

# Stereotype threat, test anxiety, and mathematics performance

Tobias Tempel · Roland Neumann

Received: 14 March 2013 / Accepted: 10 May 2014 / Published online: 31 May 2014  
© Springer Science+Business Media Dordrecht 2014

**Abstract** We investigated the combined effects of stereotype threat and trait test anxiety on mathematics test performance. Stereotype threat and test anxiety interacted with each other in affecting performance. Trait test anxiety predicted performance only in a diagnostic condition that prevented stereotype threat by stereotype denial. A state measure of fear of the test mediated this influence. However, stereotype threat reduced the performance of low test-anxious participants to the level of high test-anxious participants. Thus, stereotype threat affected persons low in test anxiety but not persons high in test anxiety. Both phenomena apparently share common mechanisms through which they impair performance.

**Keywords** Trait anxiety · State anxiety · Worry · Gender · Stereotype threat

## 1 Introduction

Poor performances in ability tests can stem from several sources. Aside from low ability and lack of effort, characteristics of the test-taking situation may impair performance. Two detrimental phenomena that are both related to such characteristics are stereotype threat and test anxiety.

Stereotypes not only shape the perception and treatment of others. If they refer to oneself they can shape self-perception and one's own behavior in a stereotype-consistent manner. Testing situations that activate stereotypes related to perceived deficient abilities can undermine performance. In a seminal study, [Steele and Aronson \(1995\)](#) demonstrated that Black participants who indicated their race prior to com-

---

T. Tempel (✉) · R. Neumann  
Fachbereich I—Psychologie, University of Trier, 54286 Trier, Germany  
e-mail: tempel@uni-trier.de

pleting a difficult verbal ability test performed significantly worse compared to both White and Black participants who were not requested to indicate their race. This finding has been replicated for several domains such as female inferiority in mathematics (e.g. Ambady et al. 2004; Spencer et al. 1999) or spatial ability (e.g. Martens et al. 2006; McGlone and Aronson 2006). A few studies have already identified dispositional factors that moderate stereotype threat effects. These are domain identification (Aronson et al. 1999; Stone et al. 1999), stigma consciousness (Brown and Pinel 2003), stereotype endorsement (Schmader et al. 2004), and in-group identification (Schmader 2002). We, for the first time, investigated a dispositional factor that is not linked to the stereotype but is related directly to the diagnosticity of a testing situation, namely trait test anxiety.

Corresponding to the conception of state and trait anxiety, test anxiety is a personality trait that possesses its own accordant state. The more test-anxious a person is (trait), the stronger he or she experiences stress when facing a testing situation (state). A thorough investigation of the different facets of test anxiety revealed two key features: emotionality, which comprises anxious feelings and physiological arousal, and the cognitive component of worry. Emotionality and worry represent two well-validated factors of inventories to assess test anxiety (Liebert and Morris 1967; Schwarzer 1984; Spielberger 1980; Wacker et al. 2008; Ware et al. 1990). Moreover, they correlate differentially with test performance. Worry predicts performance better than emotionality (Hembree 1988; Seipp 1991). To demonstrate the causation of performance impairments, typically state test anxiety is induced by presenting a task, for example solving anagrams, as a test of ability (diagnostic condition). Performance then is compared to a condition in which that same task is presented in a non-diagnostic manner, i.e. without mentioning the possibility of drawing inferences about the individual's ability. Sometimes, the non-diagnostic condition also tries to mask any potential ability assessment by instructions explicitly emphasizing that the study would not allow any inferences about the participants' abilities. In this experimental design, worry reliably predicts the amount of reduced performance in the diagnostic condition as compared to the non-diagnostic condition. Emotionality, in contrast, is associated to the actual impaired performance to a lesser degree.

## 2 The present research

Stereotype threat and test anxiety both manifest their detrimental impact in testing situations, i.e. under diagnostic conditions that imply the testing of individual abilities. Both phenomena therefore might affect performance in one and the same situation. We examined their combined effects in such situations.

