetic problem. Problems were presented in a pseudo-randomized list in which half of every 6th trial was a low demand problem and the other half a high demand problem. Participants were instructed to select the first face if they were not worried at all having solved the preceding problem correctly and sixth face if they were very worried.

4.2.2.5. Manipulation Check

After completing block 2, participants were given a number of questions to check whether the pressure manipulation was effective (see Beilock et al., 2004). Firstly, the state scale of the STAI was administered after the second block in each group assessing the level of state anxiety. Secondly, participants were asked to rate the second block regarding (a) how important they felt it was to perform at a high level, (b) how much pressure they felt to perform at a high level, (c) how highly they would rate their performance in the task. Items are described in more detail in the method section of Study 2.

4.2.2.6. EEG data acquisition

EEG was acquired through a BioSemi Active Two system (BioSemi, Amsterdam, The Netherlands) from 64 scalp electrodes placed according to the extended 10–20 system. An electrooculogram (EOG) was recorded from four additional electrodes; two placed horizontally at the outer canthi of both eyes, one placed above and one below the right eye. EEG and EOG signals were sampled at 256 Hz and filtered between DC and 128 Hz.

4.2.3. Procedure

The study was advertised as an EEG study aiming to better understand how a new math skill is acquired. Upon arrival, participants filled out a consent form and answered demographic questions. Then, participants were seated in the EEG cabin and EEG electrodes were placed. The recording started with a 3 min sequence in which participants were asked to deliberately execute eye movements followed by a resting state EEG, which are not within the scope of the present study and are therefore not reported in the following. Participants received task instructions on screen and were asked to answer as quickly and as accurately as possible in block 1. Then, six practice trials followed (three for each demand) for which feedback was provided (viz., "correct", "wrong", "too slow"). In addition, the response time was displayed for after each practice trial allowing participants to assess the maximum response interval of 30 s used in the experimental trials. At the end of the practice trials, participants were instructed on how to use the FAS and familiarized themselves with selecting a face by pressing the left and the right button of the number pad.

The first block consisted of 60 modular arithmetic problems and served as a pretest of arithmetic performance (for the balancing of task demand and truth value within each block, see Figure 12). As can be seen in Figure 13a, a trial started with a fixation period of 2 s, followed by the presentation of the problem until a response was given but for a maximum of 30 s. A blank screen was display between trials for 1 s. Every sixth trial, participants were

asked to rate their worries about having solved the preceding problem correctly using the FAS (see Figure 13b). In order to analyze the same number of worry ratings for each demand, problems were presented in a pseudo-randomized list in which half of every sixth trial was a low demand problem and the other half a high demand problem. In addition, the demand was balanced within each group of six problems to control for possible effects of the problems presented before the present problem.

The first second block consisted of 60 modular arithmetic problems and served as a posttest of arithmetic performance. While the control group was instructed to continue working through the problems as before, the pressure group was put into a high-pressure scenario. Firstly, outcome pressure was induced by informing participants that the computer had tracked their performance in block 1 and computed a performance score based on both the individual response time and accuracy. If they could manage to improve their score by 20% in the second block, they would be rewarded with an extra €10. In addition, participant received the information that all but one participant so far could manage to improve their performance. Secondly, monitoring pressure was induced by recording participant during the second block on videotape. Participants were instructed that this recording would be evaluated by a committee of teachers and scientists to better understand how modular arithmetic is learned. The experimenter then set up the video camera on a tripod directly to the right of participants, approximately 70 cm away.

At the end of the experiment, both groups filled out the state scale of the STAI and rated the importance, pressure, and success in the second block as a manipulation check. Participants were then debriefed about the true purpose of the study and those in the pressure group were informed that no videotape was recorded. Finally, participants received a code to use on the next day to fill out online questionnaires assessing test anxiety and general anxiety.

(a) Trial

2 s

until response
max. 30 s

#

$7 \equiv 3 \ (mod2)$

1 s

(b) Every 6th trial

2 s

until response
max. 30 s

#

$42 \equiv 23 \ (mod7)$

until response

1 s
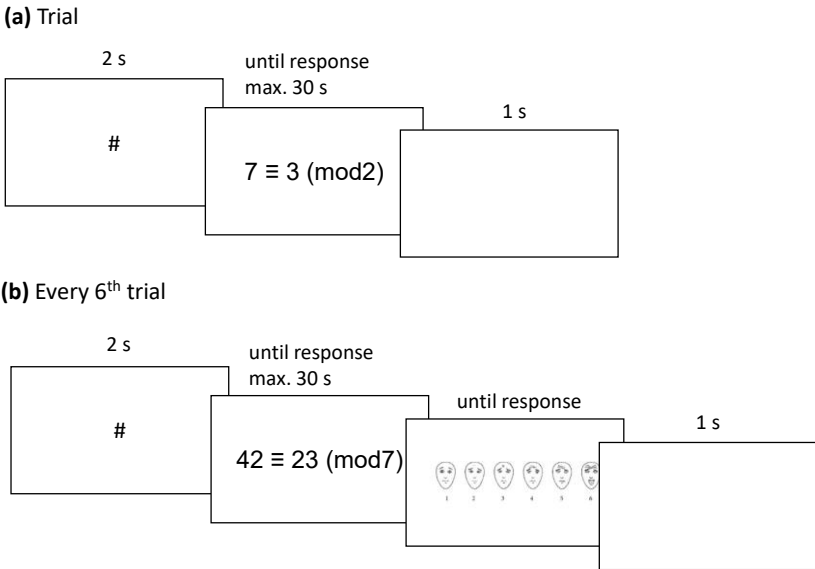
Figure 13. (a) Schematic display of a trial with a low demand problem. (b) Every sixth trial, participants rated their worries using an FAS (here, a high demand problem is displayed).

4.2.4. Data Analysis

EEG data were analyzed in the MATLAB environment (Mathworks, Inc., Natick, MA, USA, release 2014b) using EEGLAB 14.1.1 (Delorme & Makeig, 2004) for the pre-processing and Fieldtrip (version number 20171010; Oostenveld, Fries, Maris, & Schoffelen, 2011) to run cluster-based permutation tests. Further statistical analyses for both behavioral and electrophysiological data were carried out with the software SPSS 22.0 (IBM Corp. Released, 2011) using the PROCESS package for the mediation analyses (A. Hayes, 2013).

4.2.4.1. Behavioral Data

Whether participants in the pressure group perceived block 2 as more stressful than participants in the control group was checked by independent t-tests with the importance, pressure and success rating as well as the state scale of STAI as dependent variables.

Behavioral performance was analyzed by 2 (block) x 2 (demand) x 2 (group) ANOVAs with the dependent variable mean response time (in s) and mean accuracy (in percentage). The change in performance between blocks was then correlated with individual differences in test anxiety. For this, the difference in response time (RT_{block2} - RT_{block1} = ΔRT) and accuracy (AC_{block2} - AC_{block1} = ΔAC) was computed for low and high demand problems separately and correlated with test anxiety in each group.

Self-reported worries were averaged and analyzed with a 2 (block) x 2 (demand) x 2 (group) ANOVA. Again, the change in worries ($worries_{block2}$ - $worries_{block1}$ = $\Delta worries$) was correlated with test anxiety in each group.

4.2.4.2. Electrophysiological Data

The EEG signal was re-referenced to the mean (common average reference), the baseline (channel mean) was removed, and a 1 Hz high-pass filter was applied. Eye movement artifacts were corrected by means of independent component analysis (ICA). For this, non-stereotyped noise was first removed from the continuous signal by visual inspection. Then, the ICA was run using the EEGLAB routine *runica* and independent components reflecting vertical and horizontal eye movements were removed for each block and subject. The signal was then again visually inspected and remaining artifacts were manually removed from the continuous signal. Finally, bad channels were defined based on the joint kurtosis of the recorded electrodes. Channels with a kurtosis deviating more than 5 *SD* from the mean were removed and interpolated.

We then explored the event-related (de-)synchronization (ERS/ERD) (Pfurtscheller & Aranibar, 1977; Pfurtscheller & Lopes da Silva, 1999) in the theta frequency band (4-8 Hz) across the scalp. The theta frequency band was defined from 4 to 8 Hz based on the review by Hsieh and Ranganath (2014). ERS/ERD values were derived as follows: the EEG signal was band pass filtered within the theta band and then squared to obtain the power. The power was averaged across time and trials for the reference (R) and the activation (A) interval. A 1.75 s interval from 125 ms after fixation onset to 125 ms before stimulus onset served as baseline (see Figure 13). The reference interval was defined as the time from stimulus onset until a response was given. The last 20 ms of the activation interval were discarded to take temporal inaccuracy of measuring the time stamp of the button press into account. The ERS/ERD was calculated by the following formula (see Pfurtscheller & Aranibar, 1977):

$$\%ERS|ERD = \frac{(A - R)}{R} * 100$$

If the band power is increasing in the activation interval as compared to reference interval, a positive value will result, which is referred to as event-related synchronization (ERS). If the band power is decreasing, a negative value will result, which is referred to as event-related desynchronization (ERD). The ERS/ERD scores were then compared between groups for each demand during block 2 using cluster-based permutation tests (Maris & Oostenveld, 2007). For this, an independent t-test was computed for each electrode (critical alpha value = .05, uncorrected) and the probability of a cluster was determined by using a Monte-Carlo simulation with 10,000 random permutations (critical alpha value = .05, corrected; minimum cluster size = 3). To rule out that the two groups already differed in their pretest neural activity, we also compared the two groups for each demand during block 1 using the same cluster-based permutation procedure.

We predicted that the pressure group exhibit a stronger ERS in the theta frequency band than the control group at frontal midline recording sites. Based on previous studies (e.g., Ishihara & Yoshii, 1972; Luu et al., 2004; Osinsky et al., 2017; Wang et al., 2015), the following frontal-central recording sites were considered as reflecting FMΘ: Fz, FCz, and Cz. FMΘ was then defined at those frontal midline electrodes that were revealed by the cluster-based permutation tests to differ significantly between groups for both demands. Then, the differences in FMΘ between blocks (FMΘ$_{block2}$ - FMΘ$_{block1}$ = ΔFMΘ) was correlated with participants test anxiety scores.

To explore whether the predicted increase of FMΘ in test-anxious students in the pressure group was due to an increase in worries, mediation analyses were used. For this, the change between blocks was averaged across demands for FMΘ (ΔFMΘ$_{AVG}$) and worries

(ΔWorries$_{AVG}$). In each group, a mediation model was tested with test anxiety as the predictor and ΔFMΘ_{AVG} as criterion variable including ΔWorries$_{AVG}$ as mediating variable. Indirect effects were tested for significance using bias-corrected bootstrap confidence intervals based on 10,000 bootstrap samples (see A. Hayes, 2013).

4.3. Results

4.3.1. Manipulation Check

There was no significant difference between groups regarding the rating of performance pressure or state anxiety (see Table 11). Note the means tended towards the predicted direction, i.e., higher values in the pressure than in the control group but failed statistical significance. This is in contrast to previous studies that have reported significant differences in these measures using a similar pressure manipulation (Beilock & DeCaro, 2007; Beilock et al., 2004; Ramirez & Beilock, 2011). To further see whether the perceived pressure was related to individual differences in test anxiety, we correlated the rating of the pressure item with test anxiety scores. In the pressure group, students with higher levels of test anxiety reported more pressure regarding block 2 than lower test anxious students, $r = .47$, $p = .009$. However, a similar relationship was also found in the control group, $r = .44$, $p = .014$. Thus, it appears that higher test-anxious students in both groups perceived the second block of modular arithmetic as relatively stressful as compare to students with lower test anxiety.

Table 11. Means and standard deviations (in brackets) of the manipulation check for each group.

	Control Group	Pressure Group	Test statistics
Importance R.	5.65 (1.05)	5.72 (0.88)	$t(58) = -0.31$, $p = .754$, $d = -0.07$
Pressure R.	4.00 (1.34)	4.21 (1.37)	$t(58) = -0.59$, $p = .557$, $d = -0.16$
Success R.	4.65 (1.33)	4.97 (1.01)	$t(58) = -1.04$, $p = .302$, $d = -0.27$
State Anxiety	33.58 (9.07)	36.41 (8.74)	$t(58) = -1.23$, $p = .223$, $d = -0.32$

Note. R. = rating; d = Cohen's d.

Table 12. Result summary of Analyses of Variance (ANOVAs) with (a) response time (in ms) and (b) accuracy (in %) as dependent variable.

	$F(1,58)$	p	η_p^2
(a) Response time			
Main effects			
Block	79.47	< .001	.578
Demand	276.54	< .001	.827
Group	0.46	.502	.008
Interactions			
Block x Group	8.70	.004	.130
Block x Demand	40.11	< .001	.409
Demand x Group	0.13	.718	.002
Block x Demand x Group	3.49	.066	.057
(b) Accuracy			
Main effects			
Block	2.33	.132	.039
Demand	40.38	< .001	.410
Group	< 0.01	.970	<.001
Interactions			
Block x Group	1.14	.291	.019
Block x Demand	0.16	.899	<.001
Demand x Group	0.64	.428	.011
Block x Demand x Group	0.79	.379	.013

Note. η_p^2 = Partial Eta Squared.

4.3.2. Behavioral Performance

Statistical analyses of the behavioral performance are summarized in Table 12. There was a significant main effect of *Block* as well as of *Demand* on the response time. As can be seen in Figure 14a, the response time decreased between block 1 and block 2 and problems with high demand were solved slower than problems with low demand. Furthermore, there was a significant interaction *Block* x *Demand*, with high demand problems showing a stronger decrease in response time between blocks than low demand problems. Finally, the factors *Block* and *Group* showed a significant interaction with participants assigned to the pressure group having a more pronounced decrease in the response time between blocks than those assigned to the control group. While the response time in block 2 was highly similar between groups, the pressure group appeared to respond slightly slower in block 1 than the control group (see Figure 14a). At this stage of the experiment, however, there should have been no systematic difference between groups. Therefore, it remains unclear whether the interaction effect *Block* x *Group* is due to the performance pressure or to the difference between groups

in pretest response times. Finally, the decrease in the response time between blocks in the pressure group seemed to be more pronounced for high demand problems than for low demand problems as indicated by a three-way interaction *Block x Demand x Group* with a trend towards significance (see Figure 14a)*.*

Regarding the accuracy, there was a significant main effect *Demand*, with low demand problems being solved more accurately than high demand problems (see Figure 14b). Contrary to expectations, there was no significant interaction *Block x Group* or a significant three-way interaction *Block x Demand x Group*.

The correlation between the changes in performance between blocks and test anxiety are summarized in Table 13. Neither the change in response time (ΔRT) nor the change in accuracy (ΔAC) was significantly related to test anxiety in either of the two groups. Explorative analyses correlating test anxiety with the dependent measures (i.e., response time, accuracy, worry, FMΘ) in each block separately can be found in Table S 4.

Taken together, there was a significant effect of demand with high demand problems being solved slower and less accurately than problems with low demand. Moreover, a learning effect was present as indicated by decreasing response times between block 1 and block 2. This effect was more pronounced for problems with high demand than for problems with low demand. Contrary to our hypothesis, participants assigned to the pressure group did not show a decrease in the accuracy of solving high demand problems in block 2. Furthermore, there was no significant relationship between individual test anxiety and changes in performance between blocks in either group.
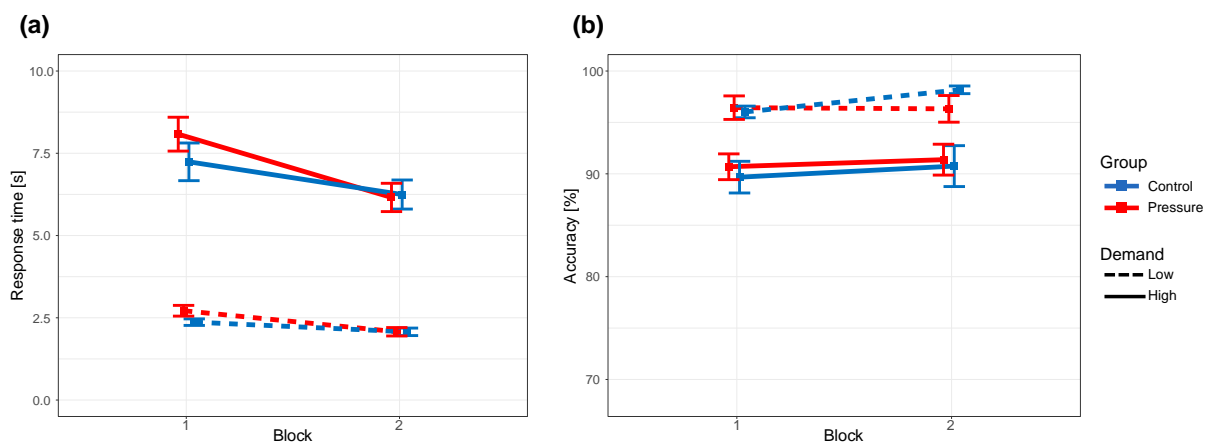
Figure 14. (a) Mean response time and (b) accuracy as a function of block, demand, and group. Error bars indicate +/- 1 standard error.

