The Journal of Experimental Education

ISSN: 0022-0973 (Print) 1940-0683 (Online) Journal homepage: http://www.tandfonline.com/loi/vjxe20

Taming Test Anxiety: The Activation of Failure-Related Concepts Enhances Cognitive Test Performance of Test-Anxious Students

Tobias Tempel & Roland Neumann

To cite this article: Tobias Tempel & Roland Neumann (2016): Taming Test Anxiety: The Activation of Failure-Related Concepts Enhances Cognitive Test Performance of Test-Anxious Students, The Journal of Experimental Education

To link to this article: http://dx.doi.org/10.1080/00220973.2015.1094649

Published online: 19 Feb 2016.

THE JOURNAL OF EXPERIMENTAL EDUCATION, *00*(00), 1–21, 2016

ISSN: 0022-0973 print/1940-0683 online
DOI: 10.1080/00220973.2015.1094649

Routledge
Taylor & Francis Group

MOTIVATION AND SOCIAL PROCESSES

Taming Test Anxiety: The Activation of Failure-Related Concepts Enhances Cognitive Test Performance of Test-Anxious Students

Tobias Tempel and Roland Neumann
University of Trier, Germany

We investigated processes underlying performance decrements of highly test-anxious persons. Three experiments contrasted conditions that differed in the degree of activation of concepts related to failure. Participants memorized a list of words either containing words related to failure or containing no words related to failure in Experiment 1. In Experiment 2, all participants initially memorized words related to failure, but one experimental condition subsequently required the intentional forgetting of these words, thereby, decreasing the words' activation in memory. The performance of highly test-anxious participants in unannounced cognitive ability tests following the memorizing phase profited from the activation of failure-related concepts. Experiment 3 replicated this finding by contrasting a condition priming failure–related concepts with a neutral control condition. The results point to a mechanism of suppressing worrying thoughts that is detrimental to test performances of highly test-anxious persons. Bringing the feared threat of failure to their awareness prevents that suppression.

Keywords *achievement, anxiety, emotion, individual differences, personality, testing*

TESTS ARE UBIQUITOUS. Everybody takes tests in school, college, or other contexts. They are not only an important tool for the assessment of abilities and efforts, they can also have beneficial effects, such as enhancing memory (Roediger & Karpicke, 2006). They also involve downsides, however, one being the pervasive problem of test anxiety. Highly test-anxious persons not only are distressed when they face tests, in general, they also perform below their ability level.

Test anxiety can be considered as both a trait and a state. The stronger trait test anxiety of a person, the stronger he or she is distressed when facing a testing situation (state test anxiety). Although there are important situational specificities and different domain-specific kinds of test

Address correspondence to Tobias Tempel, Fachbereich I—Psychologie, University of Trier, 54286 Trier, Germany. E-mail: tempel@uni-trier.de

anxiety can be distinguished (e.g. mathematics anxiety), researchers have sought to identify the general dimensions of test anxiety. There are two key features representing well-validated factors of inventories to assess test anxiety (Liebert & Morris, 1967; Schwarzer, 1984; Spielberger, 1980; Wacker, Jaunzeme, & Jaksztat, 2008; Ware, Galassi, & Dew, 1990): *emotionality*, which comprises anxious feelings and physiological arousal, and the cognitive component of *worry*. These two components of test anxiety correlate differentially with test outcomes. Worry predicts performance better than emotionality (Hembree, 1988; Seipp, 1991). The causal link between test anxiety and performance impairments typically is demonstrated experimentally, by presenting a task—for example solving verbal analogies as a test of ability (diagnostic condition). Performance then is compared to a condition in which that same task is presented in a nondiagnostic manner—that is, without mentioning the possibility of drawing inferences about the individual's ability or with instructions emphasizing the nondiagnosticity (e.g., Sarason, 1978). In this design, worry reliably predicts the amount of reduced performance in the diagnostic condition as compared to the nondiagnostic condition. Emotionality, in contrast, is associated with the actual impaired performance to a lesser degree (e.g., Deffenbacher, 1978; Morris, Davis, & Hutchings, 1981; Zeidner, 1998). Moreover, other components of text anxiety have been examined but have not turned out to predict performance decrements as well as worry. For example, it has been suggested that task-irrelevant thinking constitutes a further cognitive component but it has been found to correlate less strongly with performance than worry (e.g. Sarason, 1984). Furthermore, evidence exists that task-irrelevant thinking might not represent a separate dimension of test anxiety but should be considered as a correlate of test anxiety (Benson, Moulin-Julian, Schwarzer, Seipp, & El-Zahhar, 1992; Hodapp & Benson, 1997).

Different processes have been suggested as leading to the performance decrement. For example, Eysenck (1979; see also Eysenck & Calvo, 1992) proposed that worrying thoughts of highly test-anxious persons absorb working memory capacity. As a consequence, especially those cognitive tasks that require complex operations cannot be attended to efficiently. Several studies support the assumption of reduced working memory capacity of highly test-anxious persons (e.g., Darke, 1988; Derakshan & Eysenck, 1998; Dutke & Stöber, 2001; Ikeda, Iwanaga, & Seiwa, 1996; MacLeod & Donnellan, 1993; Richards, French, Keogh, & Carter, 2000). Similarly, related phenomena, such as, choking under pressure and stereotype threat, also have been shown to involve reduced working memory capacity available for task performance (e.g., Beilock, Kulp, Holt, & Carr, 2004; Schmader & Johns, 2003). However, it is less clear which processes account for this reduction. The seemingly widespread belief of a distraction by worrisome thoughts lacks sufficient empirical support. On the one hand, there are plenty of studies showing that evaluative situations can trigger worries in test-anxious individuals. On the other hand, it has repeatedly been shown that experimentally induced worries reduce working memory performance (for reviews see Eysenck, Derakhshan, Santos, & Calvo, 2007; Zeidner, 1998). However, evidence on the crucial link between worries of test-anxious individuals and their impact on working memory capacity is scarce. Typically, studies that aim at examining whether worries mediate test-anxiety-related performance decrements rely on questionnaires assessing anxiety and worry prospectively or in retrospect but do not measure the amount of worries triggered in an evaluative situation directly nor involve manipulations of worry (e.g., Owens, Stevenson, Hadwin, & Norgate, 2012).

A recent study by Lang and Lang (2010) actually investigated the mediation of performance decrements of test-anxious persons by the amount of worrisome thoughts. Worrisome thoughts,

measured immediately after the test, did not turn out to be a significant mediator. Since empirical evidence is missing that worrisome thoughts would consume working memory capacity of test-anxious persons via distraction, although this assumption was maintained over decades in research on test anxiety, it is about time to consider alternatives. The assumption that the active suppression of worry engrosses working memory capacity is such an alternative. Perhaps, worrisome thoughts account for reduced working memory capacity not via distraction but the suppression of worrisome thoughts instead. Probably, suppression usually would not manifest itself in measures of worrisome thoughts because these measures would yield equally low rates of worrisome thoughts for participants who suppress worry and participants who do not worry at all, which, therefore, would preclude significant correlations with outcome measures, such as working memory tasks or ability tests.