Regarding the interplay of test anxiety and stereotype threat, several theoretical models are possible. Both might influence test performance independently from each other. Their effects then should be additive. However, both also might interact because test anxiety affects the susceptibility to stereotype threat. Test anxiety then would act as a moderator of the strength of stereotype threat effects, with high test-anxious persons either being prone to stronger performance detriments due to the activation of a relevant

stereotype or being relatively protected against stereotype threat. Stereotype threat and test anxiety also might involve the same mechanisms through which performance is impaired, that is, they might share common mediator variables. Eliciting influences of either one of them in a given test situation then should decrease performance in the same way as the simultaneous elicitation of both, test anxiety and stereotype threat. Stereotype threat without test anxiety should yield the same low level of performance as test anxiety without stereotype threat, as well as their combination.

Apparently, cognitive processes that mediate the effects of test anxiety and stereotype threat are quite similar to each other or even might be identical. Eysenck (1979) proposed that worrying thoughts of high test-anxious persons absorb working memory capacity needed for test performance. As a consequence, impaired attention impacts especially those cognitive tasks that require complex operations. Several studies support this assumption (Darke 1988; Derakshan and Eysenck 1998; Dutke and Stöber 2001; Ikeda et al. 1996; MacLeod and Donnellan 1993; Richards et al. 2000). Stereotype threat also diminishes working memory capacity (Schmader and Johns 2003). In addition, stereotype threat increases the amount of worrying thoughts (Cadinu et al. 2005). An experiment by Brodisha and Devine (2009) shows probably the closest parallels of the causation of both kinds of performance impairment. Brodisha and Devine not only demonstrated an increase of worry due to stereotype threat by using self-report items of a questionnaire constructed to measure test anxiety, but they also found a significant mediation of the impaired performance of participants under stereotype threat by the amount of worry.

If the same mediators cause performance decrements of stereotype threat and test anxiety, then triggering these mediators one way or the other should produce the same result. Under diagnostic conditions without stereotype threat these mediators impair the performance of high test-anxious persons, but when stereotype threat accompanies diagnosticity these mediators should be triggered both for persons high and low in test anxiety. Only under diagnostic conditions, without stereotype activation, should performance depend on test anxiety, i.e. low test-anxious persons exhibit better performance than high test-anxious persons.

In the present experiment, participants learned that their mathematical ability was going to be tested. One of the two experimental conditions additionally emphasized that contrary to the stereotype about women's inferiority, extensive recent investigations had shown that in reality no gender differences exist in mathematical ability. This stereotype denial was supposed to prevent stereotype threat, yet leaving test anxiety as an influence on math performance. Immediately before beginning to work on the problems, participants were asked how afraid they were of the test in order to assess state anxiety. In the following, this item will be designated as *fear of the test* because it may reflect test anxiety as well as stereotype threat-induced anxiety, thereby being unambiguously object-related. We expected an interaction of test anxiety with the experimental condition. Math performance should depend on test anxiety at stereotype denial, but it should be equally low for persons high and low in test anxiety under stereotype threat.

### 3 Method

#### 3.1 Participants and design

First year psychology students at the University of Trier completed the German Test Anxiety Inventory (TAI-G; [Hodapp 1991](#)) during two lectures taking place several weeks before the experiment. These students constituted the pool from which the participants of the experiment were drawn. Sixty-three women were randomly assigned to one of two conditions, in both of which several arithmetic problems were presented as a test of mathematical ability. In one condition, the instructions explicitly stated that the stereotype about women's inferior mathematical ability does not correspond to reality (stereotype denial,  $n = 32$ ). In the other condition instructions did not mention the stereotype in order to subtly activate it (stereotype threat,  $n = 31$ ). Subtle stereotype activation has been shown to induce stronger stereotype threat effects than blatant stereotype activation involving the explicit statement of the stereotype ([Nguyen and Ryan 2008](#)).