Table 13. Pearson correlations between test anxiety and the changes in response time (ΔRT), accuracy (ΔAC), worry (ΔWorries), and FMΘ (ΔFMΘ) between blocks.

	ΔRT		ΔAC		ΔWorries		ΔFMΘ	
Demand:	Low	High	Low	High	Low	High	Low	High
Control Group	-.19	-.32	.12	-.19	-.13	.05	.23	.14
Pressure Group	-.19	.17	-.09	.01	.09	.04	.57**	.48**

Note. * = p < .05; ** = p < .01; *** = p < .001.

Table 14. Descriptive statistics of the worry rating during each block for each demand and group.

		Block 1		Block 2	
		Low Demand	High Demand	Low Demand	High Demand
Control Group	M (SD):	1.57 (0.78)	1.81 (0.66)	1.61 (0.85)	1.80 (0.73)
	[Min, Max]:	[1.00, 4.20]	[1.00, 3.60]	[1.00, 4.80]	[1.00, 4.20]
Pressure Group	M (SD):	1.68 (0.75)	2.00 (0.78)	1.70 (0.72)	1.98 (0.73)
	[Min, Max]:	[1.00, 3.20]	[1.00, 4.00]	[1.00, 3.60]	[1.00, 3.60]

4.3.3. Worry Rating

Means and standard deviations of the worry rating are summarized in Table 8. Since the means appeared to be rather low also the minimum and maximum are provided. Participants reported significantly higher worries about solving high demand problems than about solving low demand problems. However, there was no significant interaction *Block* x *Group* or a significant three-way interaction *Block* x *Demand* x *Group* (for a summary of the statistics, see Table 15). As can be seen in Table 13, the change in worries between blocks was not significantly related to test anxiety in either group.

Table 15. Result summary of Analysis of Variance (ANOVA) with worry as dependent variable.

	$F(1,58)$	p	η_p^2
Main Effects			
Block	0.02	.888	<.001
Demand	18.28	<.001	.240
Group	0.66	.420	.011
Interactions			
Block x Group	0.02	.888	<.001
Block x Demand	0.33	.566	.006
Demand x Group	0.61	.438	.010
Block x Demand x Group	< 0.01	.980	<.001

Note. η_p^2 = Partial Eta Squared.

Block 2: [Pressure - Control]
theta band (4–8 Hz)

(a) Low Demand Problems   (b) High Demand Problems
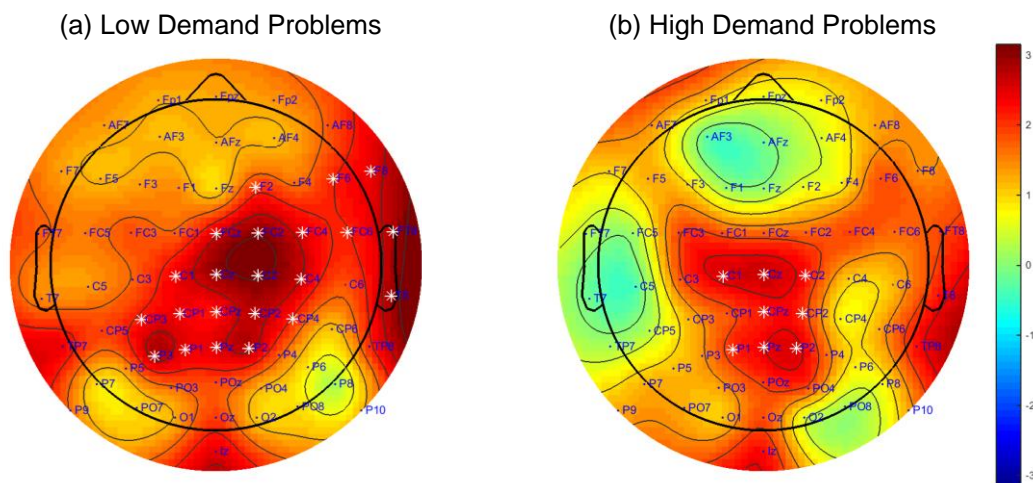
Figure 15. Topographical maps of t-values corresponding to the contrast [Pressure - Control] in block 2 for (a) low demand and (b) high demand problems with data filtered in the theta frequency band (4–8 Hz). Analyses revealed a significant cluster (p < .05, corrected) for each demand (electrodes within a cluster are marked with white asterisks). Note that there were no significant differences between groups in block 2 (see Figure S 4).

4.3.4. FMΘ under Performance Pressure

There was no significant difference between groups in the theta activity at block 1 for either task demand (see Figure S 4). In contrast, theta appeared to be increased in the pressure group across the scalp during block 2, as indicated by almost entirely positive t values (see Figure 15). The cluster-based permutation test revealed a significant cluster (cluster p <

.001) for low demand problems (i.e., $ERS_{Pressure} > ERS_{Control}$) in the theta frequency band ranging from right temporal to left parietal electrodes. The cluster peaked at fronto-central recording sites, including electrodes FCz and Cz. For high demand problems, the cluster-based permutation test revealed a significant but more focal cluster (cluster p = .036). This cluster ranged from fronto-central to fronto-parietal electrodes. This cluster showed a local peak at fronto-central recording sites, including electrode Cz. The topographical maps of the raw ERS/ERD values for each group and demand can be found in the supplementary material (block 1: Figure S 3; block 2: Figure S 4). Since the cluster of both task demands showed a peak around electrode Cz individual FMΘ was extracted from this electrode for the following analyses. This is in line with previous research examining FMΘ at electrode Cz (Cavanagh & Frank, 2014; Ishihara & Yoshii, 1972; Jacobs, Hwang, Curran, & Kahana, 2006; Liu, Woltering, & Lewis, 2014; Massar, Rossi, Schutter, & Kenemans, 2012; Missonnier et al., 2006; Neo et al., 2011; Wang et al., 2015).

Figure 16a depicts the mean FMΘ as a function of block, demand, and group. A 2 (block) x 2 (demand) x 2 (group) ANOVA showed no significant main effect *Block*, $F(1,58)$ = 0.89, p = .766, η_p^2 = .002, but a significant main effect *Group*, $F(1,58)$ = 4.41, p = .039, η_p^2 = .071, qualified by a significant interaction *Block* x *Group*, $F(1,58)$ = 6.88, p = .011, η_p^2 = .106. These results confirm the cluster-based permutation tests (see Figure 15, Figure S 4). Moreover, there was a significant main effect *Demand*, $F(1,58)$ = 11.56, p = .001, η_p^2 = .166, with low demand problems showing higher FMΘ than high demand problems. The interaction *Block* x *Demand*, $F(1,58)$ = 0.76, p = .784, η_p^2 = .001, *Demand* x *Group*, $F(1,58)$ = 0.79, p = .780, η_p^2 = .001, and the three-way interaction *Block* x *Demand* x *Group* were not significant, $F(1,58)$ = 2.56, p = .115, η_p^2 = .042. Taken together, results demonstrate that FMΘ was increased by performance pressure at fronto-central recording sites for both demands.

The relationship between the change in FMΘ between blocks (ΔFMΘ) and test anxiety is depicted in Figure 16b. Correlation analyses confirmed that the individual level of test anxiety was significantly related to ΔFMΘ for low demand (p = .001) and high demand problems (p = .008) in the pressure group (see Table 13). Note that both correlations remained significant when removing data of participant 35, which could be classified as an outlier based on Figure 16b. In the control group, in contrast, there was no significant correlation between ΔFMΘ, neither for low demand problems (p = .222) nor for high demand problems (p = .454).
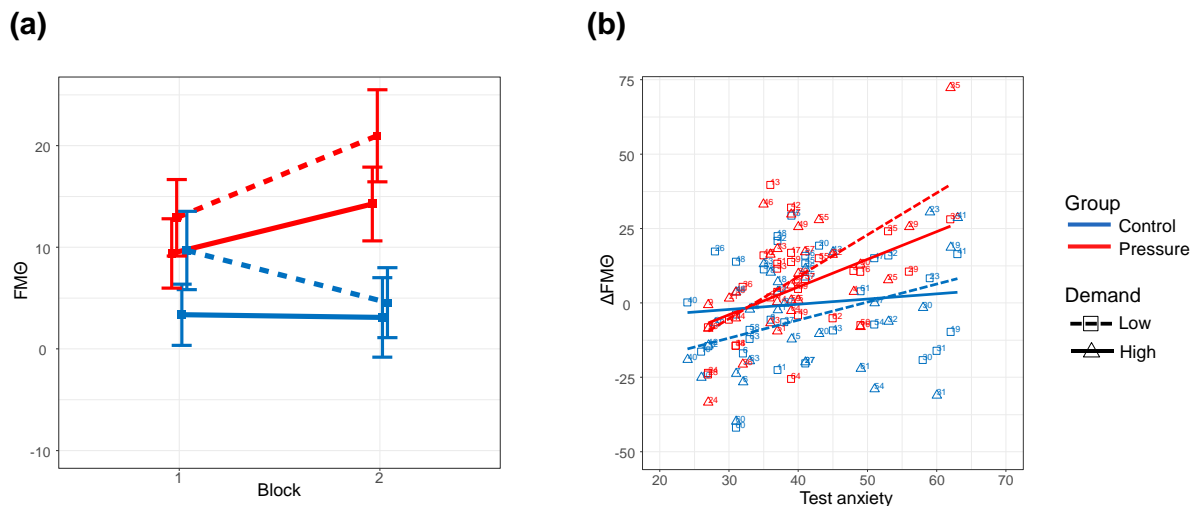
Figure 16. (a) Frontal midline theta (FMΘ) at electrode Cz as a function of block, demand, and group. Error bars indicate +/- 1 standard error. (b) Relationship of the change in FMΘ between blocks (ΔFMΘ) and test anxiety. Numbers indicate individual subjects.

4.3.5. FMΘ and Worries

Mediation models predicting the average change in FMΘ between blocks ($\Delta FM\Theta_{AVG}$) based on test anxiety for both groups are depicted in Figure 17. The average change in worries between blocks ($\Delta Worries_{AVG}$) was entered as mediating variable. On a bivariate level, test anxiety was not significantly related to $\Delta FM\Theta_{AVG}$ in the control group ($c = .21$, $p = .247$). There was no mediating effect of $\Delta Worries_{AVG}$ given that the 95% confidence interval did contain zero ($ab = .02$, $CI95 = -0.208$ to 0.129). Taking the change in worries between blocks into account, the relationship between test anxiety and $\Delta FM\Theta_{AVG}$ remained insignificant ($c' = .19$, $p = .273$). In the pressure group, test anxiety was significantly related to $\Delta FM\Theta_{AVG}$ on a bivariate level ($c = .61$, $p < .001$). However, this relationship was not significantly mediated by a change in worries between blocks ($ab = .01$, $CI95 = -0.064$ to 0.120). The direct effect of test anxiety on $\Delta FM\Theta_{AVG}$ remained highly significant ($c' = .60$, $p < .001$) when controlling for the change in worries. Thus, the increase in FMΘ during block 2 in test-anxious students under performance pressure could not be explained by an increase in worry in these students. Standard errors and model summary for all paths of the mediation analyses can be found in Table S 5.

Post-hoc analyses were conducted to test whether worries were related to the accuracy in the modular arithmetic task. For this, we correlated the $\Delta Worries_{AVG}$ with the average change in accuracy between blocks (i.e., ΔAC_{AVG}) for each group. Results revealed that participants who committed more errors in block 2 than in block 1 also reported more worries in block 2 relative to block 1 in both the control group, $r = -.45$, $p = .011$, and the pressure group, $r = -.49$, $p = .007$.
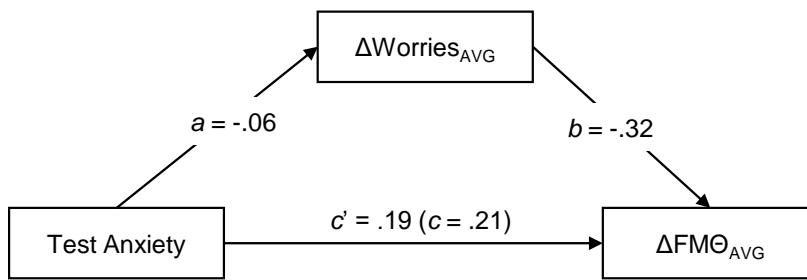
(a) Control group
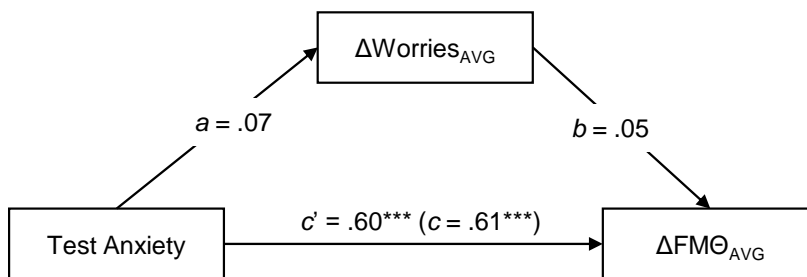
Figure 17. Mediation model predicting the average change in FMΘ between blocks ($\Delta FM\Theta_{AVG}$) based on differences in test anxiety in (a) the control and (b) the pressure group. The average change in worries between blocks ($\Delta Worries_{AVG}$) was entered as mediator variable. Decimals are the standardized regression coefficients. * = p < .05; ** = p < .01; *** = p < .001.

(b) Pressure group

4.4. Discussion

The aim of the present study was (a) to test whether test anxiety is related to the degree of *choking under pressure* and (b) to directly assess the working memory costs of performance pressure by measuring oscillatory brain responses during the task. To this end, we adapted the paradigm reported by Beilock and colleagues (2004) for an EEG study. Participants were randomly assigned to either a control group or a pressure group prior to performing two blocks of modular arithmetic with half of the problems having a low demand and the other half having a high task demand. The first block served as a pretest and the second block as a posttest of arithmetic performance. Between blocks, the two groups received different instructions. While the control group was informed that they would be performing another set of problems, the pressure group was exposed to a high pressure test scenario. Participants in both groups were asked to rate their worries about performing well at regular intervals throughout the experiment using an FAS (Bieri et al., 1990; Punaro & Reeve, 2012; Trezise & Reeve, 2014b). Working memory costs of performance pressure were assessed by analyzing the ERS/ERD in the theta frequency band (4-8 Hz) at frontal midline recording sites (FMΘ) during the task (Cavanagh & Frank, 2014; Cavanagh & Shackman, 2015; Hsieh & Ranganath, 2014; Inanaga, 1998; Klimesch, 1999; Pfurtscheller & Aranibar, 1977; Pfurtscheller & Lopes da Silva, 1999; Roux & Uhlhaas, 2014; Sauseng et al., 2010).

81

4.4.1. Manipulation Check

The manipulation check revealed that both groups perceived the second block of modular arithmetic as highly important. More specifically, the mean importance ratings were within the upper range of the 7-Point Likert scale with $M = 5.65$ in the control group and $M = 5.72$ in the pressure group. As predicted, there was no significant difference between groups in their rating of importance (see also Beilock et al., 2004). That the importance ratings were high across groups is an important indicator that participants in both groups were motivated to perform well in the second block. Differences between groups are therefore unlikely to result from a lower motivation of participants in the control group. Regarding the perceived success in the modular arithmetic task, ratings were in the average range but did not differ significantly between groups. Contrary to expectation, the rating of performance pressure was not significantly higher in the pressure than in the control group. Pressure ratings of both the control ($M = 4.00$) and the pressure group ($M = 4.21$) were more similar to the pressure rating reported by Beilock and colleagues (2004) for the control group ($M = 3.95$) than those reported for the pressure group ($M = 5.08$). Finally, the state anxiety in the present study appeared to be higher in the pressure ($M = 36.41$) than in the control group ($M = 33.58$), but this difference also failed statistical significance. The state anxiety in the control group was similar to those reported by Beilock and colleagues (2004) for the control group ($M = 32.08$). However, the pressure group showed lower state anxiety that the previously reported pressure group ($M = 42.68$). Taken together, the manipulation check could not clearly confirm that the pressure manipulation was effective in the present study. This is in contrast to previous studies that have reported significant differences in these measures using a similar pressure scenario (Beilock & DeCaro, 2007; Beilock et al., 2004; Ramirez & Beilock, 2011).