Yet, it is a well-established fact by now that thought suppression is an effortful process that depletes cognitive resources (e.g., Muraven & Baumeister, 2000; Wenzlaff & Wegner, 2000). We assume that suppression of worrisome thoughts also accounts for performance decrements of highly test-anxious persons. Support for this assumption comes from different fields of research. In clinical contexts, thought suppression has been linked to the development and perpetuation of various anxiety disorders (Purdon, 1999). Social psychological research on the phenomenon of stereotype threat showed that it involves suppression of stereotype-related thoughts. Using an implicit measure of concept activation, Brodish and Devine (2009) and Carr and Steele (2009) showed that concepts related to a negative stereotype about a social group that participants belonged to were suppressed by participants that worked on a cognitive ability test under stereotype threat. With regard to test anxiety, a recent study by Ramirez and Beilock (2011) demonstrated that a short expressive writing task immediately before a test enhanced the performance of highly test-anxious students to the level of slightly test-anxious fellow students. Ramirez and Beilock assume that writing enabled a reevaluation of the situation, thereby, stimulating the insight in the disproportionalness of extreme amounts of worry. Intrusions by worrying thoughts and rumination were diminished, setting free working memory capacity that could be used for the efficient work on the test. However, an alternative to this interpretation is that writing about worries was helpful because it eliminated the suppression of worry. In accordance with this interpretation, here, we investigated if manipulations aimed directly at eliminating the suppression of worrisome thoughts benefit test performance of persons with high test anxiety.

How to counteract thought suppression? It is pretty straightforward to assume that the activation of suppressed concepts may prevent suppression processes. In fact, this idea has been realized by many approaches in the treatment of anxiety (e.g., Hittner & Hemmo, 2009; Purdon, 1999; Wegner, 1994). Bringing to light cognitions that one usually would suppress or try to suppress is a core element in therapeutic approaches to reduce thought suppression in clinical contexts—for example, through psychoeducation or mindfulness trainings by bringing awareness to negative thoughts normally avoided and suppressed (e.g., Cohen, Wupperman, & Tau, 2013; Hepburn et al., 2009; Marino-Carper, Negy, Burns, & Lunt, 2010). Also, acceptance-based therapy (Hayes, Strosahl, & Wilson, 1999) has been shown to be an effective means to treat test anxiety specifically (Brown et al., 2011; Zettle, 2003). The activation of usually suppressed concepts generally might act as a confrontation that weakens experiential avoidance strategies such as thought suppression. Similarly, awareness of experiencing feared emotions has proven to be an effective means against emotional suppression and its negative consequences (e.g., Levitt,

Brown, Orsillo, & Brown, 2004). Research on stereotype threat also lends support to the expectation that manipulations counteracting thought suppression by activating the suppressed content itself can increase test performance or prevent performance decrements. For example, teaching participants about stereotype threat before a test enhances performance—that is, eliminates performance decrements (Johns, Schmader, & Martens, 2005). Also, replacing stereotypic thoughts by alternative content enhances test performance; whereas instructions to suppress a stereotype in question even intensifies stereotype threat-induced performance decrements (McGlone & Aronson, 2007).

Any manipulation that counteracts the suppression of worry should enhance test performances of highly test-anxious persons if this suppression caused performance decrements. This assumption matches the multitude of previous studies suggesting that worries triggered in evaluative situations account for performance decrements of test-anxious individuals. However, we assume that it is not distraction but suppression of worrying thoughts that accounts for these decrements. In three experiments, we activated concepts related to failure with the intention to eliminate a suppression of worrying thoughts. Since failure is the threat that high test-anxious persons worry about, those concepts can be considered as the cognitive core of worry. The activation of concepts usually being suppressed by highly test-anxious participants should render suppression attempts impossible. As in mindfulness trainings, directing awareness to the content of the feared threat was expected to free participants from the burden of effortful thought suppression. The activation of concepts related to failure involved reading words such as *flop*, *disappointment*, or *shame* before working on an ability test. These words were either embedded in a to-be-memorized list of words (Experiments 1 and 2) or in a word-search puzzle (Experiment 3).

EXPERIMENT 1

In Experiment 1, we manipulated the activation of concepts related to failure by presenting a list of to-be-remembered words. In one experimental condition the list contained words related to failure among others; whereas the list contained only neutral words in the other condition. Participants afterward worked on verbal analogies that were presented as being diagnostic of intelligence. Solving verbal analogies is a standard task used in research on test anxiety (e.g., Lang & Lang, 2010; Lee, 1999). To-be-completed–word stems of initially learned words were interspersed in these analogies. Trait test anxiety was measured at the beginning of the experimental session by the German Test Anxiety Inventory (TAI-G; Hodapp, 1991). At the end of the session, participants retrospectively rated one-item measures of anxiety and worry and a motivational item of self-standards. The self-standards item was included because it has been suggested that effects of test anxiety on test performance can be mediated by motivational processes, not only by anxiety- or worry-related processes (e.g., Lang & Lang, 2010). We restricted the assessment of these variables to single items in order to minimize influences of response tendencies between the measures. Similar one-item measures have been used successfully in previous investigations (e.g., Davey, Barratt, Butow, & Deeks, 2007; Tempel & Neumann, 2014).

We examined the hypothesis that the activation of failure-related concepts benefited test performance of participants high in trait worry. We expected the learning of words related to failure at the beginning to enhance the performance of highly test-anxious participants at solving the analogies. In addition, we explored whether it would be possible to model the assumed

suppression mechanism by varying the instructions for the completion of the word stems. The instructions asked participants to complete the word stems by either using any words that came to mind or using words other than those used in the previous learning phase. We assumed instructed suppression would decrease performance equally to suppression due to trait test anxiety. Thus, we expected a three-way interaction. The learning of words related to failure should not enhance the performance of highly test-anxious participants instructed to use novel words but should enhance the performance of highly test-anxious participants who were free to complete the stems using any words.

METHOD

Participants and Design

Eighty-three psychology students (67 women) at the University of Trier participated in the experiment. Two factors were manipulated between subjects: Concept activation (failure activation, neutral) and word stem completion (free, constrained). Participants were randomly assigned to the four resulting conditions. In addition, trait test anxiety was assessed by the TAI-G. A median split, performed separately for men and women, by the score on the worry scale of the TAI-G divided the sample into two groups (low worry, high worry). The sample had a considerably larger range in age (17 to 31) than most studies on test anxiety, which often use pupil samples of one specific class level. Since trait test anxiety is known to correlate with age (e.g., Wacker et al., 2008), as do cognitive performance measures, we included age as a covariate in all analyses in order to statistically control for any potential confounding.

Measures

The software PXLab (Irtel, 2007) served for programming. The experiment was conducted using Dell Optiplex 755 PCs with Eizo FlexScan S1901 monitors and standard German QWERTZ keyboards. All instructions and all items were given on the screen. Participants entered their answers via the keyboard and mouse.

Test Anxiety

The TAI-G (Hodapp, 1991) comprises four scales. Ten items measure worry (e.g., "I am thinking about the consequences of failing."), eight items measure emotionality (e.g., "My heart is pounding."), six items measure interference (e.g., "I am preoccupied by other thoughts, and thus distracted."), and six items measure lack of confidence (e.g., "I am confident concerning my own performance," reverse coded). Participants indicate how they usually feel and what they are thinking in testing situations by rating the items on a 4-point-scale (*almost never*, *seldom*, *often*, *almost always*). The interference scale and the lack of confidence scale represent extensions of the test-anxiety construct that have been shown to correlate differently with specific other self-report measures, compared to worry or emotionality (e.g., Stöber, 2004). However, direct evidence of associations of the two ancillary scales to performance outcome measures is as yet missing.

Verbal Analogies

We chose 20 analogies from the analogy subtest of the German intelligence structure test by Amthauer, Brocke, Liepmann, and Beauducel (1999; IST 2000). Participants had to select one out of five options in answering each analogy, by pressing a number between one and five on the keyboard. The time per analogy was limited. If participants did not give an answer within 10 seconds, the analogy was scored as incorrect.

State Measures

Participants rated their anxiety on three separate items with an eight-point scale (How afraid have you been of the word stems and analogies? 0 = *not at all*, 7 = *extremely*), their worry (How much did you worry about your performance? 0 = *not at all*, 7 = *extremely*), and a motivational item of self-standards (How good did you want to perform at the word stems and analogies? 0 = *average*, 7 = *excellent*)[1] during the verbal ability test.