#### 3.2 Measures

##### 3.2.1 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 as do the worry scale or the emotionality scale (e.g. [Stöber 2004](#)). However, direct evidence of associations of the two ancillary scales to performance outcome measures is as yet missing.

##### 3.2.2 Math test

Eight arithmetical problems from the Program for International Student Assessment (PISA) were printed on several sheets of paper. For example, one of the eight items comprised the translation of a given amount of Singapore dollar into South African rand at a specified exchange rate. Their order was the same for all participants who however were free to skip working on a problem and to proceed to the next whenever they felt they were not likely to arrive at a solution. Participants wrote down their answers on a separate sheet of paper. Scoring was done, (depending on the difficulty level of each task) by awarding between two and five points per problem. The maximum

of achievable points was 30. However, pretests ensured that participants would not be able to solve all problems in the allotted time.

### 3.3 Procedure

The experiment was conducted in groups of one to four people. Participants first learned about the testing of their mathematical ability. In the stereotype threat condition, instructions stated that the math test consisted of PISA problems and that it has as an objective the measuring of individual strengths and weaknesses regarding mathematical ability. The instructions of the stereotype denial condition were identical but additionally stated that contrary to the well-known stereotype about women's weaker mathematical abilities, the PISA studies had shown that in reality no gender differences in mathematical ability exist. The time to work on the test was limited to 15 min.

Participants then indicated their gender. To assess fear of the test, participants were asked: "How afraid are you of the test?" They responded on an 8-point scale ranging from 0 (*not at all*) to 7 (*extremely*). Then they started to work on the math test. After 15 minutes the experimenter gave a stop signal. Participants indicated if they had any suspicion about the investigated hypotheses. No participant stated a hypothesis-consistent suspicion. Finally, the experimenter thanked the participants for their taking part in the experiment and debriefed them about the purpose of the study.

## 4 Results

Math test performance was analyzed using multiple regression. In a first analysis, trait worry, experimental condition, and their interaction term were entered as predictors<sup>1</sup> (see Table 1). The interaction was significant. Figure 1 shows its structure. It indicates that the lower the participants' level of worry, the larger was the performance difference between the two conditions. For participants low in worry, math performance was considerably lower under stereotype threat than in the stereotype denial condition, but not for participants high in worry.

In a second analysis, fear of the test, experimental condition, and their interaction term were entered as predictors (see Table 2). In this analysis, the interaction was significant too. Figure 2 shows its structure. Parallel to the influence of trait worry, a larger performance difference between the two conditions was associated with lower fear of the test.

A third analysis used trait worry, experimental condition, and their interaction term to predict fear of the test.<sup>2</sup> Table 3 shows the results. The interaction was significant.

<sup>1</sup> In contrast to the results for trait worry, using trait emotionality and the experimental condition as predictors of math performance did not produce a significant interaction ( $p = .373$ ), nor did emotionality by itself significantly predict math performance ( $p = .588$ ).

<sup>2</sup> Trait emotionality did not significantly predict fear of the test ( $p = .540$ ), neither did the interaction term with the experimental condition ( $p = .284$ ).

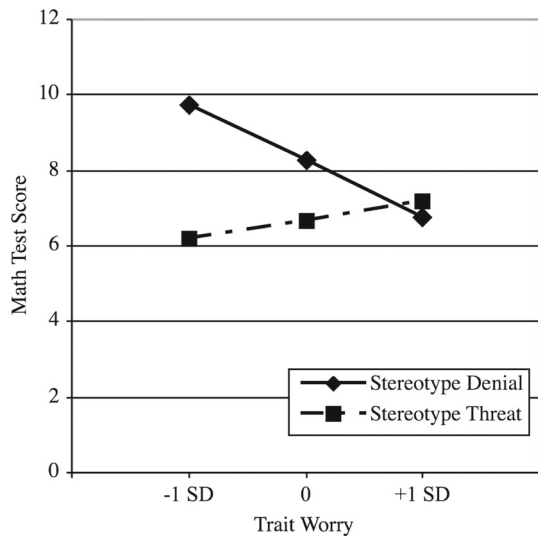
**Table 1** Math performance predicted by experimental condition and trait worry