One reason for the lack of significant differences in the manipulation check might have been that we modified the pressure manipulation in the present study as compared to previous studies (Beilock & DeCaro, 2007; Beilock et al., 2004; Ramirez & Beilock, 2011). More specifically, previous studies instructed participants in the pressure group that they had been paired with another participant and that the reward of both would depend upon the performance of the present participant. This seemed implausible to us given that the present participants could be expected to be familiar with psychological testing. Instead, participants were informed that of the previous participants all but one could manage to improve their performance. This difference in the pressure manipulation might explain why we failed to find significant differences in the manipulation check in the present study. Note, however, that we found a significant difference in Study 2 regarding the pressure rating between pressure conditions. In Study 2, we did not provide information about the performance of other participants at all as part of the pressure manipulation. Therefore, it seems unlikely that changing the cover story

regarding the team effort can account for the lack of significant differences in the present manipulation check. Another explanation might be that participants in the present sample were less sensitive to the pressure manipulation resulting in only subtle differences in the manipulation check. In line with this explanation is that the pattern of results was directionally as predicted, with the pressure group reporting higher pressure and state anxiety than the control group. Although speculative, this might indicate cultural difference between studies. While the studies by Beilock and colleagues (Beilock & DeCaro, 2007; Beilock et al., 2004; Ramirez & Beilock, 2011) were conducted with US-American students, the present study included Austrian and German students. Future studies would be needed to address potential cultural differences in responding to performance pressure as manipulated in the present study.

Post-hoc analyses were conducted to explore whether participants differed in their perception of performance pressure depending on their levels of test anxiety. In both groups, higher test-anxious students reported more performance pressure than lower test-anxious students. Thus, it appears that test-anxious students in both groups perceived the second block of modular arithmetic to some degree as an evaluative situation. In fact, this might have further contributed in diluting an already subtle difference in the pressure rating between the control group and the pressure group.

Taken together, while the effectiveness of the pressure manipulation is called into question, students with higher test anxiety generally perceived the second block of modular arithmetic as relatively stressful as compared to students with lower test anxiety.

4.4.2. Behavioral Performance

On a behavioral level, the present findings revealed a main effect of demand, with high demand problems being solved slower and less accurately than problems with low demand. Furthermore, a learning effect was present as indicated by decreasing response times between block 1 and block 2. This effect was more pronounced for problems with high demand than for problems with low demand. Contrary to our hypothesis, participants assigned to the pressure group did not show a decrease in the accuracy of solving high demand problems in block 2. Instead, the accuracy did not differ significantly between blocks for low or high demand problems in either group. This is in sharp contrast to previous studies showing a decrease in the performance of problems with high demand in the pressure group (Beilock & Carr, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004, 2007; Boere et al., 2016; Ramirez & Beilock, 2011). For instance, Beilock and colleagues (2004) reported that the accuracy for solving high demand problems was at about 80% in the pretest in both the control and the pressure group. In the posttest, this accuracy further increased in the control group but decreased by over 5%

in the pressure group. In the present study, in contrast, the accuracies of solving high demand problems were already at about 90% for both groups in block 1 (see Figure 14). The accuracy rate did not change significantly between blocks in either group.

However, previous studies reporting performance impairments in the pressure group also found significant differences between groups in the manipulation check (with the exception of Boere et al., 2016, who did not report a manipulation check). This might offer an explanation of why we failed to replicate the *choking under pressure effect* in the present study. Given that block 2 was not perceived as more stressful by the pressure group than by the control group, participants in the pressure group can also not be expected to show behavioral impairments due to performance pressure. An explanation for the lack of both differences in the manipulation check and performance impairments could be that we increased the number of trials in the present study in order to improve the signal-to-noise ratio for the EEG analysis. While Beilock and colleagues (2004) compared two blocks à 24 problems (with a middle block of 24 problems skipped in the analysis), we compared two blocks à 60 problems. Thus, participants had more practice in the present study possibly allowing them to achieve already a high level of accuracy in the first block. In the second block, participants therefore might have been skilled enough to cope with the detrimental effect of performance pressure and consequently report less pressure. However, another experiment reported by Beilock and colleagues (2004) does not support this explanation. In this experiment, the performance of solving high demand problems which were previously either infrequently (i.e., one or two times) or frequently (i.e., 50 repetitions) practiced was compared under performance pressure. Results revealed that those problems which were frequently practiced were unimpaired by performance pressure. However, performance pressure still harmed the accuracy with which infrequently practiced problems were solved. This suggest that while practicing a problem until the answer can be retrieved from memory neutralizes the effect of performance pressure, general practice on the solution algorithm alone cannot shield against choking. In the present study, we used a unique set of problems in each block. It therefore seems unlikely that the additional practice on the solution algorithm provided by increasing the number of trials can account for the present lack of significant behavioral effects. However, the overall high accuracy in the modular arithmetic task is indicating that the task was relatively easy for the present participants. Future studies should therefore consider to increase the difficulty of the task, for instance, by lowering the maximum response interval.

It should be noted that one other published study also failed to find a significant effect of performance pressure on the accuracy in modular arithmetic. In this study (Benny & Banks, 2015), students were asked to perform two blocks of modular arithmetic à 24 problems. While the first block served as a pretest of arithmetic performance, all participants were exposed to a high pressure scenario, as reported by Beilock and colleagues (2004), in the second block.

At the end of each block, participants were asked to fill out the state scale of the STAI. Similar to the present study, participants did neither report more state anxiety nor exhibited performance impairments in the second block as compared to the first block.

Finally, changes between blocks in both the response time and the accuracy were not significantly related to individual differences in test anxiety. This was hypothesized based on previous research (for reviews, see Calvo & Eysenck, 1992; Eysenck et al., 2007; Moran, 2016; Wine, 1971) according to which test-anxious students are especially prone to be distractible by performance-related worries. However, given that there was no significant difference in the accuracy between groups it is not surprising that we did not find a correlation between test anxiety and the degree of choking within the pressure group.

In summary, results revealed a higher performance for low demand than for high demand problems and a learning effect with decreasing response times between blocks. Contrary to expectations, participants assigned to the pressure group showed no performance impairments in the present study and changes in the performance between blocks were not related to individual differences in test anxiety.

4.4.3. Performance Pressure and Worries

Similar to the previous findings, participants in the pressure group did not report more worries during the second block of modular arithmetic than the control group. In general, the worry ratings in the present study were within the lower range of the 6-Point FAS (see Table 14). Worry ratings were also low in comparison to Study 2 in which we reported ratings in the average range of the scale. Furthermore, there was no significant correlation between worries and individual differences in test anxiety. Given that the overall performance in the present study was high across conditions and groups, participants might have had little reason to be worried about their performance.

4.4.4. Increased FMΘ Under Performance Pressure

In the EEG analysis, we first compared the ERS/ERD in the theta frequency band across the scalp using cluster-based permutation tests. There was no significant difference in theta activity between groups at block 1 for either task demand (see Figure S 4). This was important to rule out that the two groups already differed in their pretest theta activity. In contrast, theta appeared to be increased in the pressure group across the scalp during block 2, as indicated by almost entirely positive t values (see Figure 15). The cluster-based permutation tests revealed a significant cluster for each task demand. For low demand problems, the cluster ranged from right temporal to left parietal electrodes peaking at fronto-

central recording sites, including FCz and Cz. The cluster for high demand problems was more focal ranging from fronto-central to fronto-parietal electrodes with a local peak at fronto-central recording sites, including Cz. Since the cluster of both task demands showed a peak around electrode Cz individual FMΘ was extracted from this electrode for the following analyses. This is in line with previous research reporting FMΘ at electrode Cz (Cavanagh & Frank, 2014; Ishihara & Yoshii, 1972; Jacobs et al., 2006; Liu et al., 2014; Massar et al., 2012; Missonnier et al., 2006; Neo et al., 2011; Wang et al., 2015). For instance, in one of the first studies on FMΘ, Ishihara and Yoshii (1972) defined FMΘ at the Fz-Cz lead. In another study, Liu, Woltering, and Lewis (2014) showed that the theta activity was most pronounced at Cz during a Go/No-go task in children. Extracting individual FMΘ from Cz was further demonstrated to be inversely related to both the age and the control ability of the children. However, it should be noted that FMΘ has also been commonly reported at more frontal recording sites, especially at Fz (for a review, see Hsieh & Ranganath, 2014). An ANOVA confirmed that FMΘ at Cz did not differ between groups at block 1 and was increased at block 2 for the pressure group as compared to the control group. Moreover, theta activity was higher for low demand problems than for high demand problems across blocks and groups. This finding is in line with studies by Grabner and De Smedt (De Smedt, Grabner, & Studer, 2009; Grabner & De Smedt, 2011, 2012) showing that arithmetic problems that are retrieved from memory, display more theta activity than problems that are solved by using an arithmetic procedure. In the present study, the answer to the first step (e.g., 9 - 2) of solving low demand problems (e.g., "9 ≡ 2 (mod4)") can be expected to be retrieved from memory (see Siegler, 1988). In contrast, the answer to the first step (e.g., 62 - 37) of solving high demand problems (e.g., "62 ≡ 37 (mod5)") are likely to require an arithmetic procedure.

We then correlated the change in FMΘ between blocks (ΔFMΘ) and individual test anxiety in each group. Test anxiety was significantly related to ΔFMΘ for low demand and high demand problems in the pressure group (see Table 13). In the control group, in contrast, there was no significant correlation between ΔFMΘ and test anxiety for either task demand. Finally, mediation analyses were used to test whether the increase of FMΘ for test-anxious students in the pressure group was mediated by an increase of worries between blocks. The analyses revealed that there was no significant mediating effect of worries in explaining the relationship between test anxiety and ΔFMΘ. Instead, the change of worries seemed to be directly related to the change in accuracy, with participants of both groups who committed more errors in block 2 than in block 1 also reporting more worries in block 2 relative to block 1.

Taken together, the EEG analyses revealed that the theta activity was increased by performance pressure at frontal-midline recording sites, especially in students with higher levels of test anxiety. The findings corroborate recent studies that have addressed the effect of anxiety as well as situational stress on FMΘ in the time-frequency domain (see Osinsky et

al., 2017). Regarding situational stress, Osinsky and colleagues (2017) could demonstrate that FMϴ (recorded at electrode Fz) was increased in participants performing a simple two-choice task under threat anticipation as compared to a control group. However, this relationship was reported to only hold true for female but not for male participants. In contrast, we see an increase of FMϴ by performance pressure in the present study across female and male students. Adding gender as a factor in the ANOVA with FMϴ at electrode Cz as dependent variable did not change the effect of pressure. One difference between the two studies is that participants in the pressure group of the previous studies anticipated to be evaluated (i.e., to give a speech after the EEG recording), while they were evaluated during the task in the present study. This might offer an explanation of why we found a more robust difference in the theta activity between groups including both female and male students. Nevertheless, both studies provide evidence for the notion that performance pressure is inducing higher FMϴ while performing a cognitive task. Moreover, previous studies have shown that individuals with higher anxiety-related symptoms exhibit higher FMϴ than individuals without these symptoms (Balconi & Crivelli, 2010; Mueller et al., 2014, 2015; Neo et al., 2011). For instance, Mueller and colleagues (2014) have reported that FMϴ (recorded from electrode FCz) is increased in individuals with higher neuroticism following negative feedback. The present results add to this body of evidence by demonstrating that higher test-anxious students show a stronger increase in FMϴ by performance pressure than lower test-anxious students.

The modulation of FMϴ by both situational performance pressure and individual differences in test anxiety is in line with the *adaptive control hypothesis* by Cavanagh and Shackman (2015). According to this hypothesis, increased FMϴ in anxious individuals reflects the need for attentional control regarding uncertain or threatening events (Cavanagh & Shackman, 2015). The *adaptive control hypothesis* was formulated based on findings that ERPs following uncertain events such as errors (ERN), punishment (FRN), and conflict-inducing stimuli (N2) originate from a transient oscillatory synchronization of FMϴ (for reviews, see Cavanagh & Frank, 2014; Cavanagh & Shackman, 2015). Against this backdrop, the present findings suggest that test-anxious students increased their attentional control in the pressure group. This might have enabled them to maintain their performance in solving the modular arithmetic problems. In fact, this offers another potential explanation of why we did not see performance impairments in the present study. By increasing their attentional control (e.g. "by focusing harder on the task"), test-anxious students might have been able to compensate for the detrimental effect of performance pressure and could maintain a high performance. Future studies are needed to better understand the cognitive processes that give rise to the FMϴ in test-anxious students who are being evaluated. Such studies should consider to implement a more direct indicator of attentional control to support the present explanation of higher FMϴ reflecting a compensatory increase in attentional control. For

instance, measuring the pupil dilation during the task could provide complementary information about the dynamics of attentional control in test-anxious students (for a review, see Beatty, 1982).

4.4.5. Conclusion

The present study failed to replicate *choking under pressure* in mathematical problem solving (Beilock & Carr, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004, 2007; Boere et al., 2016; Ramirez & Beilock, 2011). Participants assigned to the pressure group did not report more pressure and performed at the same level as participants assigned to the control group. Furthermore, the performance in the modular arithmetic task was not related to individual differences in test anxiety and self-reported worries were low across groups. In contrast, the study revealed significant differences between groups in the oscillatory brain responses during the task. The theta activity was increased by performance pressure at frontal-midline recording sites, especially in students with higher levels of test anxiety. Taken together, results are in line with the *adaptive control hypothesis* (Cavanagh & Shackman, 2015), suggesting that test-anxious students increase their attentional control in an evaluative situation in order to maintain task performance.

5. General Discussion

Test anxiety can hinder students from achieving their full potential in evaluative situations, such as tests or examinations. Converging evidence suggests that performance-related worries impair the working memory of these students. However, the mechanisms by which worries affect the working memory of test-anxious students remain poorly understood. The present work aimed to fill this gap by comparing the EEG of lower and higher test-anxious students performing a cognitive task in three studies. Compared to other neuroscientific techniques (e.g., fMRI), the temporal resolution of EEG is superior (up to 1 ms) allowing us to gain fine-grained and precise temporal information about the neurocognitive mechanisms underlying test anxiety (see Schillinger, 2016). Previous research has suggested that increased neural activity in anxious individuals during a cognitive task reflects compensatory effort to maintain task performance (for reviews, see Derakshan & Eysenck, 2009; Eysenck & Derakshan, 2011; Eysenck et al., 2007).

Based on the distraction model (Calvo & Eysenck, 1992; Eysenck et al., 2007; Wine, 1971), test-anxiety was conceptualized in the present work as a personality trait that interacts with the evaluative character of a situation in inducing performance-related worries (see Figure 1). Therefore, in all three studies lower and higher test-anxious students were tested performing a cognitive task in both a high pressure condition and a low pressure control condition (see Table 1).

Two studies addressed the response monitoring in test-anxious students by analyzing the ERN – an event-related potential occurring shortly after an erroneous response. Different lines of research suggest that the amplitude of the ERN is increased in individuals who are prone to worry (for reviews, see Moser et al., 2013; Weinberg et al., 2011). At the same time, performance in tasks used to measure the ERN (e.g., Stroop) appears to be uncompromised in anxious individuals (for reviews, see Gehring et al., 2012; Moser et al., 2013). Moser and colleagues (2013) have therefore proposed that the increased ERN seen in anxious individuals reflects compensatory efforts to maintain task performance. The aim of Study 1 was to start answering the question of how performance pressure changes response monitoring and whether this depends on individual differences in test anxiety. For this, 18 female participants were tested performing a numerical Stroop in two pressure conditions (i.e., low vs. high pressure) including test anxiety as a continuous variable in the analysis. However, students test anxiety scores turned out to be mostly in the average range and we did not control for differences in general anxiety in this study. In Study 2, we therefore compared the response monitoring of low ($N = 23$) and high test-anxious ($N = 21$) students while minimizing differences in general anxiety. In addition, participants were asked to rate their worries at regular intervals

throughout the experiment using an FAS in order to directly relate the ERN to worries (Bieri et al., 1990; Punaro & Reeve, 2012; Trezise & Reeve, 2014b). Based on the results of Study 1, a number of changes in the methods of Study 2 were implemented including focusing on the numerical task of the Stroop paradigm, recording both pressure conditions within a single EEG session, checking the effectiveness of the pressure manipulation, and analyzing both the ERN and the CRN using an ANOVA approach.