Procedure

The experiment was conducted in groups of one to three participants who were tested in separate cubicles. First, participants answered the TAI-G. In the subsequent learning phase, participants were asked to memorize words. Each of these 24 words was presented for four seconds. The failure activation condition contained six words that were synonymous or semantically closely related to failure (*flop, failure, disappointment, break down, collapse, debacle*).[2] The neutral condition contained six words unrelated to achievement but to construction materials instead (*synthetical, cardboard, stone, cement, metal, plastic*) that were matched on word frequency to the words related to failure. The remaining 18 words were identical in the two conditions and neither were related to achievement. After the learning phase, participants were instructed to work on a test of verbal ability comprising two different tasks, one being verbal analogies, the other involving the completion of word stems. These word stems were to be completed either by any word that came to mind (free completion) or only by words that had not been presented previously (constrained completion). The test was described as assessing individual strengths and weaknesses in verbal ability very exactly by relying on accuracy as well as speed of answers. The items of the word stem completion and the analogies tasks were interspersed with each other. Lastly, participants rated their anxiety, worry, and self-standards during the verbal ability test and provided demographic data (major, age, gender). A final question asked about speculations on the investigated hypotheses.

RESULTS AND DISCUSSION

A 2 (high trait worry, low trait worry) × 2 (failure activation, neutral) × 2 (free word stem completion, constrained word stem completion) between subjects ANCOVA with age as the covariate examined the number of correctly solved analogies. The main effect of trait worry was not statistically significant, $F(1,74) = 2.49$, $p = .119$, neither was the main effect of word stem completion condition, $F(1, 74) = 3.06$, $p = .084$, nor the main effect of concept activation, F

< 1. However, the interaction of trait worry and concept activation was statistically significant, $F(1, 74) = 5.80$, $p = .018$, $\eta_p^2 = .07$. The performance of participants who were high in trait worry was better in the failure activation condition than in the neutral condition; whereas the performance of participants low in trait worry was slightly worse in the failure activation condition than in the neutral condition (see Figure 1, upper section). None of the other interactions was statistically significant, Fs < 1.[3] The nonsignificant, three-way interaction of trait worry, word stem completion, and item set shows that we did not model a suppression process by instructing participants to use novel words to complete the word stems. Obviously, this instruction activated concepts related to failure in the same manner as the instruction to use any words that came to mind. We assume that participants, in fact, had to explicitly think about previously presented words in order to think of alternatives for the completion task. Hence, word stem completion did not moderate the interaction of item set and trait worry, but the initial learning of words related to failure enhanced performance of participants high in trait worry independently of the later alternative of the word stem completion task.[4]

We examined the ratings of self-standards, anxiety, and worry during the test by 2 (high trait worry, low trait worry) $\times$ 2 (failure activation, neutral) $\times$ 2 (free word-stem completion, constrained word-stem completion) ANCOVAs with age as the covariate. There were no statistically significant effects with regard to self-standards, Fs < 2.33, $p > .131$. The mean self-standards rating was 4.6 ($SD = 1.6$). With regard to anxiety, the main effect of concept activation condition was statistically significant, $F(1, 74) = 4.89$, $p = .030$, $\eta_p^2 = .06$, but no other main effect or interaction was, Fs < 2.21, $p > .142$. The anxiety ratings were lower in the failure activation condition ($M = 1.1$, $SD = 1.4$) than in the neutral condition ($M = 1.9$, $SD = 1.7$). With regard to the worry ratings, the main effect of concept activation condition was also statistically significant, $F(1, 74) = 4.73$, $p = .033$, $\eta_p^2 = .06$, as were the main effect of trait worry, $F(1, 74) = 4.77$, $p = .032$, $\eta_p^2 = .06$, and the main effect of word-stem completion condition, $F(1, 74) = 4.69$, $p = .034$, $\eta_p^2 = .06$. The ratings were lower in the failure activation condition than in the neutral condition; they were higher with the free word-stem completion condition than with constrained word-stem completion, and participants high in trait worry gave higher ratings than participants low in trait worry. Also, the interaction of concept activation condition and trait worry was statistically significant, $F(1, 74) = 4.61$, $p = .035$, $\eta_p^2 = .06$; whereas the interaction of word-stem completion condition and trait worry only approached significance, $F(1, 74) = 3.88$, $p = .053$. Whereas the ratings of worry during the test were almost identical for participants high and low in trait worry in the neutral condition, participants high in trait worry rated worry during the test higher than participants low in trait worry in the failure activation condition (see Table 1). These higher ratings indicate that participants high in worry, in fact, did not suppress worrying thoughts in the failure activation condition. Indeed, higher worry ratings alternatively also might indicate unsuccessful suppression attempts or distraction by worrying thoughts. However, both suppression attempts and distraction would imply poorer performance of participants high in trait worry in the failure activation condition. Yet, the opposite effect occurred. Higher worry ratings accompanied better performance. Additionally, the low ratings of participants low in trait worry suggest that failure activation might actually have an opposite influence on them.

Taken together, the initial learning of words related to failure yielded better performance of highly test-anxious participants compared to highly test-anxious participants in the neutral condition. This finding matches the assumption that test performances of highly test-anxious

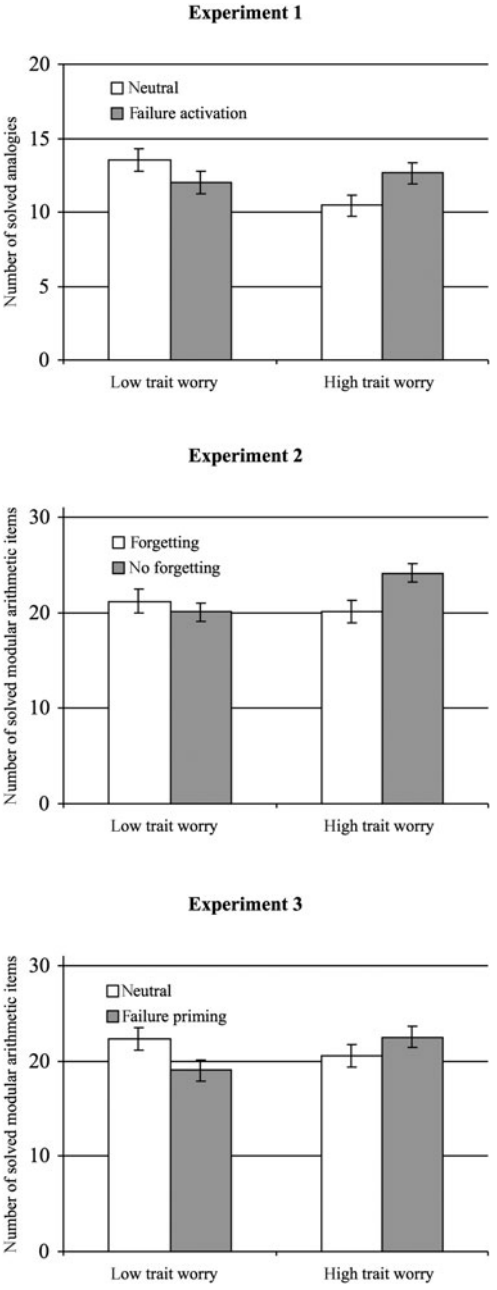

FIGURE 1 The effects of trait worry and the activation of concepts related to failure on test performance in the three experiments: Experiment 1—the effects of trait worry and concept activation condition on performance in the verbal analogies test; Experiment 2—the effects of trait worry and forgetting condition on modular arithmetic performance; Experiment 3—the effects of trait worry and priming condition on modular arithmetic performance.

