Personality and Social Psychology Bulletin

Motivated Cultural Worldview Adherence and Culturally Loaded Test Performance

Published by:

$SAGE

http://www.sagepublications.com

On behalf of:

Society for Personality and Social Psychology

>> Version of Record - Mar 5, 2009

OnlineFirst Version of Record - Jan 20, 2009

What is This?

Motivated Cultural Worldview Adherence and Culturally Loaded Test Performance

Mark J. Landau
University of Kansas

Jeff Greenberg
University of Arizona

Zachary K. Rothschild
University of Kansas

Academic tests and their conditions of administration are culturally loaded when they make salient culturally specific knowledge structures in addition to measuring the intended cognitive ability. Cultural loading demonstrably influences test performance, but why? Drawing on converging perspectives on the psychological function of culture, this article proposes that one factor is the individual's internal motivation to affirm and uphold the cultural worldview. This possibility is tested within the framework of terror management theory, which claims that cultural worldview adherence defends against mortality-related concerns. It is hypothesized that making mortality salient would (a) improve performance on standardized test items when, incidental to the problem structure, the correct answers affirm prevailing cultural stereotypes and (b) impair test performance when excelling violates stereotypic expectancies for one's group. Two studies provide support for these hypotheses. Implications for test validity are briefly discussed.

Keywords: cultural bias; motivation; standardized testing; stereotype threat; terror management theory

Academic tests such as the SAT are commonly used in education admissions and occupational entry decisions because they purportedly measure examinees' cognitive abilities and knowledge. However, beginning with the civil rights movement of the 1950s, psychometric researchers and the general public have vigorously debated whether the contents of academic test items and the conditions of administration are culturally loaded in ways that differentially advantage individuals and groups with dissimilar cultural backgrounds (see, e.g., Helms, 2006; Reynolds & Brown, 1984; Sackett, Borneman, & Connelly, 2008; Wicherts, Dolan, & Hessen, 2005). Most scholarship on this issue concentrates on social structural variables such as gender and socioeconomic status. We sought to complement this approach by addressing the psychological factors that generally influence how individuals interpret and perform on culturally loaded tests (cf. Sternberg, 1980). Cultural loading is conventionally thought to influence test performance by placing differential demands on the examinee's limited processing resources. Inspired by converging perspectives on culture's psychological function, we propose that performance may also be influenced by people's underlying motivation to adhere to their cultural worldview, that is, to affirm cultural beliefs and uphold them in action. We use one such perspective, terror management theory (TMT), to test whether experimentally inducing motivated worldview adherence influences culturally loaded test performance.

Culturally Loaded Tests and Testing Environments

Cognitive tests are culturally loaded when the content of their items reflects cultural knowledge, values, and experiences that are, strictly speaking, irrelevant

442

to the knowledge or skill being assessed. Although intended as a mere vehicle for measuring cognitive abilities, culturally loaded content can nevertheless influence test performance. For example, adults perform poorly on conditional reasoning problems couched in abstract terms and conditions (e.g., "If D then 3") but perform much better when the content (but not structure) of the problem reflects culturally familiar experiences (e.g., sending mail; Wason & Johnson-Laird, 1972).

There is some experimental evidence that cultural loading can result in cultural bias, meaning that differential degrees of familiarity with construct-irrelevant content among groups of individuals with dissimilar cultural backgrounds significantly contributes to group-level performance gaps (cf. Reynolds, 1982). For example, Milton (1959) found that men outperformed women on reasoning problems featuring stereotypically masculine experiences, but this difference was diminished when the content (but not structure) of the problems was altered to reflect stereotypically feminine experiences (see also Rosser, 1989). In another study by Hamdi, Knirk, and Michael (1982), American and Arabic children completed two versions of a test intended to measure spatial ability: one requiring pictorial depth perception of objects and scenes particularly relevant to American culture (e.g., shopping carts) and the other requiring depth perception of content especially common in Arabic culture (e.g., mosques). Although the two versions were structurally parallel, both groups scored higher on the version requiring depth perception of pictorial content representative of their respective culture than on the alternate version and compared to the other group.