Given that the working memory load of the numerical Stroop task used in Study 1 and in Study 2 is relatively low the third study probed the effect of performance pressure using a more demanding task. For this, we adapted the *choking under pressure* paradigm reported by Beilock and colleagues (2004) to meet the requirements of an EEG study. Sixty participants were randomly assigned to either a control group or a pressure group prior to performing two blocks of modular arithmetic, with half of the problems having a low demand and the other half having a high demand. The first block served as a pretest and the second block as a posttest of arithmetic performance. While the control group was informed after the first block that they would be performing another set of problems, the pressure group was exposed to a high pressure test scenario during block 2. Participants in both groups were asked to rate their worries about performing well at regular intervals throughout the experiment using an FAS (Bieri et al., 1990; Punaro & Reeve, 2012; Trezise & Reeve, 2014b). Finally, working memory costs of performance pressure were assessed by analyzing the ERS/ERD in the theta frequency band (4-8 Hz) at frontal midline recording sites (FMΘ) during the task (Hsieh & Ranganath, 2014; Inanaga, 1998; Klimesch, 1999; Pfurtscheller & Aranibar, 1977; Pfurtscheller & Lopes da Silva, 1999; Roux & Uhlhaas, 2014; Sauseng et al., 2010). An increase in FMΘ has been linked to the need for attentional control regarding uncertain or threatening events by previous research (Cavanagh & Frank, 2014; Cavanagh & Shackman, 2015)

5.1. Summary of Results

Results of Study 1 showed that the response monitoring is modulated by both performance pressure and individual difference in test anxiety (Schillinger et al., 2016). More specifically, the ΔERN (i.e., the difference score between CRN and ERN) was enhanced in the high pressure as compared to the low pressure condition and linearly increased with test anxiety in both conditions. However, there was no significant interaction between performance pressure and test anxiety in modulating the ΔERN. On a behavioral level, the study provided mixed evidence with slower response times but increased accuracies in the high pressure condition than in the low pressure condition. Contrary to predictions, students with higher test

anxiety responded significantly faster under performance pressure than students with lower test anxiety.

The manipulation check of Study 2 firstly demonstrated that the pressure manipulation was effective as indicated by significantly higher ratings of performance pressure in the high pressure than in the low pressure condition. On an electrophysiological level, the ERN was confirmed to be more negative than the CRN across all conditions. The analysis further revealed that the ERN was modulated by performance pressure in HTA students but not in LTA students. More specifically, while the amplitude of the ERN was significantly enhanced in the pressure condition as compared to the control condition in HTA students, the amplitude did not significantly differ between pressure conditions in LTA students. On a behavioral level, both response time and accuracy showed a significant effect of congruency with responses to incongruent trials being slower and less accurate than responses to neutral and congruent trials. However, there was no significant difference in the behavioral performance between test anxiety groups or pressure conditions. Worries as assessed by an FAS turned out not to be significantly related to the amplitude of the ERN in either group. Instead, worries appeared to be directly related to the performance in the numerical Stroop task for LTA but not for HTA students.

Considering the effect of performance pressure on EEG indices of response monitoring across the two studies revealed firstly that the subsample with the lowest test anxiety (i.e., the LTA group of Study 2) showed no significant modulation of the ERN or the ΔERN. Secondly, the sample of Study 1 with mostly average levels of test anxiety exhibited a subtle modulation of the ERN, which was only significant in relationship to the decreasing CRN (i.e., the ΔERN). Finally, in the subsample with the highest test anxiety (i.e., the HTA group of Study 2), the amplitude of the ERN was significantly enhanced by performance pressure. Taken together, results show that the ERN is enhanced by performance pressure in higher but not in lower test-anxious students. At the same time, task performance was not (Study 2) or not unambiguously (Study 1) impaired by either test anxiety or performance pressure as reported by previous studies on the ERN in anxious individuals (for reviews, see Gehring et al., 2012; Moser et al., 2013) or individuals being evaluated (Hajcak, Moser, et al., 2005). This pattern of results is in line with the *compensatory error-monitoring hypothesis* by Moser and colleagues (2013). The present findings suggest accordingly that test-anxious students need compensatory effort to monitor their responses and to dynamically adapt to errors in an evaluative situation.

Study 3 could not replicate the previously reported *choking under pressure effect* in mathematical problem solving (Beilock & Carr, 2005; Beilock & DeCaro, 2007; Beilock et al., 2004, 2007; Boere et al., 2016; Ramirez & Beilock, 2011). Participants assigned to the pressure group did not report more pressure and performed at the same level as participants

assigned to the control group. Furthermore, the performance in the modular arithmetic task was not related to individual differences in test anxiety and self-reported worries were low across groups. In contrast, the study revealed significant differences between groups in the oscillatory brain responses during the task. The theta activity at frontal-midline recording sites was increased by performance pressure, especially in students with higher levels of test anxiety. Taken together, results are in line with the *adaptive control hypothesis* by Cavanagh and Shackman (2015). The findings of Study 3 suggest accordingly that test-anxious students increase their attentional control in an evaluative situation in order to maintain task performance.

5.2. Increased Neural Activity as Reflecting Compensatory Effort

Eysenck and colleagues have suggested that the increase in neural activity seen in anxious individuals during a cognitive task reflects their compensatory effort to maintain task performance (Derakshan & Eysenck, 2009; Eysenck & Derakshan, 2011; Eysenck et al., 2007). This proposition is based on a pattern of results reported by recent studies using neuroimaging techniques, such as EEG or fMRI. Studies have firstly shown that anxious individuals exhibit increased neural signals in different cognitive tasks, including an enhanced N2 in a Sustained Attention to Response Task (SART) (Righi et al., 2009), more frontal Contingent Negative Variation activity (CNV) during an antisaccade task (T. L. Ansari & Derakshan, 2011), and increased activity in the DLPFC in a Stroop task (Basten et al., 2011) as well as in an N-back task (Fales et al., 2008). Secondly, the performance in these tasks were reported not to differ significantly between lower and higher anxious participants. Thus, individuals with higher anxiety appear to use additional cognitive resources giving rise to the increased neural activity in order to perform on the same level as lower anxious individuals.

In the same vein, Moser and colleagues (2013) have suggested the *compensatory error-monitoring hypothesis* to account for the increased ERN amplitude in anxious individuals. Firstly, the ERN has been reported to be enhanced for various forms of anxiety, including general anxiety disorder (see Weinberg et al. 2011), obsessive-compulsory disorder (see Weinberg et al. 2011), general anxiety (Hajcak et al., 2003), negative affect (Hajcak et al., 2004; Luu et al., 2000), math anxiety (Suárez-Pellicioni et al., 2013), perfectionism (Perrone-McGovern et al., 2017), and helplessness (Pfabigan et al., 2013). Again, the performance in tasks used to measure the ERN (e.g., Stroop) appears to be uncompromised in anxious individuals (for reviews, see Gehring et al., 2012; Moser et al., 2013). Thus, anxious individuals appear to use compensatory effort to monitor ongoing responses and to dynamically adapt to errors. This compensatory effort is thought to give rise to more neural activity after committing an error as reflected in an increased ERN.

Finally, Cavanagh and Shackman (2015) have proposed a general framework to explain an increase of FMΘ in anxious individuals. According to the *adaptive control hypothesis*, higher FMΘ reflects the need for attentional control regarding uncertain or threatening events (Cavanagh & Shackman, 2015). This accounts in a similar fashion for the link between anxiety and the ERN as the *compensatory error-monitoring hypothesis* (Moser et al., 2013). But other ERPs, such as the FRN following punishment or negative feedback and the N2 in response to a mismatch, have also been shown to originate from a transient oscillatory synchronization in the theta frequency band (Cavanagh & Frank, 2014; Cavanagh & Shackman, 2015). Most recently, studies have started to address the role of FMΘ in both anxiety and situational stress using time-frequency analyses. These studies have demonstrated that FMΘ is increased by situational stress (Osinsky et al., 2017) as well as in individuals with symptoms of anxiety (Balconi & Crivelli, 2010; Mueller et al., 2014, 2015; Neo et al., 2011).

Taken together, different theoretical accounts have linked increased neural activity to compensatory efforts in anxious individuals during cognitive tasks. Note, however, that there are also studies which have found a decrease or decoupling of neural activity in anxious individuals. For instance, Bishop (2009) reported that students with higher general anxiety showed reduced activity in the DLPFC in response to incongruent stimuli in a response-conflict task. However, higher anxious participants also responded significantly slower to the incongruent stimuli than lower anxious participants. These results are only at the first sight contradictory to the notion of increased neural activity as reflecting compensatory effort (see Eysenck & Derakshan, 2011). If anxious students do not use compensatory effort or disengage from the task, they are predicted to show performance impairments (Derakshan & Eysenck, 2009; Eysenck & Derakshan, 2011; Eysenck et al., 2007). However, increased neural activity during a cognitive task without performance impairments is likely to indicate compensatory effort.

The present findings are in line with the notion that increased neural activity is reflecting compensatory effort. Two studies have demonstrated that the ERN is enhanced by performance pressure in higher but not in lower test-anxious students without clear differences in the task performance between conditions. And in Study 3, FMΘ was increased by performance pressure, especially in students with higher levels of test anxiety. However, test-anxious students could maintain their performance in the modular arithmetic task under performance pressure. Importantly, the cognitive tasks used in the present work were rated as highly important by both low and high test-anxious students (Study 2) as well as in the low and high pressure condition (Study 2 and Study 3). Thus, it seems unlikely that the present findings of higher neural activity reflect a purely motivational effect of performance pressure or test anxiety. In summary, the present three studies demonstrate that test-anxious students exhibit

enhanced neural activity performing a cognitive task under performance pressure. At the same time, the performance of test-anxious students appeared not to be impaired. Thus, students with higher test anxiety are likely to have used compensatory efforts to maintain their task performance under performance pressure.

Given that the present work suggests that test-anxious students use compensatory efforts to uphold task performance in evaluative situation, the question arises what these compensatory efforts exactly are. According to the attentional control theory, compensatory efforts comprise both the increase in attentional control in the first place and the usage of auxiliary strategies (e.g. articulatory rehearsal when reading). Given that in the present studies usage of auxiliary strategies seems to be unlikely, test-anxious students probably increased their attentional control in performing the cognitive task. However, future studies are needed to better characterize the cognitive processes that give rise to the increased neural activity in test-anxious students being evaluated.

5.3. Present Results within the Distraction Model

The present research was framed within the distraction model of the effect of test anxiety on cognitive performance (see Figure 1). This model was synthesized based on the distraction theory by Wine (1971) and the attentional control theory by Eysenck and colleagues (Calvo & Eysenck, 1992; Eysenck et al., 2007). In this model, test-anxiety is conceptualized as a personality trait that interacts with the evaluative character of a situation in inducing performance-related worries in students. These worries, in turn, are thought to impair the working memory of students including the executive functions and the phonological loop (see Baddeley, 2003). As a consequence, working memory resources have to be split between the task at hand and the distracting worries. If the combined demands of task-related and worry-related processes exceed the available working memory resources, performance impairments will result. How do the present findings relate to the distraction model?

Firstly, the present work provides evidence that there is indeed an interaction between the personality trait of test anxiety and the evaluative character of a situation. In Study 2, test anxiety and performance pressure were shown to interact in modulating the ERN. In the same vein, there was a linear relationship between the increase of FMΘ in the pressure but not in the control group of Study 3. Note, however, that we did not find a significant interaction between test anxiety and performance pressure in modulating the ERN in Study 1. This might have been due to the fact that individual differences in general anxiety were not controlled for. The test-anxious students in Study 1 are therefore likely to also have been more anxious in general and thus to show a similar response monitoring in both pressure conditions. Taken

together, the present findings are in line with the distraction model according to which test anxiety affects the cognition depending on the evaluative character of a situation.

With respect to the induction of worries in test-anxious students by being evaluated as assumed by the distraction model, the present studies appear to be in discordance with the model since we did not find evidence that higher test-anxious students experience more worries than lower test-anxious students. Worries were assessed in Study 2 and Study 3 by an FAS presented at regular intervals throughout the experiment (Bieri et al., 1990; Punaro & Reeve, 2012; Trezise & Reeve, 2014b). The FAS consisted of six facial expressions (see Figure 7) and participants were asked to select the first face (neutral expression) if they were not worried at all about their performance and the sixth face (expressing discomfort) if they were very worried about their performance. This worry rating was not significantly higher for HTA than for LTA students in either pressure condition of Study 2. Similarly, there was no significant difference between the worries reported by participants assigned to the pressure or the control group in Study 3. Instead, worries were rather directly related to the performance in the cognitive task used in the present studies – irrespective of participants' test anxiety. In Study 2, worry ratings increased with the number of errors in the numerical Stroop task for LTA but not for HTA students. Similarly, students in both groups of Study 3 reported a higher level of worries in the posttest when they had committed more errors than in the pretest. Thus, the FAS appeared to have rather captured participants' awareness of having made errors than their more abstract worries about performing well. Possible alternatives to using an FAS to assess worries are therefore discussed in the next section.

The next part of the distraction model is the impairment of working memory processes by worries in test-anxious students. Here, the present neural data are in line with a detrimental effect of performance pressure on the working memory of test-anxious students. Both the stronger modulation of the ERN and the increase of FMΘ between blocks indicate that test-anxious students needed compensatory effort to maintain their performance.

Finally, the distraction model aims to account for cognitive impairments of test-anxious students and eventually for their lower academic achievements. In contrast, we did not see clear behavioral differences between lower and higher test-anxious students in the present work. However, it is important to note that the detrimental effects of performance pressure are not conceptualized to directly translate into a lower performance within the distraction model. Instead, students are thought to have a limited working memory capacity (see Baddeley, 2003) that can allow to compensate for the detrimental effect of performance pressure. Only when the combined demands of task-related and worry-related processes exceed this working memory capacity performance impairments will result. Thus, if enough cognitive resources are available, task performance (i.e., the performance effectiveness) will remain unaffected. However, given that more cognitive resources have been used to maintain task performance,

the processing efficiency is reduced (see Calvo & Eysenck, 1992; Eysenck et al., 2007). In fact, the present findings of increased neural activity in test-anxious students under performance pressure is indicating that such additional cognitive resources have been used to maintain performance (see Berggren & Derakshan, 2013; Derakshan & Eysenck, 2009; Eysenck & Derakshan, 2011; Eysenck et al., 2007). Thus, while the performance effectiveness is unimpaired in the present studies, the processing efficiency seems to be lower in test-anxious students under performance pressure.

In summary, the present findings are by and large in line with the distraction model. Test anxiety and the evaluative character of a situation were shown to interact in modulating the ERN and FMƟ. Both neural signals, in turn, indicate a compensation of the detrimental effects of performance pressure on the working memory of test-anxious students. However, the present studies remain inconclusive regarding the role of worries in mediating the relationship between performance pressure and impaired working memory. Finally, while there were no clear performance impairments in the present studies, the neural data suggest that test-anxious students needed compensatory effort to maintain task performance, hence showing reduced processing efficiency.

5.4. Limitations and Future Research

As discussed in the previous section, performance-related worries were not related to individual differences in test anxiety or to the EEG measures in the present studies. Instead, worries appeared to be directly related to the performance in the cognitive tasks. This raises the question of whether the FAS we used in the present work captures worries about performing well or rather captures participants' awareness of having made errors. The present FAS is based on the Faces Pain Scale by Bieri and colleagues (1990) and has been previously used to assess worries in children related to literacy (Punaro & Reeve, 2012) and math performance (Punaro & Reeve, 2012; Trezise & Reeve, 2014a, 2014b). Furthermore, the original FAS (Bieri et al., 1990) has been shown to be a reliable measure of the intensity of pain in adults. To the best of our knowledge, the present studies are the first to use an FAS to assess performance-related worries in adults. Future research is therefore needed to evaluate the FAS for assessing worries about cognitive performance in adults.

Another approach to assess performance-related worries is to use an open answer format during or after an evaluative situation. For instance, Beilock and colleagues (2004) asked participants to recall their thoughts during the high pressure scenario after completing the last block of modular arithmetic. This verbal thought questionnaire revealed that worrisome thoughts accounted for more than half of all thoughts reported by participants. In another study (Benny & Banks, 2015), participants reported their thoughts while performing the modular

arithmetic tasks under performance pressure using five response options, including both positive and negative thoughts. Negative but not positive thoughts turned out to be inversely related to the performance in the modular arithmetic task. Both procedures are promising in assessing performance-related worries with more facets than by using an FAS.

A further limitation of the present work is that a high pressure test situation was only simulated in the laboratory. By using a manipulation check, we could demonstrate that the pressure manipulation was effective in Study 2. However, the manipulation check in Study 3 could not clearly confirm that the high pressure condition was perceived as more stressful than the low pressure condition. In general, modeling a real-life test situation in the laboratory is limited. Because of ethical considerations, participants cannot be actually evaluated, for instance, by giving them a real grade. Moreover, most students participating in laboratory experiments are volunteers who participate more often in such experiments either for money or course credits. As such, they can be expected to have some implicit or explicit knowledge about psychological testing. Especially students of psychology are therefore less likely to believe that they are actually evaluated in studies like the present.