	<i>B</i>	95 % CI
Condition <sup>a</sup>	−1.57 <sup>+</sup>	[−3.21, 0.07]
Trait worry	−1.48*	[−2.75, −0.22]
Condition × trait worry	1.98*	[0.31, 3.66]

*N* = 63. *CI* confidence interval

<sup>+</sup>  $p < .10$ ; \*  $p < .05$

<sup>a</sup> Coded as 0 = stereotype denial, 1 = stereotype threat

**Fig. 1** Trait worry as a moderator of the effect of experimental condition on math performance**Table 2** Math performance predicted by experimental condition and fear of the test

	<i>B</i>	95 % CI
Condition <sup>a</sup>	−1.28 <sup>+</sup>	[−2.76, 0.20]
Fear of the test	−2.56***	[−3.70, −1.42]
Condition × fear of the test	1.88*	[0.38, 3.38]

*N* = 63. *CI* confidence interval

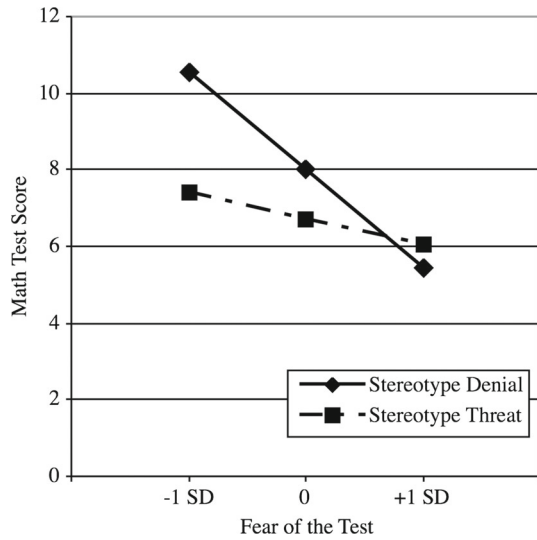
<sup>+</sup>  $p < .10$ ; \*  $p < .05$ ; \*\*\*  $p < .001$

<sup>a</sup> Coded as 0 = stereotype denial, 1 = stereotype threat

In the stereotype denial condition trait worry and fear of the test correlated positively, but in the stereotype threat condition they did not (see Fig. 3).

The above regression analyses correspond to the first steps of a mediation analysis (Baron and Kenny 1986). The interaction of the experimental condition with trait worry predicted the math performance, as did the interaction of the experimental condition with fear of the test, the potential mediator. The interaction of the experimental condition with trait worry also predicted fear of the test. To complete the mediation

**Fig. 2** Fear of the test as a moderator of the effect of experimental condition on math performance



**Table 3** Fear of the test predicted by experimental condition and trait worry

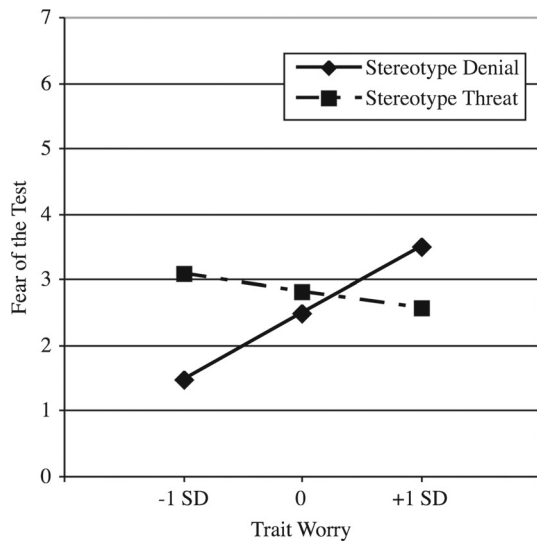
	<i>B</i>	95 % CI
Condition <sup>a</sup>	0.34	[-0.63, 1.31]
Trait worry	1.02**	[0.27, 1.77]
Condition × trait worry	-1.27*	[-2.26, -0.28]