TABLE 1
Mean Ratings of Worry During the Test in Dependence of Experimental Condition and Trait Worry in Experiments 1 and 3

	Low trait worry	*High trait worry*
Experiment 1		
Neutral	3.2	3.2
Failure activation	1.4	3.2
Experiment 3		
Neutral	2.7	2.1
Failure priming	1.3	3.6

persons suffer from the active suppression of worrying thoughts and that the activation of concepts related to failure is able to eliminate this suppression process.

EXPERIMENT 2

Experiment 1 provided first evidence for the counterintuitive idea that activating concepts related to failure enhances test performances of highly test-anxious persons. In Experiment 2, we examined this idea further by manipulating concept activation in a different way. We transferred a paradigm from experimental memory research for this purpose.

Directed forgetting is the phenomenon in which the intention to forget previously learned information actually renders this information less accessible compared to information that is not intended to be forgotten. There are different experimental paradigms in investigating directed forgetting. In the list method, participants first receive a list of to-be-remembered items, usually presented one after another. After several items (list 1) are presented, half of the participants are instructed to forget the items presented thus far and to remember only those items that are to be presented subsequently (list 2), whereas the other half receives no instruction to forget. After list 2 has been presented, a test assesses memory for all items. Directed forgetting is present if participants given an instruction to forget recall significantly fewer list 1 items than participants who have been given no forgetting instruction. Current theoretical models explain this effect either by assuming a change in mental context (Sahakyan & Kelley, 2002) or by an inhibitory mechanism that lowers the accessibility of list 1 (Bjork, 1989). We utilized the list method to decrease the activation of concepts related to failure and thus to manipulate suppression. The initial activation of concepts related to failure should counteract suppression of worrying thoughts only if the thought still is present during work on an ability test. A reduction in the accessibility after their initial activation by directed forgetting, however, should prevent concepts related to failure from exerting any influence on the assumed suppression process.

A further modification pertained to the words used to activate concepts related to failure. In Experiment 1 we had selected words that were synonymous with or semantically closely related to failure. In Experiment 2, however, we used the most common associations to failure that were generated in a pilot study. Besides, we additionally included words representing neutral associations to achievement. These words were not generated in response to the cue *failure*, but in response to the cue *achievement*. Only associations were selected that were neither positively

nor negatively valenced. The neutral associations to achievement were not subjected to directed forgetting. Thus, we were able to disentangle effects of the activation of concepts related to failure from effects of the activation of achievement-related concepts in general. We assume that not the activation of any achievement-related concept has the potential to eliminate worry suppression, but that the activated concepts necessarily have to be related to failure. In addition, we changed the ability domain to visual-spatial and mathematical reasoning in order to test the generalization of the findings over different kinds of cognitive tasks. Since worry-related processes, such as interference or suppression, are basically verbal or, at least, operate on a verbal level, it is possible that tasks relying mainly on visual working memory might be unaffected by a manipulation activating concepts related to failure.

Participants learned two lists of words. List 1 contained associations to failure among other words, list 2 contained neutral associations to achievement among other words. In one condition of the experiment, participants received a forgetting instruction for list 1 after its presentation—but not in the other condition. Then, performance on two tasks (Raven's matrices and modular arithmetic) was assessed. Both tasks have been used in previous investigations of test anxiety (e.g., Meichenbaum, 1972; Ramirez & Beilock, 2011). By changing the kind of test, we intended to examine the generalization of the observed findings. Raven's matrices and modular arithmetic depend to a lesser degree on verbal abilities than solving verbal analogies does. Hence, we examined whether the activation of failure-related concepts also influenced performance in these tasks probably involving less verbal processing. The two tasks differ from each other, however, by concerning different reasoning abilities. Whereas modular arithmetic represents a test of mathematical abilities, it has been suggested that performance at Raven's matrices depends on visual-spatial abilities besides representing general reasoning ability (Colom & Garcia-Lopez, 2002; Colom, Escorial, & Rebollo, 2004; Flynn, 1998). We assumed the activation of failure-related concepts to influence performance on both tasks.

We examined the hypothesis that reducing the activation of failure-related concepts by directed forgetting hampered test performance of participants high in trait worry. Accordingly, we expected an interaction of the condition with trait worry. The performance of participants high in trait worry in the no-forgetting condition should be better than the performance of participants high in trait worry in the forgetting condition.

METHOD

Participants and Design

Eighty-seven psychology students (64 women) at the University of Trier participated in the experiment. Participants were randomly assigned to two conditions: forgetting, no forgetting. In addition, trait test anxiety was assessed by the TAI-G. A median split, performed separately for men and women, by the score on the worry scale of the TAI-G divided the sample into two groups (low worry, high worry). The sample again had a large range in age (18 to 28). Therefore, we again included age as a covariate in all analyses.

Measures

The software PXLab (Irtel, 2007) again served for programming. The experiment was conducted using Dell Optiplex 755 PCs with Eizo FlexScan S1901 monitors and standard German QWERTZ

keyboards. Participants entered their answers via the keyboard and mouse, with the exception of a free recall memory test. Again, the TAI-G assessed trait test anxiety.

Raven's Matrices

We chose 16 matrices from the Standard Progressive Matrices (Raven, 1960), eight from set C and eight from set D. Participants had to select one out of eight options in answering each matrix by pressing a number between one and eight on the keyboard. The time per matrix was limited to 20 seconds for the set C matrices and to 30 seconds for the set D matrices. If participants did not give an answer within 20 or 30 seconds, respectively, the matrix was scored as incorrect.

Modular Arithmetic

Gauss's modular arithmetic involves problem statements such as $30 \equiv 12 \pmod 3$. The statement is true if dividing the difference of the first and the second number (18 : 3) results in a whole number (cf. Beilock et al., 2004). We generated 31 items of this kind (13 true statements). Participants had to judge the truth of each statement by pressing one out of two marked keys on the keyboard. The time per statement was limited to 2 seconds. If participants did not give an answer within 2 seconds, the item was scored as incorrect.

State Measures

Participants rated on six separate items using an eight-point scale their anxiety, worry, and self-standards during the matrices and during the modular arithmetic test.

Procedure

The experiment was conducted in groups of one to three participants who were tested in separate cubicles. First, participants answered the TAI-G. In the subsequent learning phase, participants were asked to memorize words. Each word was presented for 4 seconds. After the presentation of 18 words (list 1), participants either read the instruction that they should forget the previously presented words and only memorize those words that would subsequently be presented (forgetting condition) or that they should continue to try to memorize the subsequently presented words in the same manner as they did with regard to the previously presented words (no forgetting condition). Then, 18 more words were presented (list 2). List 1 contained six words that had been generated as associations to failure in a pilot study (*anger, disappointment, shame, sorrow, misfortune, failure*).[5] List 2 contained six words that had been generated as associations to achievement and that were neither of negative nor positive valence (*effort, occupation, evaluation, university, learning, study*). The remaining words were unrelated to achievement. After the learning phase, participants first worked on the Raven's matrices and then on the modular arithmetic. Both tests were described as assessing individual strengths and weaknesses in cognitive abilities very exactly by relying on accuracy as well as speed of answers. After the modular arithmetic test, participants received a sheet of paper and were instructed to write down all words from the learning phase they could recall. The instructions emphasized also to recall list 1 words in the forgetting condition. Lastly, participants rated their anxiety, worry, and self-standards during the matrices and during

the modular arithmetic test and provided demographic data (major, age, gender). A final question involved speculations on the investigated hypotheses.

RESULTS AND DISCUSSION

Since performances in visual-spatial as well as mathematical tasks are subject to gender differences (Inzlicht & Schmader, 2011), we included gender as a control factor in all analyses. Analyses of the number of recalled words in the memory test proved the manipulation success. Statistically significantly fewer list 1 words were recalled in the forgetting condition than in the no-forgetting condition $t(85) = 2.60$, $p = .011$. In contrast, the recall of list 2 words did not differ significantly between the two conditions, $t < 1$.