To overcome this limitation future studies could consider to address the effect of performance pressure in regard to real-world tests or examinations. The gold standard would certainly be to investigate test-anxious students while being in a real-world test situation, such as a written or oral exam. However, to the present day, it is not possible to record EEG in such situations, especially without interfering with the test situation. Alternatively, studies could investigate the cognition of test-anxious students before or after an important exam. For instance, Ramirez and Beilock (2011) tested an intervention of expressive writing for alleviating the effect of choking under pressure in a field experiment. For this, ninth graders were asked just before their final high-school exam in biology to either write 10 min about their feelings towards the upcoming exam (expressive writing group) or to think about a topic that would not be covered on the exam (control group). Results revealed that the relationship between test anxiety and the exam scores that was present in the control condition turned insignificant in the expressive writing group. Furthermore, Wu and colleagues (2014) compared the ERN between students undergoing preparation for a major exam and students without upcoming exams. While the two groups did not differ in in their ERN amplitude, the group facing the exam showed an increased positivity following the ERN. However, simulating performance pressure in the laboratory, as in the present studies, remain without alternative to investigate the neurocognitive mechanisms underlying test anxiety in an evaluative situation.

Finally, the present results are limited by the fact that the studies did not include an independent measure of attentional control. The conclusion according to which the increase in neural activity in test-anxious students under performance pressure is due to compensatory effort and specifically to an increase of attentional control is therefore to some degree based

on reverse inference (see De Smedt et al., 2011). Future studies could support this reasoning by implementing an independent measure of attentional control. For instance, measuring the pupil dilation during the task could provide complementary information about the dynamics of attentional control in test-anxious students (for a review, see Beatty, 1982). Another idea would be to use a mental effort rating as common in research within the cognitive load theory (for reviews, see Brünken, Plass, & Leutner, 2003; Paas, Tuovinen, Tabbers, & Van Gerven, 2003). For instance, Hadwin and colleagues (2005) asked children to rate their mental effort using a visual analog scale while performing different working memory tasks. Comparing lower and higher anxious children revealed that they did not differ in their accuracy across working memory tasks. However, higher anxious children reported higher mental effort in some of the working memory tasks.

In summary, future research should consider assessing performance-related worries by using an open answer format as well as including an independent measure of attentional control. This could help to corroborate the present account of the increased neural activity in test-anxious students being evaluated as reflecting compensatory effort to maintain task performance.

5.5. Concluding Remarks

Test anxiety can hinder students from achieving their full potential in evaluative situations, such as tests or examinations. Converging evidence suggests that performance-related worries impair the working memory of these students. The aim of the present work was to better understand the neurocognitive mechanisms by which worries affect the working memory of test-anxious students. In three EEG studies, higher test-anxious students were shown to display increased neural activity performing a cognitive task under performance pressure as compared to lower test-anxious students. At the same time, the performance of higher test-anxious students was not impaired in the present studies. Furthermore, self-reported worries were not related to individual differences in test anxiety or the EEG measures. Instead, worries appeared to be directly related to participants' awareness of having committed errors. Taken together, results demonstrate that test-anxious students exhibit increased neural activity under performance pressure, suggesting that they need compensatory effort to maintain performance in evaluative situations. Future research should aim to corroborate the present findings by improving the assessment of worries and by including an independent measure of attentional control.

6. References

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*(2), 219–245. https://doi.org/10.1037//0033-2909.121.2.219

American Psychiatric Association. (2013). *American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders (5th ed.). American Journal of Psychiatry.* https://doi.org/10.1176/appi.books.9780890425596.744053

Ansari, D., Grabner, R. H., Koschutnig, K., Reishofer, G., & Ebner, F. (2011). Individual differences in mathematical competence modulate brain responses to arithmetic errors: An fMRI study. *Learning and Individual Differences*, *21*(6), 636–643. https://doi.org/10.1016/j.lindif.2011.07.013

Ansari, T. L., & Derakshan, N. (2011). The neural correlates of cognitive effort in anxiety: Effects on processing efficiency. *Biological Psychology*, *86*(3), 337–348. https://doi.org/10.1016/j.biopsycho.2010.12.013

Ashcraft, M. H. (2002). Math Anxiety: Personal, Educational, and Cognitive Consequences. *Current Directions in Psychological Science*, *11*(5), 181–185. https://doi.org/10.1111/1467-8721.00196

Atkinson, R. C., & Geiser, S. (2009). Reflections on a Century of College Admissions Tests. *Educational Researcher*, *38*(9), 665–676. https://doi.org/10.3102/0013189X09351981

Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature Reviews. Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Balconi, M., & Crivelli, D. (2010). Veridical and false feedback sensitivity and punishment-reward system (BIS/BAS): ERP amplitude and theta frequency band analysis. *Clinical Neurophysiology*, *121*(9), 1502–1510. https://doi.org/10.1016/j.clinph.2010.03.015

Basten, U., Stelzel, C., & Fiebach, C. J. (2011). Trait anxiety modulates the neural efficiency of inhibitory control. *Journal of Cognitive Neuroscience*, *23*(10), 3132–3145. https://doi.org/10.1162/jocn_a_00003

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: linear mixed-effects models using S4 classes. R package version 1.1-6. *R*. https://doi.org/http://CRAN.R-project.org/package=lme4

Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, *46*, 610–620. https://doi.org/10.1037/0022-3514.46.3.610

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Beidel, D. C., Turner, M. W., & Trager, K. N. (1994). Test anxiety and childhood anxiety disorders in African American and white school children. *Journal of Anxiety Disorders*, *8*(2), 169–179. https://doi.org/10.1016/0887-6185(94)90014-0

Beilock, S. L. (2008). Math Performance in Stressful Situations. *Current Directions in Psychological Science*, *17*(5), 339–343. https://doi.org/10.1111/j.1467-8721.2008.00602.x

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, *130*(4), 701–725. https://doi.org/10.1037//0096-3445.130.4.701

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: working memory and "choking under pressure" in math. *Psychological Science*, *16*(2), 101–5. https://doi.org/10.1111/j.0956-7976.2005.00789.x

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(6), 983–98. https://doi.org/10.1037/0278-7393.33.6.983

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: choking under pressure in mathematical problem solving. *Journal of Experimental Psychology. General*, *133*(4), 584–600. https://doi.org/10.1037/0096-3445.133.4.584

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spillover. *Journal of Experimental Psychology. General*, *136*(2), 256–76. https://doi.org/10.1037/0096-3445.136.2.256

Benny, B., & Banks, J. B. (2015). Under Pressure. *Journal of Individual Differences*, *36*(2), 93–100. https://doi.org/10.1027/1614-0001/a000160

Berggren, N., & Derakshan, N. (2013). Attentional control deficits in trait anxiety: Why you see them and why you don't. *Biological Psychology*, *92*(3), 440–446. https://doi.org/10.1016/j.biopsycho.2012.03.007

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, *17*(5), 467–472. https://doi.org/10.1016/0028-3932(79)90053-8

Bieri, D., Reeve, R. A., Champion, G. D., Addicoat, L., & Ziegler, J. B. (1990). The faces pain scale for the self-assessment of the severity of pain experienced by children: Development, initial validation, and preliminary investigation for ratio scale properties. *Pain*, *41*(2), 139–150. https://doi.org/10.1016/0304-3959(90)90018-9

Birenbaum, M., & Pinku, P. (1997). Effects of test anxiety, information organization, and testing situation on performance on two test formats. *Contemporary Educational Psychology*, *22*(1), 23–38. https://doi.org/10.1006/ceps.1997.0923

Bishop, S. J. (2009). Trait anxiety and impoverished prefrontal control of attention. *Nature Neuroscience*, *12*(1), 92–98. https://doi.org/10.1038/nn.2242

Boere, J. J., Fellinger, L., Huizinga, D. J. H., Wong, S. F., & Bijleveld, E. (2016). Performance pressure and caffeine both affect cognitive performance, but likely through independent mechanisms. *Brain and Cognition*, *102*, 26–32. https://doi.org/10.1016/j.bandc.2015.11.006

Bruch, M. A. (1981). Relationship of test-taking strategies to test anxiety and performance: Toward a task analysis of examination behavior. *Cognitive Therapy and Research*, *5*(1), 41–56. https://doi.org/10.1007/BF01172325

Bruch, M. A., Juster, H. R., & Kaflpwitz, N. G. (1983). Relationships of Cognitive Components of Test Anxiety to Test Performance : Implications for Assessment and Treatment. *Journal of Counseling Psychology*, *30*(4), 527–536. https://doi.org/10.1037/0022-0167.30.4.527

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 53–61. https://doi.org/10.1207/S15326985EP3801_7

Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition*, *118*(1), 35–47. https://doi.org/10.1016/j.cognition.2010.09.005

Byrne, K. A., Silasi-Mansat, C. D., & Worthy, D. A. (2015). Who chokes under pressure? The Big Five personality traits and decision-making under pressure. *Personality and Individual Differences*, *74*, 22–28. https://doi.org/10.1016/j.paid.2014.10.009

Calvo, M. G., & Eysenck, M. W. (1992). Anxiety and Performance: The Processing Efficiency Theory. *Cognition and Emotion*, *6*(6), 409–434. https://doi.org/10.1080/02699939208409696

Calvo, M. G., Ramos, P. M., & Estevez, A. (1992). Test anxiety and comprehension efficiency: The role of prior knowledge and working memory deficits. *Anxiety Stress Coping*, *5*, 125–138. https://doi.org/10.1080/10615809208250492

Carey, E., Hill, F., Devine, A., & Szücs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology*, *6*(JAN), 1–6. https://doi.org/10.3389/fpsyg.2015.01987

Carter, C. S. (1998). Anterior Cingulate Cortex, Error Detection, and the Online Monitoring of Performance. *Science*, *280*(5364), 747–749. https://doi.org/10.1126/science.280.5364.747

Cassady, J. C. (2004). The influence of cognitive test anxiety across the learning-testing cycle. *Learning and Instruction*, *14*(6), 569–592. https://doi.org/10.1016/j.learninstruc.2004.09.002

Cassady, J. C., & Johnson, R. E. (2002). Cognitive Test Anxiety and Academic Performance. *Contemporary Educational Psychology*, *27*(2), 270–295. https://doi.org/10.1006/ceps.2001.1094

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Cavanagh, J. F., & Shackman, A. J. (2015). Frontal midline theta reflects anxiety and cognitive control: Meta-analytic evidence. *Journal of Physiology-Paris*, *109*(1–3), 3–15. https://doi.org/10.1016/j.jphysparis.2014.04.003

Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test Anxiety and Academic Performance in Undergraduate and Graduate Students. *Journal of Educational Psychology*, *97*(2), 268–274. https://doi.org/10.1037/0022-0663.97.2.268

Chib, V. S., De Martino, B., Shimojo, S., & O'Doherty, J. P. (2012). Neural mechanisms underlying paradoxical performance for monetary incentives are driven by loss aversion. *Neuron*, *74*(3), 582–94. https://doi.org/10.1016/j.neuron.2012.02.038

Coles, M. G., Scheffers, M. K., & Holroyd, C. B. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological Psychology*, *56*(3), 173–189. https://doi.org/10.1016/S0301-0511(01)00076-X

Darke, S. (1988). Anxiety and working memory capacity. *Cognition & Emotion*, *2*(2), 145–154. https://doi.org/10.1080/02699938808408071

De Smedt, B., Ansari, D., Grabner, R. H., Hannula-Sormunen, M., Schneider, M., & Verschaffel, L. (2011). Cognitive neuroscience meets mathematics education: It takes two to Tango. *Educational Research Review*, *6*(3), 232–237. https://doi.org/10.1016/j.edurev.2011.10.003

De Smedt, B., Grabner, R. H., & Studer, B. (2009). Oscillatory EEG correlates of arithmetic strategy use in addition and subtraction. *Experimental Brain Research*, *195*(4), 635–642. https://doi.org/10.1007/s00221-009-1839-9

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal*

of *Experimental Child Psychology*, *103*(4), 469–479.
https://doi.org/10.1016/j.jecp.2009.01.010

DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: multiple routes to skill failure. *Journal of Experimental Psychology. General*, *140*(3), 390–406. https://doi.org/10.1037/a0023466

Deffenbacher, J. L. (1978). Worry, emotionality, and task-generated interference in test anxiety: An empirical test of attentional theory. *Journal of Educational Psychology*, *70*(2), 248–254. https://doi.org/10.1037/0022-0663.70.2.248

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Derakshan, N., & Eysenck, M. W. (2009). Anxiety, processing efficiency, and cognitive performance: New developments from attentional control theory. *European Psychologist*, *14*(2), 168–176. https://doi.org/10.1027/1016-9040.14.2.168

Dew, K., Galassi, J. P., & Galassi, M. D. (1983). Mathematics Anxiety: Some Basic Issues. *Journal of Counseling Psychology*, *30*(3), 443–446. https://doi.org/10.1037/0022-0167.30.3.443

Eysenck, M. W., & Derakshan, N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences*, *50*(7), 955–960. https://doi.org/10.1016/j.paid.2010.08.019

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, *7*(2), 336–53. https://doi.org/10.1037/1528-3542.7.2.336

Fales, C. L., Barch, D. M., Burgess, G. C., Schaefer, A., Mennin, D. S., Gray, J. R., & Braver, T. S. (2008). Anxiety and cognitive efficiency: Differential modulation of transient and sustained neural activity during a working memory task. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(3), 239–253. https://doi.org/10.3758/CABN.8.3.239

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*(6), 447–455.

Folin, O., Denis, W., & Smillie, W. G. (1914). Some observations on "emotional glycosuria" in man. *Journal of Biological Chemistry*, *17*, 519–520.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Formazin, M., Schroeders, U., Köller, O., Wilhelm, O., & Westmeyer, H. (2011). Studierendenauswahl im fach psychologie testentwicklung und validitätsbefunde. *Psychologische Rundschau*, *62*(4), 221–236. https://doi.org/10.1026/0033-3042/a000093

Ganushchak, L. Y., & Schiller, N. O. (2008). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *NeuroImage*, *39*(1), 395–405. https://doi.org/10.1016/j.neuroimage.2007.09.001

Gärtner, M., Grimm, S., & Bajbouj, M. (2015). Frontal midline theta oscillations during mental arithmetic: effects of stress. *Frontiers in Behavioral Neuroscience*, *9*(April), 1–8. https://doi.org/10.3389/fnbeh.2015.00096

Gärtner, M., Rohde-Liebenau, L., Grimm, S., & Bajbouj, M. (2014). Working memory-related frontal theta activity is decreased under acute stress. *Psychoneuroendocrinology*, *43*, 105–113. https://doi.org/10.1016/j.psyneuen.2014.02.009

Gehring, W. J., Goss, B., & Coles, M. G. H. (1993). A neural system for error detection and compensation. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A Neural System for Error Detection and Compensation. *Psychological Science*, *4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (2012). The Error-Related Negativity (ERN/Ne). In *The Oxford Handbook of Event-related Potential Components*.

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, *7*(4), 374–385. https://doi.org/10.1093/cercor/7.4.374

Gorsuch, R. L. (1966). The general factor in the test anxiety questionnaire. *Psychological Reports*, *19*, 308.