*N* = 63. *CI* confidence interval

\*  $p < .05$ ; \*\*  $p < .01$

<sup>a</sup> Coded as 0 = stereotype denial, 1 = stereotype threat

**Fig. 3** Trait worry as a moderator of the effect of experimental condition on fear of the test



analysis, we conducted separate regression analyses for each condition. First, only trait worry was used to predict math performance. The prediction was significant in the stereotype denial condition,  $\beta = -.35$ ,  $p = .048$ , but not in the stereotype threat condition,  $\beta = .20$ ,  $p = .280$ . Secondly, fear of the test additionally entered the regression in the stereotype denial condition,  $\beta = -.58$ ,  $p = .002$ . As a consequence, trait worry no longer predicted math performance significantly,  $\beta = -0.06$ ,  $p = .739$ . Lastly, a Sobel Test (Preacher and Hayes 2004) confirmed that the mediation was significant,  $z = 2.29$ ,  $p = .022$ .

## 5 Discussion

We examined the interplay of stereotype threat and test anxiety by contrasting two conditions: one subtly inducing stereotype threat, the other counteracting stereotype threat by stereotype denial. Test anxiety predicted performance only in the stereotype denial condition, but under stereotype threat performance was generally low, irrespective of test anxiety. The significant interaction shows that stereotype threat and test anxiety do not influence test performance independently. The negative link between test anxiety and performance only appeared when tasks to be solved were presented as a test of a cognitive ability, and stereotype threat was prevented at the same time. Whereas this result could reflect lower susceptibility to stereotype threat of high test-anxious persons, the mediation by fear of the test suggests common mechanisms in the causation of stereotype threat and test anxiety effects.

It is most probable that the fear of the test item represented state worry. Stereotype threat negatively affected only the performance of those participants who were not affected by test anxiety in the first place. Fear of the test mediated the influence of trait worry on math performance in the stereotype denial condition. Trait worry, however, did not predict performance in the stereotype threat condition. Neither did fear of the test. These results are in accordance with the work of Cadinu et al. (2005) and Brodish and Devine (2009), and suggest that worry caused the performance impairments that occurred. This worry was triggered by the alleged diagnosticity of the math test for high test-anxious participants in both experimental conditions as well as by stereotype threat for low test-anxious participants. However, low test-anxious participants in the stereotype denial condition presumably did not experience worry and hence were able to attain a higher performance. If a person already worries about his or her performance in a cognitive ability test, an additional trigger of worrying thoughts would be less detrimental compared to a person who otherwise would not worry at all. When a high test-anxious person is confronted with a stereotype that implies his or her low performance in an upcoming test, the resulting worry concurs with his or her worry concerning ability tests in general. When a low test-anxious person is confronted with a stereotype that implies his or her low performance in an upcoming test, the triggered worry comes out of the blue.

The present investigation is not without limitations. Since we selected only one specific stereotype, namely the stereotype about women's inferior mathematical abilities, we cannot be sure that the present finding would generalize to a different stereotype. Stereotype threat has been proven to be a phenomenon that can be triggered by a vari-



ety of gender stereotypes as well as racial stereotypes. Future studies should examine whether test anxiety also moderates the impact of stereotype threat that is linked to other stereotypes, for example, about women's inferior spatial reasoning ability. A further, yet related, limitation pertains to the test performance measure we used in the present investigation. Stereotype threat does not only affect cognitive performance measures. In fact, quite different kinds of stereotype threat have been reported and, probably, not all stereotype threat effects involve the same mediator variables. [Schmader et al. \(2008\)](#) distinguish different classes of mediators that, beside anxious feelings and thoughts, also contain self-monitoring processes. When performance in a sensorimotor task, such as, golf putting, is impaired by stereotype threat, self-monitoring processes account for this impairment ([Beilock et al. 2006](#); [Stone et al. 1999](#)). We, therefore, would not assume trait test anxiety to moderate stereotype threat effects with regard to sensorimotor tasks.