A 2 (high trait worry, low trait worry) × 2 (forgetting, no forgetting) × 2 (gender) between subjects ANCOVA with age as the covariate examined the number of correctly solved matrices. Neither the main effect of worry nor the interaction of worry and condition was statistically significant, Fs < 1, nor was any of the other main effects or interactions, Fs < 3.19, ps $> .078$.

A further 2 (high trait worry, low trait worry) × 2 (forgetting, no forgetting) × 2 (gender) between subjects ANCOVA with age as the covariate examined the number of correctly solved modular arithmetic items. Here, the interaction of trait worry and condition was statistically significant, $F(1, 78) = 4.69$, $p = .033$, $\eta_p^2 = .06$. Corresponding to Experiment 1, the performance of participants high in trait worry was better in the no-forgetting condition than in the forgetting condition; whereas the performance of participants low in trait worry was slightly worse in the no-forgetting condition than in the forgetting condition (see Figure 1, middle section). The main effect of gender also was statistically significant, $F(1, 78) = 9.10$, $p = .003$, $\eta_p^2 = .11$. Men answered more items correctly than women. The remaining main effects were not statistically significant, neither were any of the interactions, Fs < 3.32, ps $> .072$.

We examined the ratings on the items asking for self-standards, anxiety, and worry during the test separately for the ratings with regard to the matrices and the ratings for the modular arithmetic. Thus, six dependent variables were examined. Six 2 (high trait worry, low trait worry) × 2 (forgetting, no forgetting) × 2 (gender) ANCOVAs with age as the covariate examined the ratings. The only statistically significant results were main effects of gender with regard to anxiety and worry during the test, Fs > 7.09, $p < .010$, $\eta_p^2 > .08$. Men gave lower ratings. No other main effect or interaction was statistically significant, Fs < 3.19, $p > .078$. With regard to the matrices, the mean self-standards rating was 4.0 ($SD = 2.0$), the mean anxiety rating was 2.2 ($SD = 2.0$), and the mean worry rating was 3.0 ($SD = 2.0$). With regard to modular arithmetic, the mean self-standards rating was 4.1 ($SD = 2.0$), the mean anxiety rating was 2.3 ($SD = 2.0$), and the mean worry rating was 2.8 ($SD = 2.0$).

Taken together, participants high in trait worry again profited from the activation of concepts related to failure. Modular arithmetic performance of participants high in trait worry in the no-forgetting condition was better than in the forgetting condition. Concept activation influenced only modular arithmetic and not the performance at solving Raven's matrices. There was no main effect of trait worry either. Perhaps, visual-spatial reasoning in general is influenced by test anxiety to a lesser degree. In fact, Raven's matrices have not often been used in research on test anxiety and the existing findings are mixed (e.g., Meichenbaum, 1972; Paulman & Kennelly, 1984). It is, of course, also possible that we did select test items that were not suited to reflect

effects of test anxiety—for example, because of their difficulty level. In any case, we have to state that the matrices used in the present study were insensitive to influences of test anxiety.

EXPERIMENT 3

The results of the first two experiments already demonstrated correspondingly that the activation of concepts related to failure was beneficial to test performances of participants high in trait worry. We tested this assumption by using a further manipulation of concept activation in Experiment 3. Instead of including an initial learning phase operationalizing concept activation, we adopted a priming procedure that did not involve instructions to memorize words for a later memory test. Participants solved a word search puzzle that either comprised words related to failure or neutral control words. We expected that the activation of concepts related to failure through the work on that puzzle would equally enhance performance in a subsequent modular arithmetic test. We chose modular arithmetic again because it had proved to reflect influences of test anxiety and concept activation in Experiment 2. A second aim of Experiment 3 was to examine the dependency of the investigated processes of test anxiety and concept activation on different working memory systems. Perhaps, performance on the Raven's matrices had been unaffected in Experiment 2 because it depended on visual-spatial ability, which might imply that solving the matrices relied on visual working memory. It has been suggested, however, that manipulations affecting the extent of state worry only influence tasks relying primarily on verbal working memory (Rapee, 1993). Beilock, Rydell, and Connell (2007) showed that stereotype-threat-related worries only impaired performance of math tasks that were presented in a format requiring verbal-working-memory capacity, whereas tasks presented in a format requiring visual-working-memory resources were unaffected. Modular arithmetic tasks were presented horizontally—that is, with the elements of one task displayed in one row, one next to the other (verbal-working-memory format) or vertically—that is, with the elements of one task over two rows, one below the other (visual-working-memory format). We transferred this manipulation and presented one half of modular arithmetic tasks in horizontal format and the other half of modular arithmetic tasks in vertical format.

We examined the hypothesis that the activation of failure-related concepts benefited test performance of participants high in trait worry primarily for tasks relying on verbal processing. Thus, we expected influences of test anxiety and concept activation only on the performance for tasks in horizontal tasks if the Raven's matrices in Experiment 2, in fact, had been unaffected because their solution relied primarily on visual working memory.

METHOD

Participants and Design

Seventy-five psychology students (55 women) at the University of Trier participated in the experiment. Participants were randomly assigned to two conditions: failure priming, control. In addition, trait test anxiety was assessed by the TAI-G. A median split, performed separately for men and women, by the score on the worry scale of the TAI-G divided the sample in two (low worry, high worry). The sample again had a large range in age (18 to 29). Therefore, we again

included age as a covariate in all analyses. Three participants were excluded from the analyses because they stated at the end that they had assumed a relationship between the priming procedure and the ability test.

Measures

The software PXLab (Irtel, 2007) served for programming. The experiment was conducted using Dell Optiplex 755 PCs with Eizo FlexScan S1901 monitors and standard German QWERTZ keyboards. Participants entered their answers via the keyboard and mouse, with the exception of completing a word search puzzle in pen and paper format. Again, the TAI-G assessed trait test anxiety.

Modular Arithmetic

Participants worked on 60 modular arithmetic tasks. Thirty tasks were presented in a horizontal format with the elements of the tasks displayed in one row with one beside the other, whereas the other half was presented in a vertical format with the elements of the tasks displayed in two rows below each other. The assignment of the tasks to a vertical or horizontal format was counterbalanced. As in Experiment 2, participants had to judge the truth of each statement by pressing one out of two marked keys on the keyboard. The time per statement was limited to 2 seconds. If participants did not give an answer within 2 seconds, the item was scored as incorrect.

State Measures

Participants rated on three separate items with an eight-point scale their anxiety, worry, and self-standards during the modular arithmetic test.

Procedure

The experiment was conducted in groups of one to three participants who were tested in separate cubicles. First, participants answered the TAI-G. Subsequently, they worked on a word search puzzle. This puzzle comprised 24 × 22 letters pasted individually in a rectangular grid printed on a sheet of paper. In the failure priming condition, 20 words were hidden vertically or horizontally within the puzzle. Participants searched the words and connected the constituting letters using a pen. The words were also printed at the bottom of the page in order to ensure a high detection rate. All 20 words were related to failure (*anxiety, debacle, pressure, disappointment, miscarriage, consequences, flop, breakdown, pessimism, collapse, letdown, setback, failure, self-doubt, sorrows, sadness, insecurity, fizzling, future, doubt*).[6] In the neutral condition, 20 words unrelated to failure and of neutral emotional valence (*anise, baking soda, butter, fruit, frosting, hazelnut, yeast, cocoa, almonds, margarine, flour, cloves, candied orange peel, sugar powder, raisins, cream, chocolate, vanilla sugar, cinnamon, lemon)* had to be detected within a different grid instead. The experimenter interrupted the participants after 10 minutes and showed them the remaining words in the puzzle if they had not found all of them yet. Then, the modular arithmetic test began. The instructions described it as assessing individual strengths and weaknesses in cognitive abilities very exactly by relying on accuracy as well as speed of answers. Two testing

orders were established to which participants were randomly assigned: The test started either with the vertical or with the horizontal format. Lastly, participants rated their anxiety, worry, and self-standards during the modular arithmetic test and provided demographic data (major, age, gender). A final question asked about speculations on the investigated hypotheses.