Grabner, R. H., & Ansari, D. (2010). Promises and potential pitfalls of a "cognitive neuroscience of mathematics learning." *ZDM The International Journal on Mathematics Education*, *42*(6), 655–660. https://doi.org/10.1007/s11858-010-0283-4

Grabner, R. H., & De Smedt, B. (2011). Neurophysiological evidence for the validity of verbal strategy reports in mental arithmetic. *Biological Psychology*, *87*(1), 128–36. https://doi.org/10.1016/j.biopsycho.2011.02.019

Grabner, R. H., & De Smedt, B. (2012). Oscillatory EEG Correlates of Arithmetic Strategies: A Training Study. *Frontiers in Psychology*, *3*(10), 428. https://doi.org/10.3389/fpsyg.2012.00428

Hadwin, J. A., Brogan, J., & Stevenson, J. (2005). State anxiety and working memory in children: A test of processing efficiency theory. *Educational Psychology*, *25*(4), 379–393. https://doi.org/10.1080/01443410500041607

Hajcak, G., McDonald, N., & Simons, R. F. (2003). Anxiety and error-related brain activity. *Biological Psychology*, *64*(1–2), 77–90. https://doi.org/10.1016/S0301-0511(03)00103-0

Hajcak, G., McDonald, N., & Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain and Cognition*, *56*(2), 189–97. https://doi.org/10.1016/j.bandc.2003.11.001

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology*, *42*(2), 151–60. https://doi.org/10.1111/j.1469-8986.2005.00270.x

Hajcak, G., Nieuwenhuis, S., Ridderinkhof, K. R., & Simons, R. F. (2005). Error-preceding brain activity: Robustness, temporal dynamics, and boundary conditions. *Biological Psychology*, *70*(2), 67–78. https://doi.org/10.1016/j.biopsycho.2004.12.001

Hayes, A. (2013). Introduction to mediation, moderation, and conditional process analysis. *New York, NY: Guilford*, 3–4. https://doi.org/978-1-60918-230-4

Hayes, S., Hirsch, C., & Mathews, A. (2008). Restriction of working memory capacity during worry. *Journal of Abnormal Psychology*, *117*(3), 712–717. https://doi.org/10.1037/a0012908

Hembree, R. (1988). *Correlates, Causes, Effects, and Treatment of Test Anxiety*. *Review of Educational Research* (Vol. 58). https://doi.org/10.3102/00346543058001047

Hembree, R. (1990). The Nature, Effects, and Relief of Mathematics Anxiety. *Journal for Research in Mathematics Education*, *21*(1), 33. https://doi.org/10.2307/749455

Henik, A., & Tzelgov, J. (1982). Is three greater than five: the relation between physical and

semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395. https://doi.org/10.3758/BF03202431

Hill, K. T., & Sarason, S. B. (1966). The Relation of Test Anxiety and Defensiveness to Test and School Performance over the Elementary-School Years: A Further Longitudinal Study. *Monographs of the Society for Research in Child Development*, *31*(2), 1–76. https://doi.org/10.2307/1165770

Hill, K. T., & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. *The Elementary School Journal*, *85*(1), 105–126. https://doi.org/10.1086/461395

Hirsh, J. B., & Inzlicht, M. (2010). Error-related negativity predicts academic performance. *Psychophysiology*, *47*(1), 192–6. https://doi.org/10.1111/j.1469-8986.2009.00877.x

Hodapp, V. (1991). Das Prüfungsängstlichkeitsinventar TAI-G: Eine erweiterte und modifizierte Version mit vier Komponenten. *Zeitschrift Für Pädagogische Psychologie*, *5*(2), 121–130.

Hodapp, V., & Benson, J. (1997). The multidimensionality of test anxiety: A test of different models. *Anxiety, Stress and Coping*, *10*(3), 219–244. https://doi.org/10.1080/10615809708249302

Hodapp, V., Laux, L., & Spielberger, C. (1982). Theorie und Messung der emotionalen und kognitiven Komponente der Prüfungsangst. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, *3*(3), 169–184.

Hodapp, V., Rohrmann, S., & Ringeisen, T. (2011). PAF -Prüfungsangstfragebogen. *Göttingen: Hogrefe*.

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

Hsieh, L. T., & Ranganath, C. (2014). Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *NeuroImage*, *85*, 721–729. https://doi.org/10.1016/j.neuroimage.2013.08.003

IBM Corp. Released. (2011). IBM SPSS Statistics for Windows, Version 22.0. *2011*.

Inanaga, K. (1998). Frontal midline theta rhythm and mental activity. *Psychiatry and Clinical Neurosciences*, *52*(6), 555–566. https://doi.org/10.1046/j.1440-1819.1998.00452.x

Ishihara, T., & Yoshii, N. (1972). Multivariate analytic study of EEG and mental activity in Juvenile delinquents. *Electroencephalography and Clinical Neurophysiology*, *33*(1), 71–80. https://doi.org/10.1016/0013-4694(72)90026-0

Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision making. *NeuroImage*, *32*(2), 978–987. https://doi.org/10.1016/j.neuroimage.2006.02.018

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *Neuroscience*, *15*(8), 1395–1399. https://doi.org/10.1046/j.1460-9568.2002.01975.x

Kalcher, J., & Pfurtscheller, G. (1995). Discrimination between phase-locked and non-phase-locked event-related EEG activity. *Electroencephalography and Clinical Neurophysiology*, *94*(5), 381–384. https://doi.org/10.1016/0013-4694(95)00040-6

Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., … Ischebeck, A. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task : an event-related fMRI study. *NeuroImage*, *25*, 888–898.

https://doi.org/10.1016/j.neuroimage.2004.12.041

Keith, N., Hodapp, V., Schermelleh-Engel, K., & Moosbrugger, H. (2003). Cross-sectional and longitudinal confirmatory factor models for the german test anxiety inventory: A construct validation. *Anxiety, Stress & Coping*, *16*(3), 251–270. https://doi.org/10.1080/1061580031000095416

Kim, E. Y., Iwaki, N., Imashioya, H., Uno, H., & Fujita, T. (2007). Error-related negativity in a visual go/no-go task: children vs. adults. *Developmental Neuropsychologyeuropsychology*, *31*(2), 181–191. https://doi.org/10.1080/87565640701190775

Kim, E. Y., Iwaki, N., Uno, H., & Fujita, T. (2005). Error-related negativity in children: effect of an observer. *Developmental Neuropsychology*, *28*(3), 871–83. https://doi.org/10.1207/s15326942dn2803_7

Kirkland, K., & Hollandsworth, J. G. (1980). Effective test taking: skills-acquisition versus anxiety-reduction techniques. *Journal of Consulting and Clinical Psychology*, *48*(4), 431–439. https://doi.org/10.1037//0022-006X.48.4.431

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, *1*, 1–12. https://doi.org/10.3389/fpsyg.2010.00238

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, *29*(2–3), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old students. *Cognition*, *93*(2), 99–125. https://doi.org/10.1016/j.cognition.2003.11.004

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. (1981). Das State-Trait-Angstinventar: STAI. *Weinheim: Beltz*.

Lee, T. G., & Grafton, S. T. (2014). Out of control: Diminished prefrontal activity coincides with impaired motor performance due to choking under pressure. *NeuroImage*, *105*, 145–155. https://doi.org/10.1016/j.neuroimage.2014.10.058

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety. *Psychological Reports*, *20*(1963), 975–978.

Liu, Z. X., Woltering, S., & Lewis, M. D. (2014). Developmental change in EEG theta activity in the medial prefrontal cortex during response control. *NeuroImage*, *85*, 873–887. https://doi.org/10.1016/j.neuroimage.2013.08.054

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213

Luu, P., Collins, P., & Tucker, D. M. (2000). Mood, personality, and self-monitoring: Negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *Journal of Experimental Psychology: General*, *129*(1), 43–60. https://doi.org/10.1037/0096-3445.129.1.43

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation. *Clinical Neurophysiology*, *115*(8), 1821–1835. https://doi.org/10.1016/j.clinph.2004.03.031

Maloney, E. a., Sattizahn, J. R., & Beilock, S. L. (2014). Anxiety and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(August), 403–411. https://doi.org/10.1002/wcs.1299

Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal Psychology*, *47*(2), 166–173. https://doi.org/10.1037/h0062855

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Markman, A. B., Maddox, W. T., & Worthy, D. a. (2006). Choking and excelling under pressure. *Psychological Science*, *17*(11), 944–8. https://doi.org/10.1111/j.1467-9280.2006.01809.x

Massar, S. A. A., Rossi, V., Schutter, D. J. L. G., & Kenemans, J. L. (2012). Baseline EEG theta/beta ratio and punishment sensitivity as biomarkers for feedback-related negativity (FRN) and risk-taking. *Clinical Neurophysiology*, *123*(10), 1958–1965. https://doi.org/10.1016/j.clinph.2012.03.005

Mattarella-Micke, A., Mateo, J., Kozak, M. N., Foster, K., & Beilock, S. L. (2011). Choke or thrive? The relation between salivary cortisol and math performance depends on individual differences in working memory and math-anxiety. *Emotion*, *11*(4), 1000–1005. https://doi.org/10.1037/a0023224

McDonald, A. S. (2001). The Prevalence and Effects of Test Anxiety in School Children. *Educational Psychology*, *21*(1), 89–101. https://doi.org/10.1080/01443410020019867

Middendorf, E., Apolinarski, B., Poskowsky, J., Kandulla, M., Netz, N., Naumann, H., & Buck, D. (2016). Die wirtschaftliche und soziale Lage der Studierenden in der Bundesrepublik Deutschland 2016. *Sozialerhebung Des Deutschen …*, 1–516. Retrieved from http://www.sozialerhebung.de/download/20/soz20_hauptbericht_gesamt.pdf

Missonnier, P., Deiber, M.-P., Gold, G., Millet, P., Gex-Fabry Pun, M., Fazio-Costa, L., … Ibáñez, V. (2006). Frontal theta event-related synchronization: comparison of directed attention and working memory load effects. *Journal of Neural Transmission*, *113*(10), 1477–1486. https://doi.org/10.1007/s00702-005-0443-9

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Mobbs, D., Hassabis, D., Seymour, B., Marchant, J. L., Weiskopf, N., Dolan, R. J., & Frith, C. D. (2009). Choking on the money: reward-based performance decrements are associated with midbrain activity. *Psychological Science*, *20*(8), 955–62. https://doi.org/10.1111/j.1467-9280.2009.02399.x

Moran, T. P. (2016). Anxiety and Working Memory Capacity: A Meta-Analysis and Narrative Review. *Psychological Bulletin*, *142*(5), 831–864. https://doi.org/10.1037/bul0000051

Moser, J. S., Moran, T. P., & Jendrusina, A. a. (2012). Parsing relationships between dimensions of anxiety and action monitoring brain potentials in female undergraduates. *Psychophysiology*, *49*(1), 3–10. https://doi.org/10.1111/j.1469-8986.2011.01279.x

Moser, J. S., Moran, T. P., Schroder, H. S., Donnellan, M. B., & Yeung, N. (2013). On the relationship between anxiety and error monitoring: a meta-analysis and conceptual framework. *Frontiers in Human Neuroscience*, *7*, 466. https://doi.org/10.3389/fnhum.2013.00466

Mowbray, T. (2012). Working memory, test anxiety and effective interventions: a review. *The Australian Educational and Developmental Psychologist*, *29*, 141–156. https://doi.org/10.1017/edp.2012.16

Mueller, E. M., Burgdorf, C., Chavanon, M. L., Schweiger, D., Wacker, J., & Stemmler, G. (2014). Dopamine modulates frontomedial failure processing of agentic introverts versus

extraverts in incentive contexts. *Cognitive, Affective and Behavioral Neuroscience*, *14*(2), 756–768. https://doi.org/10.3758/s13415-013-0228-9

Mueller, E. M., Pechtel, P., Cohen, A. L., Douglas, S. R., & Pizzagalli, D. A. (2015). Potentiated processing of negative feedback in depression is attenuated by anhedonia. *Depression and Anxiety*, *32*(4), 296–305. https://doi.org/10.1002/da.22338

Naveh-Benjamin, M. (1991). A Comparison of Training Programs Intended for Different Types of Test-Anxious Students: Further Support for an Information-Processing Model. *Journal of Educational Psychology*, *83*(1), 134–139. https://doi.org/10.1037/0022-0663.83.1.134

Naveh-Benjamin, M., McKeachie, W. J., Lin, Y., & Holinger, D. P. (1981). Test anxiety: Deficits in information processing. *Journal of Educational Psychology*, *73*(6), 816–824. https://doi.org/10.1037/0022-0663.73.6.816

Neo, P. S. H., Thurlow, J. K., & McNaughton, N. (2011). Stopping, goal-conflict, trait anxiety and frontal rhythmic power in the stop-signal task. *Cognitive, Affective and Behavioral Neuroscience*, *11*(4), 485–493. https://doi.org/10.3758/s13415-011-0046-x

Neumann, O. (2005). Carl friedrich gauss, disquisitiones arithmeticae (1801). In *Landmark Writings in Western Mathematics 1640-1940* (pp. 303–315). https://doi.org/10.1016/B978-044450871-3/50103-0

OECD. (2013). PISA 2012 Results: Ready to Learn – Students' Engagement, Drive and Self-Beliefs (Volume III), PISA, OECD Publishing.

Olvet, D., & Hajcak, G. (2009a). Reliability of error-related brain activity. *Brain Research*, *1284*, 89–99. https://doi.org/10.1016/j.brainres.2009.05.079

Olvet, D., & Hajcak, G. (2009b). The stability of error related brain activity with increasing trials. *Psychophysiology*, *46*, 957–961. https://doi.org/10.1111/j.1469-8986.2009.00848.x

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Osinsky, R., Karl, C., & Hewig, J. (2017). Dispositional Anxiety and Frontal-Midline Theta: On the Modulatory Influence of Sex and Situational Threat. *Journal of Personality*, *85*(3), 300–312. https://doi.org/10.1111/jopy.12241

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pekrun, R. (1992). Kognition und Emotion in studienbezogenen Lern- und Leistungssituationen: Explorative Analysen. *Unterrichtswissenschaft*, *20*, 308–324.

Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M., & Molfenter, S. (2004). Beyond test anxiety: Development and validation of the Test Emotions Questionnaire (TEQ). *Anxiety, Stress and Coping*, *17*(3), 287–316. https://doi.org/10.1080/10615800412331303847

Pekrun, R., & Götz, T. (2006). Emotionsregulation: Vom Umgang mit Prüfungsangst. In *Handbuch Lernstrategien* (pp. 248–258). Göttingen: Hogrefe.

Perrone-McGovern, K., Simon-Dack, S., Esche, A., Thomas, C., Beduna, K., Rider, K., … Matsen, J. (2017). The influence of emotional intelligence and perfectionism on Error-

Related Negativity: An event related potential study. *Personality and Individual Differences, 111*(January), 65–70. https://doi.org/10.1016/j.paid.2017.02.009

Pfabigan, D. M., Pintzinger, N. M., Siedek, D. R., Lamm, C., Derntl, B., & Sailer, U. (2013). Feelings of helplessness increase ERN amplitudes in healthyindividuals. *Neuropsychologia, 51*(4), 613–621. https://doi.org/10.1016/j.neuropsychologia.2012.12.008

Pfurtscheller, G., & Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalography and Clinical Neurophysiology, 42*(6), 817–826. https://doi.org/10.1016/0013-4694(77)90235-8

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology, 110*(11), 1842–1857. https://doi.org/10.1016/S1388-2457(99)00141-8

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Punaro, L., & Reeve, R. (2012). Relationships between 9-Year-Olds' Math and Literacy Worries and Academic Abilities. *Child Development Research, 2012*, 1–11. https://doi.org/10.1155/2012/359089

Putwain, D., & Daly, A. L. (2014). Test anxiety prevalence and gender differences in a sample of English secondary school students. *Educational Studies, 40*(5), 554–570. https://doi.org/10.1080/03055698.2014.953914

Putwain, D. W. (2007). Test anxiety in UK schoolchildren: prevalence and demographic patterns. *The British Journal of Educational Psychology, 77*(Pt 3), 579–93. https://doi.org/10.1348/000709906X161704

Putwain, D. W. (2008). Test anxiety and GCSE performance: the effect of gender and socio-economic background. *Educational Psychology in Practice, 24*(4), 319–334. https://doi.org/10.1080/02667360802488765

R Core Team. (2015). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*, URL http://www.R-project.org/.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*, 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

Ramirez, G., & Beilock, S. L. (2011). Writing about testing worries boosts exam performance in the classroom. *Science, 331*(6014), 211–3. https://doi.org/10.1126/science.1199427

Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale : psychometric data. *Journal of Counseling Psychology, 19*, 551–554.

Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition, 56*(2), 129–140. https://doi.org/10.1016/j.bandc.2004.09.016

Riesel, A., Weinberg, A., Endrass, T., Meyer, A., & Hajcak, G. (2013). The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks. *Biological Psychology, 93*(3), 377–385. https://doi.org/10.1016/j.biopsycho.2013.04.007

Righi, S., Mecacci, L., & Viggiano, M. P. (2009). Anxiety, cognitive self-evaluation and performance: ERP correlates. *Journal of Anxiety Disorders, 23*(8), 1132–1138. https://doi.org/10.1016/j.janxdis.2009.07.018

Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics

learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition*, *102*(3), 361–395. https://doi.org/10.1016/j.cognition.2006.01.005

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: Alpha-gamma versus theta-gamma codes for distinct WM information? *Trends in Cognitive Sciences*, *18*(1), 16–25. https://doi.org/10.1016/j.tics.2013.10.010

Running, D. M., Ligon, J. B., & Miskioglu, I. (1997). Thinking about choking? Attentional processes and paradoxical performance. *Personality and Social Psychology Bulletin*, *23*, 937–944. https://doi.org/0803973233

Sarason, I. G. (1958). Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *Journal of Abnormal and Social Psychology*, *56*(3), 339–344. https://doi.org/10.1037/h0048353

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: reactions to tests. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.46.4.929

Sarason, I. G., & Ganzer, V. J. (1962). Anxiety, reinforcement, and experimental instructions in a free verbalization situation. *Journal of Abnormal and Social Psychology*, *65*(5), 300–307. https://doi.org/10.1037/h0048977

Sarason, S. B. (1959). What research says about test anxiety in elementary school children. *NEA Journal*, *48*, 26–27.

Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R. R., & Ruebush, B. K. (1960). *Anxiety in elementary school children*. Wiley, New York.

Sarason, S. B., & Mandler, G. (1952). Some correlates of test anxiety. *J. Abnorm. Soc. Psychol.*, *47*, 810–117. https://doi.org/10.1037/h0060009

Sassenrath, J. M. (1964). A factor analysis of rating-scale items on the test anxiety questionnaire. *Journal of Consulting Psychology*, *28*(4), 371–377.

Sauseng, P., Griesmayr, B., Freunberger, R., & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1015–1022. https://doi.org/10.1016/j.neubiorev.2009.12.006

Savostyanov, A. N., Tsai, A. C., Liou, M., Levin, E. a., Lee, J.-D., Yurganov, A. V., & Knyazev, G. G. (2009). EEG-correlates of trait anxiety in the stop-signal paradigm. *Neuroscience Letters*, *449*(2), 112–116. https://doi.org/10.1016/j.neulet.2008.10.084

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*(2), 103–113. https://doi.org/10.1111/j.2041-210X.2010.00012.x

Schillinger, F. L. (2016). Event-related potentials (ERPs) reflecting feedback and error processing in the context of education. *Zeitschrift Fur Psychologie / Journal of Psychology*, *224*(4), 286–289. https://doi.org/10.1027/2151-2604/a000264

Schillinger, F. L., De Smedt, B., & Grabner, R. H. (2016). When errors count: an EEG study on numerical error monitoring under performance pressure. *ZDM - Mathematics Education*, *48*(3), 351–363. https://doi.org/10.1007/s11858-015-0746-8

Schillinger, F. L., Vogel, S. E., Diedrich, J., & Grabner, R. H. (2018). Math anxiety, intelligence, and performance in mathematics: Insights from the German adaptation of the Abbreviated Math Anxiety Scale (AMAS-G). *Learning and Individual Differences*, *61*, 109–119. https://doi.org/10.1016/j.lindif.2017.11.014

Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., & Pfurtscheller, G. (2007). A fully automated correction method of EOG artifacts in EEG recordings. *Clinical Neurophysiology*, *118*(1), 98–104. https://doi.org/10.1016/j.clinph.2006.09.003

Schroder, H. S., Moran, T. P., & Moser, J. S. (2018). The effect of expressive writing on the error-related negativity among individuals with chronic worry. *Psychophysiology*, *55*(2), 1–11. https://doi.org/10.1111/psyp.12990

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, *4*(1), 27–41. https://doi.org/10.1080/08917779108248762

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*(3), 258–275. https://doi.org/10.1037/0096-3445.117.3.258

Simons, R. F. (2010). The way of our errors: theme and variations. *Psychophysiology*, *47*(1), 1–14. https://doi.org/10.1111/j.1469-8986.2009.00929.x

Sommer, M., & Arendasy, M. E. (2014). Comparing different explanations of the effect of test anxiety on respondents' test scores. *Intelligence*, *42*(1), 115–127. https://doi.org/10.1016/j.intell.2013.11.003

Spielberger, C. D. (1972). Anxiety as an emotional state. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (pp. 23–49). New York: Academic Press.

Spielberger, C. D. (1980). Test Anxiety Inventory ("Test Attitude Inventory"). Preliminary professional manual. *Palo Alto, CA: Consulting Psychologists Press.*

Spielberger, C. D., Anton, W. D., & Bedell, J. (1976). The nature and treatment of test anxiety. In M. Zuckerman & C. D. Spielberger (Eds.), *Emotions and anxiety: New concepts, methods, and applications* (Hillsdale, pp. 317–344).

Spielberger, C. D., & Vagg, P. R. (Eds.). (1995). *Test anxiety: Theory, assessment, and treatment.* Washington, DC: Taylor & Francis.

Stöber, J., & Pekrun, R. (2004). Advances in test anxiety research. *Anxiety, Stress and Coping*, *17*(3), 205–211. https://doi.org/10.1080/1061580412331303225

Stuppy, D. J. (1998). The faces pain scale: Reliability and validity with mature adults. *Applied Nursing Research*, *11*(2), 84–89. https://doi.org/10.1016/S0897-1897(98)80229-2

Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, A. (2013). Abnormal error monitoring in math-anxious individuals: evidence from error-related brain potentials. *PloS ONE*, *8*(11), e81143. https://doi.org/10.1371/journal.pone.0081143

Szafranski, D. D., Barrera, T. L., & Norton, P. J. (2012). Test anxiety inventory: 30 years later. *Anxiety, Stress and Coping*, *25*(6), 667–677. https://doi.org/10.1080/10615806.2012.663490

Tobias, S. (1985). Test Anxiety: Interference, Defective Skills, and Cognitive Capacity. *Educational Psychologist.* https://doi.org/10.1207/s15326985ep2003_3

Trezise, K., & Reeve, R. a. (2014a). Cognition-emotion interactions: patterns of change and implications for math problem solving. *Frontiers in Psychology*, *5*(July), 840. https://doi.org/10.3389/fpsyg.2014.00840

Trezise, K., & Reeve, R. a. (2014b). Working memory, worry, and algebraic ability. *Journal of Experimental Child Psychology*, *121*, 120–136. https://doi.org/10.1016/j.jecp.2013.12.001

Turner, B. G., Beidel, D. C., Hughes, S., & Turner, M. W. (1993). Text anxiety in African American school children. *School Psychology Quarterly*, *8*(2), 140–152. https://doi.org/10.1037/h0088835

Vogel, S. E., Remark, A., & Ansari, D. (2015). Differential processing of symbolic numerical magnitude and ordr in 1st grade children. *Journal of Experimental Child Psychology*, *129*, 26–39. https://doi.org/10.1016/j.jecp.2014.07.010

Wang, C. H., Lo, Y. H., Pan, C. Y., Chen, F. C., Liang, W. K., & Tsai, C. L. (2015). Frontal midline theta as a neurophysiological correlate for deficits of attentional orienting in children with developmental coordination disorder. *Psychophysiology*, *52*(6), 801–812. https://doi.org/10.1111/psyp.12402

Weinberg, A., & Hajcak, G. (2011). Longer term test-retest reliability of error-related brain activity. *Psychophysiology*, *48*(10), 1420–5. https://doi.org/10.1111/j.1469-8986.2011.01206.x

Weinberg, A., Riesel, A., & Hajcak, G. (2011). Integrating multiple perspectives on error-related brain activity: The ERN as a neural indicator of trait defensive reactivity. *Motivation and Emotion*, *36*(1), 84–100. https://doi.org/10.1007/s11031-011-9269-y

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, *76*, 92–104. https://doi.org/10.1037/h0031332

World Health Organization. (1992). The ICD-10 Classification of Mental and Behavioural Disorders. *International Classification*, *10*, 1–267. https://doi.org/10.1002/1520-6505(2000)9:5<201::AID-EVAN2>3.3.CO;2-P

Wu, J., Yuan, Y., Duan, H., Qin, S., Buchanan, T. W., Zhang, K., & Zhang, L. (2014). Long-term academic stress increases the late component of error processing: An ERP study. *Biological Psychology*, *99*, 77–82. https://doi.org/10.1016/j.biopsycho.2014.03.002

Yu, R. (2015). Choking under pressure: the neuropsychological mechanisms of incentive-induced performance decrements. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00019

Zeidner, M. (1998). *Test anxiety: The state of the art. Test anxiety: The state of the art.* New York: Kluwer Academic Publishers.

Zeidner, M. (2007). Test Anxiety in Educational Contexts. Concepts, Findings, and Future Directions. In *Emotion in Education* (pp. 165–184). https://doi.org/10.1016/B978-012372545-5/50011-3

7. Supplementary Material

7.1. Study 1 – Response Monitoring Under Performance Pressure
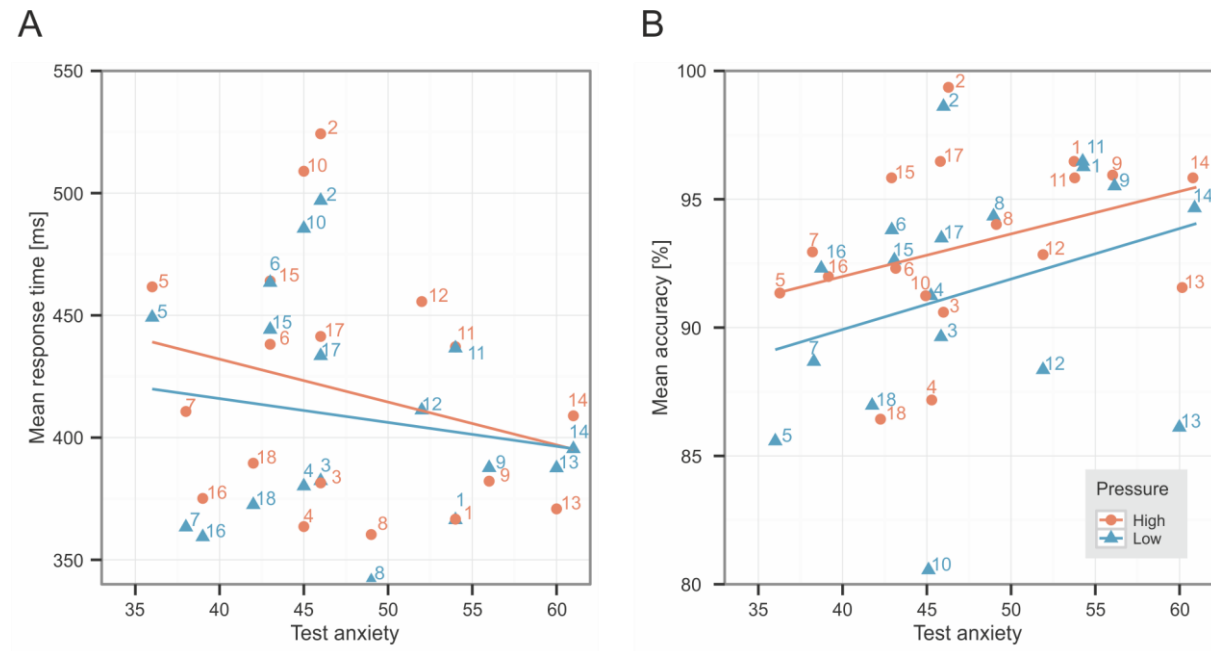
Figure S 1. Scatter plot of mean response times (A) and accuracy (B) as a function of performance pressure and test anxiety (raw values). Numbers indicate individual subjects. Linear regression lines were fitted for each pressure condition for illustrative purposes.

7.2. Study 2 – Response Monitoring in High Test-anxious Students

Supp. Material 1.

B

Fragen zum ersten Abschnitt

1.) Es gibt viele Aufgaben, bei denen die Leistung einer Person bewertet wird – mal durch Sie selbst, mal durch Andere, oder auch durch beide. Bitte geben Sie an, wie wichtig es für Sie war, eine gute Leistung bei der Zahlenvergleichsaufgabe im ersten Abschnitt zu zeigen.

☐	☐	☐	☐	☐	☐	☐
1	**2**	**3**	**4**	**5**	**6**	**7**
Gar nicht wichtig						Sehr wichtig

Bitte beschreiben Sie kurz, warum es für Sie wichtig bzw. nicht wichtig war, eine gute Leistung im ersten Abschnitt zu erbringen.

2.) Bitte geben Sie an, wie sehr Sie sich im ersten Abschnitt unter Druck gesetzt gefühlt haben, eine gute Leistung abzugeben.

☐	☐	☐	☐	☐	☐	☐
1	**2**	**3**	**4**	**5**	**6**	**7**
Gar kein Druck						Sehr viel Druck

3.) Bitte geben Sie an, wie gut Ihre Leistung im ersten Abschnitt Ihrer Meinung nach war.

☐	☐	☐	☐	☐	☐	☐
1	**2**	**3**	**4**	**5**	**6**	**7**
Sehr Schlecht						Sehr gut

3

Table S 1. List of the group assignment (LTA vs. HTA) and anxiety measures.

Group	Gender	Test anxiety			General anxiety
		Sum score	T values	Category	Sum score
LTA	m	21	21	very low	44
LTA	f	28	26	very low	43
LTA	m	25	27	very low	31
LTA	f	29	28	very low	34
LTA	m	27	29	very low	29
LTA	f	30	30	low	34
LTA	f	31	32	low	29
LTA	m	29	33	low	33
LTA	f	32	34	low	33
LTA	f	33	36	low	31
LTA	f	33	36	low	28
LTA	m	32	38	low	32
LTA	m	32	38	low	26
LTA	f	35	39	low	31
LTA	f	35	39	low	35
LTA	f	35	39	low	31
LTA	f	36	40	average	33
LTA	m	34	42	average	35
LTA	m	35	43	average	37
LTA	m	35	43	average	48
LTA	f	38	43	average	28
LTA	f	38	43	average	40
LTA	f	39	44	average	41
HTA	f	51	56	average	40
HTA	f	55	60	average	42
HTA	f	55	60	average	35
HTA	f	55	60	average	45
HTA	f	56	61	high	46
HTA	f	56	61	high	33
HTA	m	51	62	high	41
HTA	m	51	62	high	57
HTA	m	53	63	high	49
HTA	f	58	63	high	35
HTA	f	58	63	high	32
HTA	m	56	65	high	60
HTA	m	56	65	high	55
HTA	m	56	65	high	42
HTA	f	60	65	high	48
HTA	m	58	67	high	58
HTA	m	60	68	high	38
HTA	f	63	68	high	46
HTA	f	65	69	high	58
HTA	m	65	72	very high	58
HTA	f	70	74	very high	40

Note. Scores are sorted by test anxiety t values. m = male, f = female, LTA = low test anxiety, HTA = high test anxiety. Categories based on t values with < 30 = very low, 30-39 = low, 40-60 = average, 61-70 = high, >70 = very high.

Table S 2. Pairwise comparisons of the means between two conditions (I and J) with the peak amplitude (in μV) as dependent variable.

	I	J	Mean difference	*SE*	*p*
Low Pressure					
CRN	LTA	HTA	-0.29	1.13	.800
ERN	LTA	HTA	-2.25	1.41	.119
High Pressure					
CRN	LTA	HTA	0.17	1.15	.882
ERN	LTA	HTA	0.36	1.45	.807
LTA					
CRN	Low Pressure	High Pressure	-0.13	0.39	.751
ERN	Low Pressure	High Pressure	0.38	0.76	.623
HTA					
CRN	Low Pressure	High Pressure	0.34	0.41	.418
ERN	Low Pressure	High Pressure	2.98	0.80	< .001
LTA					
Low Pressure	CRN	ERN	4.51	0.70	<.001
High Pressure	CRN	ERN	5.00	0.64	<.001
HTA					
Low Pressure	CRN	ERN	2.55	0.73	.001
High Pressure	CRN	ERN	5.19	0.67	<.001

Note. LTA = low test anxiety, HTA = high test anxiety, SE = Standard error, CRN = correct response negativity, ERN = error-related negativity.

Table S 3. Result summary of Analysis of Variance (ANOVA) within HTA students including the factor order (high pressure condition first vs. low pressure condition first) with the peak amplitude (in μV) as dependent variable.

	$F(1,19)$	*p*	η_p^2
Main Effects			
Response Type	48.31	< .001	.718
Pressure	7.34	.014	.279
Order	1.23	.281	.061
Interactions			
Response Type x Pressure	7.00	.016	.269
Response Type x Order	0.97	.337	.049
Pressure x Order	1.10	.308	.055
Response Type x Pressure x Order	0.95	.342	.048

Note. η_p^2 = Partial Eta Squared, HTA = high test anxiety.

7.3. Study 3 – Working Memory Costs of Performance Pressure

Table S 4. Pearson correlations between test anxiety and (a) response time, (b) accuracy, (c) worry, and (d) FMΘ for each group, block, and demand.

		Block 1		Block 2	
		Low Demand	High Demand	Low Demand	High Demand
(a) Response Time	Control	.36*	.46**	.23	.43*
	Pressure	-.24	-.26	-.46*	-.20
(b) Accuracy	Control	-.01	-.04	.23	-.17
	Pressure	-.03	-.20	-.08	-.16
(c) Worry	Control	.32	.27	.22	.27
	Pressure	.33	.40*	.42*	.45*
(d) FMΘ	Control	-.05	.07	.12	.18
	Pressure	-.10	-.32	.35	.09

Note. * = p < .05; ** = p < .01; *** = p < .001.

Table S 5. Standardized regression coefficients, standard errors, and model summaries (R^2 and F statistics) for the mediation models predicting the average change in FMΘ based on test anxiety. The change in worries between blocks was entered as a mediator variable.