In addition to item content, the context of test administration can be culturally loaded when it makes salient construct-irrelevant knowledge or expectations relevant to the examinee's sociocultural identity. This possibility is addressed in a large body of research on stereotype threat (Steele, 1997), which shows that situational reminders of a negative cultural stereotype associated with one's group can increase the individual's apprehension that his or her test performance will confirm that stereotype, and this apprehension can in turn impair performance. For example, studies show that African American students—stereotyped as intellectually inferior to Caucasian students—underperform on a verbal ability test when it is framed as diagnostic of general intelligence, whereas they perform as well as Caucasian students when the same test is framed as nondiagnostic (Steele & Aronson, 1995). Subsequent research has reconfirmed and extended this finding to various groups, performance domains, and means of activating stereotyped sociocultural identities

associated with poor performance (for a review, see Steele, Spencer, & Aronson, 2002).

Why Does Cultural Loading Influence Performance?

The conventional answer to this question is that cultural loading places differential demands on processing resources or working memory. It is less demanding to reason about familiar topics independent of the problem structure (Gigerenzer & Hug, 1992; Wason, 1969), and there is evidence that the increased apprehension elicited by stereotype threat consumes the cognitive resources needed for problem solving (Schmader, Johns, & Forbes, 2008; Steele et al., 2002). We believe this explanation is well supported but may be only one part of the picture. Several theoretical perspectives converge on the notion that people are not passively impinged on by culture; rather, they are internally motivated to adhere to cultural beliefs and ideologies (cultural worldviews) as a means of satisfying other, more distal psychological motives. Although these perspectives emphasize different underlying motives, they have inspired multiple lines of research showing that heightening motivated worldview adherence (usually by way of threat inductions) intensifies people's efforts to affirm cultural worldviews and minimize threats to their validity.

One such perspective is TMT (Solomon, Greenberg, & Pyszczynski, 1991), which proposes that investment in a cultural worldview, and the perception that one is living up to the worldview's standards of conduct, help shield individuals from often unconscious existential fears surrounding death by enabling them to view themselves as valuable members of a cultural reality that persists beyond their own physical demise. Supporting this analysis is a large body of evidence that shows that when thoughts of mortality have been activated but are no longer conscious (i.e., mortality salience [MS]), people respond by bolstering faith in their worldview and defending it against threats (Greenberg, Solomon, & Arndt, 2008, review this work). For example, MS intensifies positive reactions to people who uphold the individual's cultural worldview (e.g., admired political leaders) as well as negative reactions to those who criticize or transgress against one's worldview (e.g., prostitutes among those morally opposed to prostitution; Greenberg et al., 1990; Landau et al., 2004; Rosenblatt, Greenberg, Solomon, Pyszczynski, & Lyon, 1989).

System justification theory (Jost & Hunyady, 2002) also addresses the motivations underlying worldview adherence and proposes that people are motivated to maintain the economic, political, and social status quo by way of various justification strategies that legitimize current cultural systems and practices. Studies have

found that threatening the stability of the prevailing system leads to increased endorsement of system-justifying ideologies and increased reliance on cultural stereotypes (Jost & Banaji, 1994). According to uncertainty management theory (McGregor, Zanna, Holmes, & Spencer, 2001; van den Bos, 2001), cultural worldviews manage feelings of personal uncertainty by providing a definite and reliable conception of the world and oneself. Studies assessing this perspective show, for example, that heightening uncertainty by having participants imagine how a childhood memory could change over time led them to express more negative attitudes toward someone who criticized their university. More recently, Heine, Proulx, and Vohs (2006) proposed the meaning maintenance model, which claims that cultural worldviews provide an extensive network of expected relations that serve the individual's underlying motivation to maintain meaningful conceptions of reality. Supporting evidence shows, for example, that threatening an individual's sense of meaning by presenting a perceptual anomaly leads to increased punativeness toward a prostitute (Proulx & Heine, in press).

Although the effects of motivated worldview adherence have been shown almost exclusively within the realm of social attitudes, we believe that similar motivational processes may influence performance on culturally loaded tests. For example, test items reflecting prevailing stereotypes are not only easier to think about, they also endorse the broader culture and legitimize its dominant beliefs (Zimbardo, Ebbesen, & Maslach, 1977). Likewise, situationally salient stereotypes are not only potentially distracting but may also connect to a stable and socially recognized sociocultural identity. Therefore, performance on culturally loaded tests may be influenced by the (implicit) motivation to affirm cultural stereotypes and to avoid acting in ways that violate them. The current research is, to our knowledge, the first to assess this possibility.