RESULTS AND DISCUSSION

As in the previous experiment, we included gender as a control factor in all analyses. A 2 (high trait worry, low trait worry) × 2 (failure priming, neutral) × 2 (gender) × 2 (testing order: vertical first, horizontal first) × 2 (presentation format: vertical items, horizontal items) mixed-model ANCOVA with repeated measures on the last factor and with age as the covariate examined the number of correctly solved modular arithmetic items. The interaction of trait worry and priming condition was statistically significant, $F(1, 57) = 5.29$, $p = .025$, $\eta_p^2 = .09$. The performance of participants high in trait worry was better in the failure priming condition than in the neutral condition; whereas the performance of participants low in trait worry was worse in the failure priming condition than in the neutral condition (see Figure 1, lower section). The main effect of gender also was statistically significant, $F(1, 57) = 6.35$, $p = .015$, $\eta_p^2 = .10$. Men answered more items correctly than women. There was a marginal three-way interaction of the testing order with condition and gender, $F(1, 57) = 3.98$, $p = .051$. The remaining main effects were not statistically significant, neither were any of the interactions, Fs < 3.33, ps $> .073$.

We examined the ratings of self-standards, anxiety, and worry during the test by 2 (high trait worry, low trait worry) × 2 (failure priming, neutral) × 2 (gender) ANCOVAs with age as the covariate. There were no statistically significant effects with regard to self-standards, Fs < 1.85, $p > .178$. The mean self-standards rating was 4.2 ($SD = 2.0$). With regard to anxiety, the interaction of gender and trait worry was statistically significant, $F(1, 65) = 4.51$, $p = .038$, $\eta_p^2 = .07$, but no other main effect or interaction was, Fs < 2.83, $p > .097$. Women high in trait worry indicated higher anxiety ($M = 3.2$, $SD = 2.1$) than men high in trait worry ($M = 1.1$, $SD = 1.8$) but women low in trait worry ($M = 1.9$, $SD = 2.1$) indicated slightly lower anxiety than men low in trait worry ($M = 2.3$, $SD = 2.6$). With regard to worry during the test, the main effect of gender was statistically significant, $F(1, 65) = 8.74$, $p = .004$, $\eta_p^2 = .12$. Men gave lower ratings. More importantly, the interaction of priming condition and trait worry also was statistically significant, $F(1, 65) = 7.84$, $p = .007$, $\eta_p^2 = .11$. Whereas the ratings of worry during the test only slightly differed between participants high and low in trait worry in the neutral condition, participants high in trait worry rated worry during the test higher than participants low in trait worry in the failure activation condition (see Table 1). As in Experiment 1, these higher ratings indicate that participants high in worry, in fact, did not suppress worrying thoughts in the failure activation condition. In contrast, the low ratings of participants low in trait worry again suggest that they were influenced by failure activation in an opposing manner, perhaps, inducing suppression of worry that normally would not have occurred.

In summary, Experiment 3 replicates the finding that the activation of concepts related to failure was beneficial to participants high in trait worry, although the effect was less pronounced as in the preceding experiments. In contrast, the performance of participants low in trait worry seemed

to be impaired by failure priming. This detrimental effect of failure priming might originate in a generally stronger concept activation caused by the word search puzzle that was able to induce worries or suppress worries in participants that would not have worried otherwise. Support for this interpretation comes from the fact that three participants had to be excluded because they identified the priming purpose of the experiment. The lack of statistically significant interactions of the presentation format (horizontal, vertical) with any of the other independent variables indicates that the investigated processes do not pertain to a specific subsystem of working memory but that verbal and visual processing is affected similarly. Indeed, it remains unclear why the Raven's matrices had not been affected in Experiment 2, thus, but the present results demonstrate that processing that taxes visual working memory resources is not protected against influences of test anxiety or failure priming.

GENERAL DISCUSSION

Three experiments provide converging evidence. The activation of concepts related to failure enhances test performances of highly test-anxious persons. Experiment 1 compared the initial activation of concepts related to failure with a control condition devoid of an initial activation of these concepts; whereas Experiment 2 compared the reduction of an initial activation of concepts related to failure with a control condition that did not reduce the initial activation of these concepts. Experiment 3 activated concepts related to failure through a word search puzzle. In Experiment 1, memorizing words that were synonymous or semantically closely related to failure yielded better performance of participants high in trait worry, as compared to memorizing words unrelated to achievement. In Experiment 2, memorizing associations to failure without a subsequent instruction to forget these words yielded better performance of participants high in trait worry, as compared to memorizing associations to failure followed by the instruction to forget the words. This directed forgetting ought not to be confused with the suppression of worrying thoughts. The forgetting instruction concerned initially learned words that partly were associated with failure, before a measure of cognitive ability was even announced. Thereby, the accessibility of the corresponding concepts was diminished. Only in succession of the forgetting instruction then the suppression of worrying thoughts during the cognitive ability tests was able to occur because no activation of concepts related to failure prevented the suppression. In Experiment 3, failure priming interacted with trait worry, enhancing the performance of participants high in trait worry but impairing the performance of participants low in trait worry.

Retrospectively rated items assessed self-standards, anxiety, and worry during the tests. In Experiment 1, the initial learning of words related to failure was associated with a higher amount of state worry reported by highly test-anxious participants compared to slightly test-anxious participants. This result matches the assumption that the activation of concepts related to failure eliminates the suppression of worrying thoughts. The finding was replicated in Experiment 3. In Experiment 2, we did not find a parallel result in the amount of reported state worry, however. Perhaps, the used items were not sensitive enough to assess the actual amount of worry because, different from experiments 1 and 3, the items were not rated immediately after the cognitive ability test but only after a memory test for the initially learned items. Also, asking for separate ratings concerning the matrices test and the arithmetic test might have, contrary to

the intended higher specificity, actually blurred the ratings given by the participants. The lack of any statistically significant effect on the ratings except for a gender difference suggests a general insensitivity of the items to test-anxiety-related processes in Experiment 2. The lower ratings of state anxiety compared to worry in all three experiments indicate that participants did not experience strong anxiety. Given the artificial testing situation in a laboratory experiment, the low anxiety ratings are not surprising. However, a floor effect might preclude effects on anxiety by the examined independent variables that would show up under more-realistic conditions—for example, in an exam. Although the worry ratings were somewhat higher, they did not exceed the mean of the used 8-point-scale either. We assume worry ratings also to be higher in a real-life exam. Hence, it should be examined in future studies whether the present findings generalize to real-life exams involving higher levels of anxiety and worry and perhaps also motivation.

The activation of concepts related to failure was supposed to eliminate the suppression of worrying thoughts. We intended to vitiate any suppression by bringing those concepts to awareness that constitute the core of worry. The idea that bringing failure to awareness can help highly test anxious persons certainly has a counterintuitive appeal. It does not correspond to actions that rely on the activation of positive concepts related to achievement such as competence priming (Lang & Lang, 2010) or relaxation procedures (e.g., Russell & Sipich, 1973). It matches, however, the finding that expressive writing can alleviate test anxiety (Ramirez & Beilock, 2011). The crucial aspect of expressive writing might not be a reevaluation of the situation, though, as Ramirez and Beilock argue, but becoming fully aware of the threatening outcome of failure. This awareness then in turn counteracts a working-memory-capacity-consuming process of suppressing worrying thoughts. The causal role of suppression of worrisome thoughts also might explain why, in the course of many decades of research on test anxiety, it has not been shown convincingly yet that distraction by worry consumes working memory capacity (cf. Eysenck et al., 2007).