Practical implications of the reported findings pertain to examinations of pupils, apprentices, students, applicants, and so forth. Examiners might find it useful to consider multiple sources of performance impairments that can all contribute to a specific test score. Trying to eliminate one unwanted influence potentially facilitates another unwanted influence that is not taken account of. A gender-fair test description, for example, may prevent stereotype threat but thereby enable performance differences due to test anxiety to emerge. Efforts to provide conditions that prevent certain ability-independent performance detriments thus could promote other ability-independent performance detriments to become a relatively stronger influence.

We conclude that trait test anxiety is an important influence on stereotype threat effects. With regard to mathematics test performance, stereotype threat apparently triggers state worry, as does trait worry. The origin of the performance impairment differs, but the mechanism is the same.

## References

- Ambady, N., Paik, S. K., Steele, J., Owen-Smith, A., & Mitchell, J. P. (2004). Deflecting negative self-relevant stereotype activation: The effects of individuation. *Journal of Experimental Social Psychology*, 40, 401–408.
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35, 29–46.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality & Social Psychology*, 51, 1173–1182.
- Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin*, 32, 1059–1071.
- 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.
- Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, 39, 626–633.
- Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? *Psychological Science*, 16, 572–578.
- Darke, S. (1988). Anxiety and working memory capacity. *Cognition & Emotion*, 2, 145–154.

- Derakshan, N., & Eysenck, M. W. (1998). Working memory capacity in high trait-anxious and represser groups. *Cognition & Emotion*, 12, 697–713.
- Dutke, S., & Stöber, J. (2001). Test anxiety, working memory, and cognitive performance: Supportive effects of sequential demands. *Cognition & Emotion*, 15, 381–389.
- Eysenck, M. W. (1979). Anxiety, learning, and memory: A reconceptualization. *Journal of Research in Personality*, 13, 363–385.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58, 47–77.
- 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.
- Ikeda, M., Iwanaga, M., & Seiwa, H. (1996). Test anxiety and working memory system. *Perceptual & Motor Skills*, 82, 1223–1231.
- 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.
- Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, 42, 236–243.
- McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, 27, 486–493.
- MacLeod, C., & Donnellan, A. M. (1993). Individual differences in anxiety and the restriction of working memory capacity. *Personality and Individual Differences*, 15, 163–173.
- Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93, 1314–1334.
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36, 717–731.
- Richards, A., French, C. C., Keogh, E., & Carter, C. (2000). Test anxiety, inferential reasoning and working memory load. *Anxiety, Stress & Coping: An International Journal*, 13, 87–109.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38, 194–201.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85, 440–452.
- Schmader, T., Johns, M., & Barquiseau, M. (2004). The cost of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles*, 50, 835–850.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336–356.
- Schwarzer, R. (1984). Worry and emotionality as separate components in test anxiety. *International Review of Applied Psychology*, 33, 205–220.
- Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety, Stress & Coping: An International Journal*, 4, 27–41.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28.
- Spielberger, C. D. (1980). *Test anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811.
- Stöber, J. (2004). Dimensions of test anxiety: Relations to ways of coping with pre-exam anxiety and uncertainty. *Anxiety, Stress & Coping*, 17, 213–226.
- Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology*, 77, 1213–1227.
- 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.
- Ware, W. B., Galassi, J. P., & Dew, K. M. H. (1990). The test anxiety inventory: A confirmatory factor analysis. *Anxiety Research*, 3, 205–212.

**Tobias Tempel** is research associate in the department of psychology at the University of Trier. His research interests include stereotypes, self-relevant information processing, and memory.

**Roland Neumann** is professor of General Psychology at the University of Trier. His research interests include stereotypes, emotional behavior, and motivation.