The Current Research

As suggested by the preceding review, cultural worldview adherence is a complex, multidetermined phenomenon, and it has been shown to be affected by diverse psychological motives. In the current research we chose to experimentally induce motivated worldview adherence using TMT's MS paradigm for two reasons. First, a substantial amount of research on the parameters and cognitive processes associated with MS supports the unique role of death concerns in MS effects (see Arndt, Cook, & Routledge, 2004; Greenberg et al., 2008, for more complete discussions). MS has been manipulated in a variety of ways (e.g., explicit death reminders followed by a delay, subliminal "death" primes) and has

been compared with the salience of a variety of topics that are anxiety provoking (e.g., pain, paralysis), future oriented (e.g., upcoming events), self-relevant (e.g., embarrassment, social exclusion), or existential in nature (e.g., uncertainty, meaninglessness). MS-induced worldview adherence is unmediated by subjective arousal or emotion, heightened self-awareness, or high cognitive load (e.g., Greenberg, Simon, Harmon-Jones, et al., 1995). Also, supraliminal MS instigates a unique set of immediate proximal defenses followed by delayed distal terror management defenses involving worldview adherence and self-esteem striving. The latter defenses are triggered by increased accessibility of death-related thought outside of current focal attention, which signals a heightened potential for anxiety. Furthermore, threats to central aspects of an individual's worldview and self-worth increase the accessibility of death-related thought, but not other negative thoughts, and bolstering these structures reduces this accessibility back to baseline levels (e.g., Arndt, Greenberg, Solomon, Pyszczynski, & Simon, 1997; Schimel, Hayes, Williams, & Jahrig, 2007). In short, although some studies have shown that threats other than MS can have similar effects under specific conditions (e.g., McGregor et al., 2001, Study 3; van den Bos, 2001), a much larger body of TMT research strongly indicates the convergent and discriminant validity of MS inductions.

Our second reason for using the MS paradigm is that MS has been shown to intensify worldview adherence in specific ways pertaining to culturally loaded test performance. For example, MS leads Euro-American participants to prefer outgroup (e.g., African American) members who exemplify cultural stereotypes over those who do not (Schimel et al., 1999) and increases people's reliance on stereotypes to describe the ambiguous behavior of others (Renkema, Stapel, Maringer, & van Yperen, 2008). Extrapolating from these findings, we might expect MS to heighten motivation to answer test items in ways that affirm prevailing cultural stereotypes. Also, MS has been uniquely shown to dissuade people from completing behavioral tasks in ways that violate their worldview. Specifically, Greenberg, Simon, Porteus, Pyszczynski, and Solomon (1995) showed that under MS people are less willing to desecrate cultural icons (e.g., using a U.S. flag to sift sand) even when doing so was advantageous in task completion. MS might likewise motivate individuals to perform on tests in ways that uphold salient cultural stereotypes for their group, even when those stereotypes prescribe underachievement. Although these prior findings are suggestive, assuming that similar motivational processes influence academic test performance is, in our opinion, too far an inferential leap to make without direct empirical testing. We therefore designed two studies examining whether

the intensified adherence to one's cultural worldview engendered by MS (a) improves performance on test items when the problem solutions affirm cultural stereotypes in addition to measuring the intended cognitive abilities (Study 1) and (b) impairs test performance when excellence is possible but inconsistent with a salient stereotypic expectancy to perform poorly (Study 2).

STUDY 1

Study 1 was designed to test whether MS would lead participants to respond more quickly and accurately to test items when, incidental to the problem structure, the correct answers affirmed prevailing stereotypes about social groups and institutions. To test this hypothesis, we primed participants with mortality or a control topic (an upcoming exam) and then had them answer six items, adapted from the Graduate Record Examination (GRE), assessing performance in quantitative reasoning, sentence completion, and analogical reasoning. We manipulated cultural loading by adjusting the content (but not the structure) of one of each item type such that, depending on the condition, correct answers affirmed or did not affirm stereotypes about male math superiority, African American intellectual inferiority, and the protective function of the legal system. All participants also completed one of each item type that was neutral with respect to prevailing stereotypes.