A limitation of the present study is that the suppression of worry was not directly measured. Therefore, alternative processes potentially accounting for the present results should be considered in addition. Those processes might comprise a greater level of reactive effort invested by test-anxious participants. The activation of failure concepts might have triggered more effort in order to avoid failure and to compensate for working-memory deficits (Eysenck & Calvo, 1992; Pekrun, 2006; Putwain, Shah, & Lewis, 2014). However, in the experimental conditions with an activation of failure concepts, highly test anxious participants only rated state worry higher compared to slightly test anxious participants, but ratings on the motivational item of self-standards did not differ. Yet, it is possible that the item assessing state worry was more sensitive and an alternative measure of motivation would be able to display a difference in the amount of invested effort. A scale assessing motivation with more than one item could possess a higher reliability. A further limitation is the relatively low number of men in the examined samples. Thus, the present findings were largely based on female participants.

Implications of the findings reported here should pertain to test-anxiety treatments. Interventions relying on the activation of the feared threat of failure may prove to be effective tools for alleviating test anxiety also in real-world settings. If this activation is the crucial factor in overcoming performance decrements, then other elements of instruments intended to enhance performances may be replaceable. It might suffice to face the threat.

AUTHOR NOTES

Dr. Tobias Tempel is research associate in the department of psychology at the University of Trier. His research interests include test anxiety, stereotypes, and memory. **Dr. Roland Neumann** is professor of General Psychology at the University of Trier. His research interests include stereotypes, emotional behavior, and test anxiety.

NOTES

1. The originally German items to assess self-standards, anxiety, and worry were *Wie gut wollten Sie bei der Ergänzung der Wortstämme und der Lösung der Analogien sein?* (0 = *durchschnittlich*, 7 = *ausgezeichnet*); *Wie viel Angst hatten Sie vor den Wortstämmen und Analogien?* (0 = *überhaupt keine*, 7 = *extrem große*); and *Wie stark haben Sie Sorgen über Ihre Leistung beschäftigt?* (0 = *überhaupt nicht*, 7 = *extrem stark*)
2. The originally German failure-related words in Experiment 1 were *Misserfolg*, *Versagen*, *Enttäuschung*, *Scheitern*, *Pleite*, *Debakel*. The neutral words were: *Kunststoff*, *Pappe*, *Stein*, *Zement*, *Metall*, *Plastik.*
3. It might be of interest whether gender influenced the results in Experiment 1. Unfortunately, the low number of men (16) violates prerequisites of ANOVA tests and can distort results if gender is entered as a factor. We nevertheless included gender as an additional factor in an exploratory analysis. Gender did not moderate any effect, $F(1, 67) < 1.47$, $p > .231$.
4. As an alternative statistical approach, multiple regression with interaction terms was used (Jaccard & Turrisi, 2003). The same variables as in the above ANCOVAs entered as predictors. Trait worry was z standardized and entered as a continuous predictor. In Experiment 1, the predictor concept activation was coded with -1 and 1, as was the word-stem completion condition. Age entered as a covariate. The interaction of concept activation and trait worry was also in this analysis statistically significant, $\beta = .26$, $p = .022$. In Experiment 2, the predictor condition was coded with -1 and 1, as was gender. Age entered as a covariate. With regard to modular arithmetic performance, the interaction of condition and trait worry was also in this analysis statistically significant, $\beta = .27$, $p = .019$. With regard to the number of correctly solved matrices, it was not, $\beta = .13$, $p = .328$. In Experiment 3, the predictor priming condition was coded with -1 and 1, as were gender and testing order. Age entered as a covariate. The interaction of condition and trait worry was also in this analysis statistically significant, $\beta = .26$, $p = .033$.
5. The originally German failure-related words in Experiment 2 were *Ärger*, *Enttäuschung*, *Scham*, *Trauer*, *Unglück*, *Versagen*. The neutral achievement-related words were *Anstrengung*, *Beruf*, *Bewertung*, *Hochschule*, *Lernen*, *Studieren*
6. The originally German failure-related words in Experiment 3 were *Angst, Debakel, Druck, Enttäuschung, Fehlschlag, Konsequenzen, Misserfolg, Misslingen, Pessimismus, Pleite, Reinfall, Rückschlag, Scheitern, Selbstzweifel, Sorgen, Traurigkeit, Unsicherheit, Versagen, Zukunft, Zweifel*. The neutral words were *Anis, Backpulver, Butter, Frucht, Guss, Haselnuss, Hefe, Kakao, Mandeln, Margarine, Mehl, Nelken, Orangeat, Puderzucker, Rosinen, Sahne, Schokolade, Vanillezucker, Zimt, Zitrone*.

REFERENCES

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (1999). *Intelligenz-struktur-test 2000: IST 2000*. Göttingen, Germany: Hogrefe.

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*, 584–600. doi: 10.1037/0096-3445.133.4.584

Beilock, S. L., Rydell, R. J., & Connell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*, 256–276. doi:10.1037/0096-3445.136.2.256

Benson, J., Moulin-Julian, M., Schwarzer, C., Seipp, B., & El-Zahhar, N. (1992). Cross-validation of a revised test anxiety scale using multi-national samples. In K. A. Hagtvet & T. B. Johnsen (Eds.), *Advances in test anxiety research* (Vol. 7, pp. 62–83). Amsterdam, The Netherlands: Swets & Zeitlinger.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (S. 309–330). Hillsdale, NJ: Erlbaum.

Brodish, A. B., & Devine, P. G. (2009). The role of performance-avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology*, *45*, 180–185. doi:10.1016/j.jesp.2008.08.005

Brown, L. A., Forman, E. M., Herbert, J. D., Hoffman, K. L., Yuen, E. K., & Goetter, E. M. (2011). A randomized controlled trial of acceptance-based behavior therapy and cognitive therapy for test anxiety: A pilot study. *Behavior Modification*, *35*, 31–53. doi:10.1177/0145445510390930

Carr, P. B., & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology*, *45*, 853–859. doi:10.1016/j.jesp.2009.03.003

Cohen, M. G., Wupperman, P., & Tau, G. (2013). Mindfulness in the treatment of adolescents with problem substance use. *Adolescent Psychiatry*, *3*, 172–183.

Colom, R., Escorial, S., & Rebollo, I. (2004). Sex differences on the progressive matrices are influenced by sex differences on spatial ability. *Personality and Individual Differences*, *37*, 1289–1293 doi:10.1016/j.paid.2003.12.014

Colom, R., & Garcia-Lopez, O. (2002). Sex differences in fluid intelligence among high school graduates. *Personality and Individual Differences*, *32*, 445–451. doi:10.1016/S0191-8869(01)00040-X

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

Davey, H. M., Barratt, A. L., Butow, P. N., & Deeks, J. J. (2007). A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. *Journal of Clinical Epidemiology*, *60*, 356–360. doi:10.1016/j.jclinepi.2006.07.015

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

Derakshan, N., & Eysenck, M. W. (1998). Working memory capacity in high trait-anxious and repressor groups. *Cognition & Emotion*, *12*, 697–713. doi:10.1080/026999398379501

Dutke, S., & Stöber, J. (2001). Test anxiety, working memory, and cognitive performance: Supportive effects of sequential demands. *Cognition & Emotion*, *15*, 381–389. doi:10.1080/02699930125922

Eysenck, M. W. (1979). Anxiety, learning, and memory: A reconceptualization. *Journal of Research in Personality*, *13*, 363–385. doi:10.1016/0092-6566(79)90001-1

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion*, *6*, 409–434. doi:10.1080/02699939208409696

Eysenck, M. W., Derakhshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*, 336–353. doi:10.1037/1528-3542.7.2.336

Flynn, J. R. (1998). Israeli military IQ tests: Gender differences small—IQ gains large. *Journal of Biosocial Science*, *30*, 541–553. doi:10.1017/S0021932098005410

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York, NY: Guilford.