Model	Variables	*Path*	*β*	SE	*p*	R^2	F statistics
(a)	Test anxiety						
	ΔWorries	a	-.06	.19	.747	<.01	$F_{(1,29)} = 0.11$, *p* = .747
	ΔFMΘ						
	ΔWorries	b	-.32	.17	.073		
	Test anxiety	c'	.19	.17	.273	.15	$F_{(2,28)} = 2.49$, *p* = .101
(b)	Test anxiety						
	ΔWorries	a	.07	.20	.715	.01	$F_{(1,27)} = 0.14$, *p* = .715
	ΔFMΘ						
	ΔWorries	b	.05	.15	.762		
	Test anxiety	c'	.60	.16	<.001	.25	$F_{(2,26)} = 7.68$, *p* = .002

Note. FMΘ = Fontal midline theta
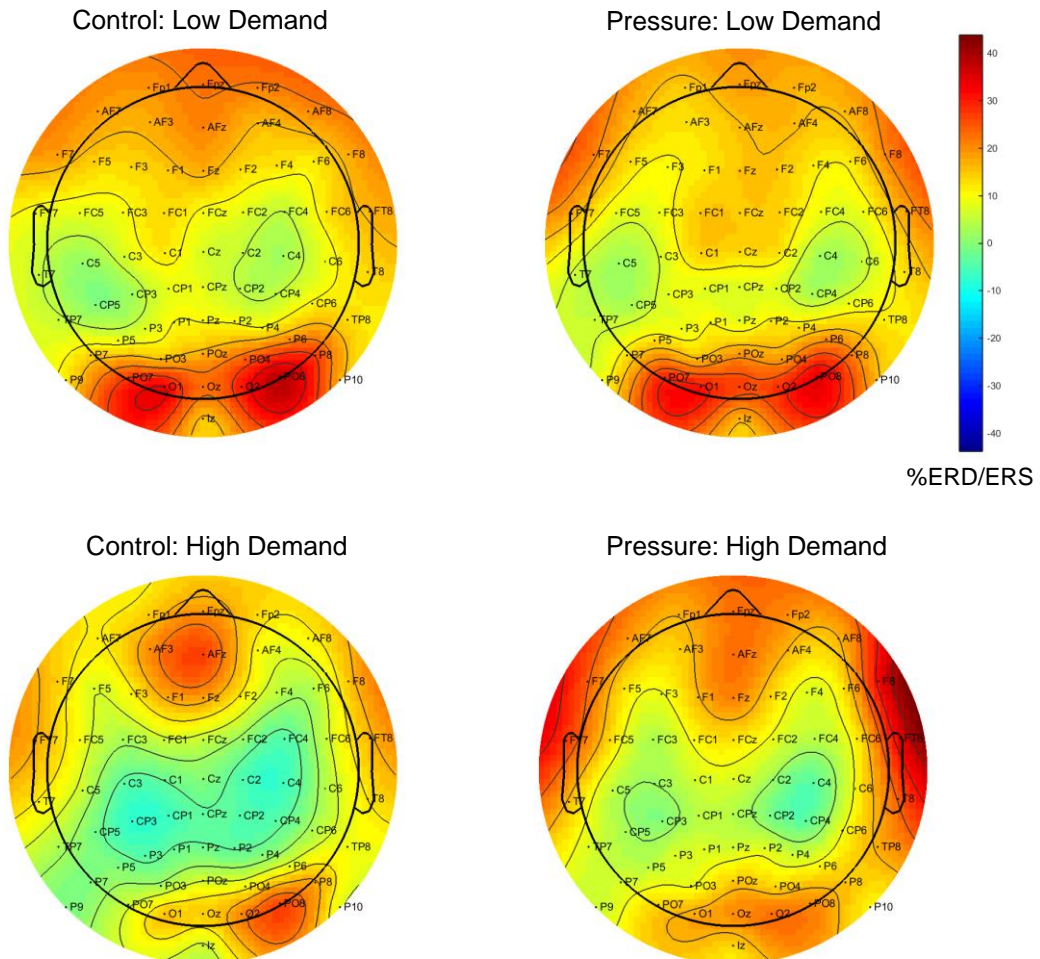
%ERS/ERD [4-8 Hz]
Block 1

Figure S 2. Topographical maps of ERS/ERD (%) for groups (control, pressure) and demands (low, high) during block 1 in the theta frequency band (4–8 Hz). Warm colors represent an increase in theta (ERS) and cold color a decrease in theta (ERD) relative to the reference interval.

%ERS/ERD [4-8 Hz]
Block 2

Control: Low Demand

Pressure: Low Demand

%ERD/ERS
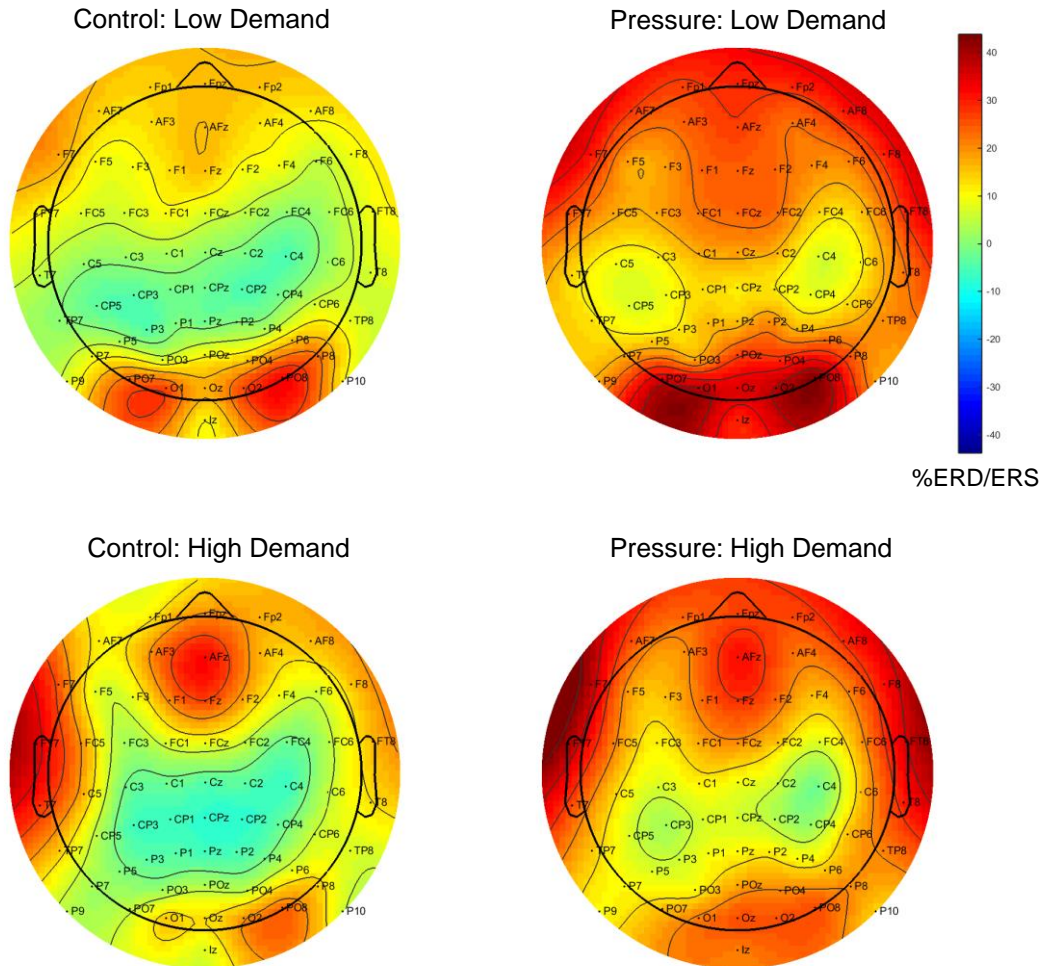
Control: High Demand

Pressure: High Demand

Figure S 3. Topographical maps of ERS/ERD (%) for groups (control, pressure) and demands (low, high) during block 2 in the theta frequency band (4–8 Hz). Warm colors represent an increase in theta (ERS) and cold color a decrease in theta (ERD) relative to the reference interval.

Block 1: [Pressure - Control]
theta band (4–8 Hz)

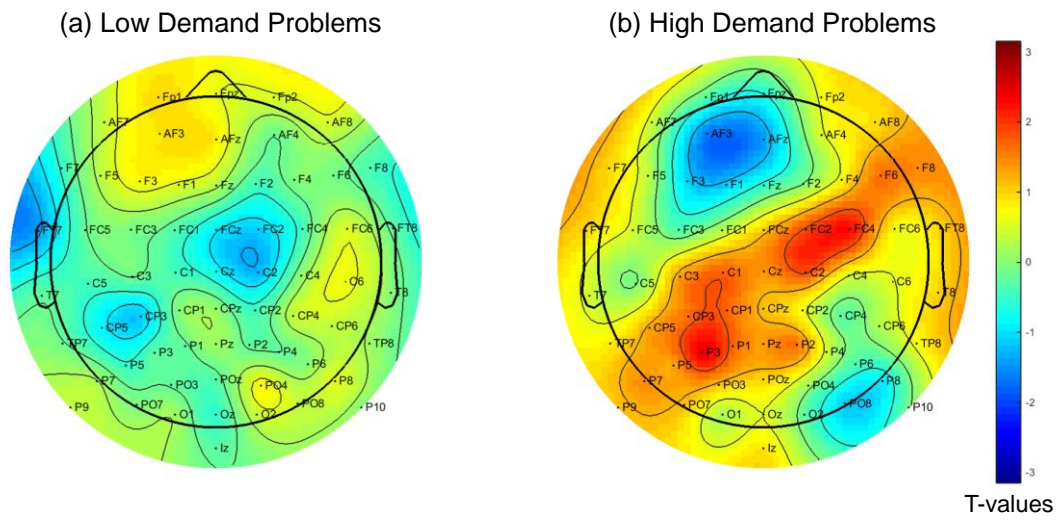(a) Low Demand Problems　　　　　　　(b) High Demand Problems

T-values

Figure S 4. Topographical maps of t-values corresponding to the contrast [Pressure - Control] in block 1 for (a) low demand and (b) high demand problems with data filtered in the theta frequency band (4–8 Hz). Analyses revealed no significant cluster for either demand.

Curriculum Vitae

Vorname:	Frieder Leon
Nachname:	Schillinger
Geburtsdatum:	01.02.1984
Staatsbürgerschaft:	Deutsch
Email:	frieder.schillinger@gmail.com

Akademische Ausbildung

12|2013 – Aktuell	Promotion (Dr.rer.nat) in der Georg-August University School of Science (GAUSS) der Georg-August Universität Göttingen
04|2004 – 07|2011	Studium der Psychologie auf Diplom an der Eberhart Karls Universität Tübingen
04|2006 – 10|2011	Studium der Philosophie auf Bakkalaureus Artium (B.A) an der Eberhart Karls Universität Tübingen
04|2003 – 04|2004	Studium generale am Leibniz Kolleg, Tübingen

Arbeitserfahrung

04|2015 – 12|2017	Wissenschaftlicher Mitarbeiter in der Abteilung für Begabungsforschung, Karl-Franzens-Universität Graz
03|2013 – 04|2015	Wissenschaftlicher Mitarbeiter in der Abteilung für Pädagogische Psychologie, Georg-August-Universität Göttingen
10/2009 – 05/2011	Wissenschaftliche Hilfskraft in der Abteilung Wahrnehmung, Kognition und Handlung, Max-Planck-Institut für biologische Kybernetik, Tübingen
03/2009 – 09/2009	Wissenschaftliche Hilfskraft in der Abteilung Sozial-motivationale Prozesse, Leibniz-Institut für Wissensmedien (IWM), Tübingen

Forschungsaufenthalt im Ausland

11/2011 – 07/2012	Forschungsaufenthalt im Neuroimaging Laboratory, Santa Lucia Foundation und an der La Sapienza Universität, Rom, Italien
04/2008 – 10/2008	Praxissemester (6 Monate) am National Institute of Mental Health, Unit on Cognitive Neurophysiology and Imaging, Bethesda, USA

Peer-reviewed

Schillinger, F. L., Vogel, S. E., Diedrich, J., Grabner, R. H. (2018). Math anxiety, intelligence, and performance in mathematics: Insights from the German adaptation of the Abbreviated Math Anxiety Scale (AMAS-G). Learning and Individual Differences, 61, 109-119.

Schillinger, F. L. (2016). Event-related potentials (ERPs) reflecting feedback and error processing in the context of education. Zeitschrift für Psychologie, 224(4), 286–289. http://doi.org/10.1027/2151-2604/a000264

Schillinger, F. L., De Smedt, B., & Grabner, R. H. (2016). When errors count: an EEG study on numerical error monitoring under performance pressure. ZDM - Mathematics Education, 48(3), 351–363. http://doi.org/10.1007/s11858-015-0746-8

de la Rosa*, S., Schillinger, F. L.*, Bülthoff, H. H., Schultz, J., & Uludag, K. (2016). fMRI Adaptation between Action Observation and Action Execution Reveals Cortical Areas with Mirror Neuron Properties in Human BA 44/45. Frontiers in Human Neuroscience, 10, 1–10. http://doi.org/10.3389/fnhum.2016.00078 (*Authors contributed equally)

Müller, K. M., Schillinger, F., Do, D. H., & Leopold, D. A. (2009). Dissociable perceptual effects of visual adaptation. PLoS ONE, 4(7). http://doi.org/10.1371/journal.pone.0006183

Diplomarbeit

Schillinger, F. L. (2010). Whole-brain fMRI using repetition suppression between action and perception reveals cortical areas with mirror neuron properties, Eberhard-Karls-Universität Tübingen. Link

Konferenzbeiträge

Schillinger, F. L., Vogel, S. E., Diedrich, J., Grabner, R. H. (2017). Math Anxiety Is Related to Performance in Mathematics Beyond Individual Differences in Numerical Intelligence and Arithmetic Abilities: Insights from the German Adaptation of the Abbreviated Math Anxiety Scale (AMAS). International Convention of Psychological Science (ICPS). Vienna, Austria.

Schillinger, F. L., Vogel, S. E., Koschutnig, K., Ansari, D., Grabner, R. H. (2016). Brain response to arithmetic errors is modulated by individual differences in mathematical competence in the inferior frontal gyrus (IFG). Conference of the Special Interest Group (SIG) 22 "Neuroscience and Education" of the EARLI. Amsterdam, Netherlands.

Schillinger, F. L., De Smedt, B. & Grabner, R. H. (2015). When errors count: An EEG study investigating error processing under performance pressure. Conference of the European Association for Research on Learning and Instruction (EARLI). Limassol, Cyprus.

Schillinger, F. L., Hinze, A., de Smedt, B., Grabner, R. H. (2014). When errors count: An EEG study investigating arithmetic error processing under performance pressure. Conference of Junior Neuroscientists of Tübingen (NeNa). Schramberg, Germany.

Schillinger, F. L., de la Rosa, S., Schultz, J., & Uludag, K. (2010). Whole-brain fMRI using repetition suppression between action and perception reveals cortical areas with mirror neuron properties, Perception, 39 (ECVP Abstract Supplement) 54.

Schillinger, F., Müller, K., Do, D. H. & Leopold, D. A. (2008). Dissociable perceptual effects of visual adaptation. National Institute of Mental Health (NIMH) Retreat, Gettysburg, USA.

Lehrerfahrung

Lehrtätigkeit an der Karl-Franzens-Universität Graz

Sommersemester 2017	Übung im Bachelor Psychologie: Statistische Verfahren am Computer (2 SWS)
Sommersemester 2016	Übung im Bachelor Psychologie: Statistische Verfahren am Computer (2 SWS)

Lehrtätigkeit an der Georg-August-Universität Göttingen

Wintersemester 2014/15	Seminar im Bachelor Psychologie: Förderung individueller und institutioneller Lehr-Lern-Prozesse (2 SWS)
Sommersemester 2014	Seminar im Bachelor Psychologie: Pädagogisch-psychologische Diagnostik und Beratung (2 SWS)
Wintersemester 2013/14	Seminar im Bachelor Psychologie: Förderung individueller und institutioneller Lehr-Lern-Prozesse (2 SWS)
Sommersemester 2013	Seminar im Bachelor Psychologie: Pädagogisch-psychologische Diagnostik und Beratung (2 SWS)

Lehrtätigkeit an der privaten Pädagogischen Hochschule der Diözese Linz

Wintersemester 2017/18	Seminar im Master „Neurowissenschaften und Bildung": Prüfungsangst: Ursachen, Mechanismen und Interventionsmöglichkeiten (2 SWS kompakt, 01/02.12.2017)