Hittner, J. B., & Hemmo, R. (2009). Psychosocial predictors of dental anxiety. *Journal of Health Psychology*, *14*, 53–59. doi:10.1177/1359105308097945

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58*, 47–77. doi:10.3102/00346543058001047

Hepburn, S. R., Crane, C., Barnhofer, T., Duggan, D. S., Fennell, M. J. V., & Williams, J. M. G. (2009). Mindfulness-based cognitive therapy may reduce thought suppression in previously suicidal participants: Findings from a preliminary study. *British Journal of Clinical Psychology*, *48*, 209–215 doi:10.1348/014466509X414970

Hodapp, V. (1991). Das Prüfungsängstlichkeitsinventar TAI-G: Eine erweiterte und modifizierte Version mit vier Komponenten [The Test Anxiety Inventory TAI-G: An expanded and modified version with four components]. *Zeitschrift für Pädagogische Psychologie*, ;*5*, 121–130.

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

Ikeda, M., Iwanaga, M., & Seiwa, H. (1996). Test anxiety and working memory system. *Perceptual & Motor Skills*, *82*, 1223–1231. doi:10.2466/pms.1996.82.3c.1223

Inzlicht, M., & Schmader, T. (2011). *Stereotype threat: Theory, process, and application*. New York, NY: Oxford University Press.

Irtel, H. (2007). *PXLab: The psychological experiments laboratory [online]. Version 2.1.11*. Mannheim, Germany: University of Mannheim. Retrieved from http://www.pxlab.de

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression (*2nd *ed.)*. Thousands Oaks, CA: Sage.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, *16*, 175–179. doi:10.1111/j.0956-7976.2005.00799.x

Lang, J. W. B., & Lang, J. (2010). Priming competence diminishes the link between cognitive test anxiety and test performance. *Psychological Science*, *21*, 811–819. doi:10.1177/0956797610369492

Lee, J. H. (1999). Test anxiety and working memory. *Journal of Experimental Education*, *67*, 218–240. doi:10.1080/00220979909598354

Levitt, J. T., Brown, T. A., Orsillo, S. M., & Barlow, D. H. (2004). The effects of acceptance versus suppression of emotion on subjective and psychophysiological response to carbon dioxide challenge in patients with panic disorder. *Behavior Therapy*, *35*, 747–766. doi:10.1016/S0005-7894(04)80018-2

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, *20*, 975–978. doi:10.2466/pr0.1967.20.3.975

MacLeod, C., & Donnellan, A. M. (1993). Individual differences in anxiety and the restriction of working memory capacity. *Personality and Individual Differences*, *15*, 163–173. doi:10.1016/0191-8869(93)90023-V

Marino-Carper, T., Negy, C., Burns, G., & Lunt, R. A. (2010). The effects of psychoeducation on thought-action fusion, thought suppression, and responsibility. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*, 289–296. doi: 10.1016/j.jbtep.2010.02.007

McGlone, M. S., & Aronson, J. (2007). Forewarning and forearming stereotype-threatened students. *Communication Education*, *56*, 119–133. doi:10.1080/03634520601158681

Meichenbaum, D. H. (1972). Cognitive modification of test-anxious college students. *Journal of Consulting and Clinical Psychology*, *39*, 370–380. doi:10.1037/h0033896

Morris, L. W., Davis, M. A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology*, *73*, 541–555. doi:10.1037/0022-0663.73.4.541

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, *126*, 247–259. doi:10.1037/0033-2909.126.2.247

Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2012). Anxiety and depression in academic performance: An exploration of the mediating factors of worry and working memory. *School Psychology International, 33*, 443–449. doi:10.1177/0143034311427433

Paulman, R. G., & Kennelly, K. J. (1984). Test anxiety and ineffective test taking: Different names, same construct? *Journal of Educational Psychology*, *76*, 279–288. doi:10.1037/0022-0663.76.2.279

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*, 315–341. doi:10.1007/s10648-006-9029-9

Purdon, C. (1999). Thought suppression and psychopathology. *Behaviour Research and Therapy*, *37*, 1029–1054. doi:10.1016/S0005-7967(98)00200-9

Putwain, D. W., Shah, J., & Lewis, R. (2014). Performance-evaluation threat does not adversely affect verbal working memory in high test-anxious persons. *Journal of Cognitive Education and Psychology*, *13*, 120–136. doi:10.1891/1945-8959.13.1.120

Ramirez, G., & Beilock, S. L. (2011). Writing about testing worries boosts exam performance in the classroom. *Science*, *331*, 211–213. doi:10.1126/science.1199427

Rapee, R. M. (1993). The utilisation of working memory by worry. *Behaviour Research and Therapy*, *31*, 617–620. doi:10.1016/0005-7967(93)90114-A

Raven, J. C. (1960). *Guide to using the standard progressive matrices*. London, UK: H. K. Lewis.

Richards, A., French, C. C., Keogh, E., & Carter, C. (2000). Test anxiety, inferential reasoning and working memory load. *Anxiety, Stress & Coping*, *13*, 87–109 doi:10.1080/10615800008248335

Roediger, H. L., III., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Russell, R. K., & Sipich, J. F. (1973). Cue-controlled relaxation in the treatment of test anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *4*, 47–49. doi:10.1016/0005-7916(73)90038-4

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1064–1072. doi:10.1037/0278-7393.28.6.1064

Sarason, I. G. (1978). The Test Anxiety Scale: Concept and research. In C. D. Spielberger & I. G. Sarason (Eds.), *Stress and anxiety* (Vol. 5, pp. 193–216). Washington, DC: Hemisphere.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology*, *46*, 929–938. doi:10.1037/0022-3514.46.4.929

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440–452. Retrieved from http://dx.doi.org/10.1037/0022-3514.85.3.440

Schwarzer, R. (1984). Worry and emotionality as separate components in test anxiety. *International Review of Applied Psychology*, *33*, 205–220 doi:10.1111/j.1464-0597.1984.tb01429.x

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

Spielberger, C. D. (1980). *Test Anxiety Inventory*: Palo Alto, CA: Consulting Psychologists Press.

Stöber, J. (2004). Dimensions of test anxiety: Relations to ways of coping with pre-exam anxiety and uncertainty. *Anxiety, Stress & Coping*, *17*, 213–226. doi:10.1080/10615800412331292615

Tempel, T., & Neumann, R. (2014). Stereotype threat, test anxiety, and mathematics performance. *Social Psychology of Education*, *17*, 491–501. doi:10.1007/s11218-014-9263-9

Wacker, A., Jaunzeme, J., Jaksztat, S. (2008). Eine Kurzform des Prüfungsängstlichkeitsinventars TAI-G [A short version of the test anxiety inventory TAI-G]. *Zeitschrift für Pädagogische Psychologie*, *22*, 73–81. doi:10.1024/1010-0652.22.1.73

Ware, W. B., Galassi, J. P., & Dew, K. M. H. (1990). The test anxiety inventory: A confirmatory factor analysis. *Anxiety Research*, *3*, 205–212. doi:10.1080/08917779008248753

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*, 34–52. doi:10.1037/0033-295X.101.1.34

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, *51*, 59–91 doi:10.1146/annurev.psych.51.1.59

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

Zettle, R. D. (2003). Acceptance and commitment therapy (ACT) vs. systematic desensitization in treatment of mathematics anxiety. *The Psychological Record*, *53*, 197–215.