Method

Participants and Procedure

Eighty-eight (48 women and 40 men) undergraduates took part in two purportedly unrelated investigations of personality and test taking. They first completed a questionnaire packet in which the MS manipulation followed two filler questionnaires included to bolster the cover story. Participants in the MS condition responded to two open-ended items pertaining to their mortality (used in previous studies; e.g., Rosenblatt et al., 1989): "Please briefly describe the emotions that the thought of your own death arouses in you" and "Jot down, as specifically as you can, what you think will happen to you physically as you die and once you are physically dead." To control for the possibility that the effect of this induction is merely a generalized reaction to reminders of any aversive, uncertain, or future experience, participants in the control condition responded to parallel items pertaining to an upcoming exam. Because MS effects have been found to be strongest after a short delay between the MS induction and the dependent variable assessment (Greenberg, Pyszczynski,

Solomon, Simon, & Breus, 1994), participants completed a self-report mood scale (the 60-item Positive and Negative Affect Schedule–Expanded Form [PANAS-X]; Watson & Clark, 1992) following the MS manipulation. Use of the PANAS-X also allowed us to assess the possibility of affective consequences of the MS inductions and whether or not mood played any mediating role in the hypothesized effects.[1]

Test Items

Participants were then presented with test items and instructed to work as quickly as possible. They were not aware they were being timed (in milliseconds). Six items were presented in a fixed random order: neutral quantitative, stereotype-relevant quantitative, stereotype-relevant sentence completion, neutral sentence completion, neutral analogical reasoning, and stereotype-relevant analogical reasoning. To clarify, all participants completed both stereotype-relevant and stereotype-neutral versions of all three item formats. The conditions differed as to whether or not the stereotype-relevant items were stereotype affirming. Question details follow; correct answers are marked with an asterisk.

Quantitative reasoning. The stereotype-relevant quantitative item presented a graph depicting the temporal progression of Charlie's (*stereotype affirming*) vs. Cindy's (*stereotype nonaffirming*) math education and engineering career. Participants were asked to calculate the target person's total years spent in statistical training. Answer options were: (a) 1.5, (b) 2, (c) 3.5, (d) 4, (e) *4.5. The stereotype-neutral quantitative item presented sales figures for Manufacturers 1, 2, and 3 and asked what percentage of Manufacturer 2's sales was made in a certain month.

Sentence completion. The stereotype-relevant sentence completion item was as follows: "Tyrone (*stereotype affirming*) vs. William (*stereotype nonaffirming*) realized that he had been ___ in his English class. If he had been more ___, the disaster may have been avoided." Answer options: (a) irreproachable . . . careful, (b) *derelict . . . vigilant, (c) neglectful . . . insensible, (d) arbitrary . . . interested, (e) unparalleled . . . careful. The stereotype-neutral item was as follows: "Satire is a marvelous reflection of the spirit of an age; the subtle ___ of Swift's epistles mirrored the Eighteenth century's delight in elegant ___." Answer options: (a) profundity . . . ditties, (b) incongruity . . . pejoratives, (c) contempt . . . anachronisms, (d) provinciality . . . rusticity, (e) *vitriol . . . disparagement.

Analogical reasoning: The stereotype-relevant analogy item presented the analogy stem: "DAM : WATER"

and the following answer options: (a) ENTERTAINMENT : PROHIBITION, (b) CATALYST : REACTION, (c) DUPLICITY : IMPARTIALITY, (d) *LAW : CHAOS (*stereotype affirming*) vs. LAW : SELF-GOVERNANCE (*stereotype nonaffirming*), (e) SUPERFLUITY : BREVITY. The stereotype-neutral analogy item's stem was "NOD : ASSENT" followed with answer options: (a) *SHRUG : INDIFFERENCE, (b) SHUDDER : RUDENESS, (c) WINK : MYSTIFICATION, (d) GLANCE : BENEFICENCE, (e) FROWN : CAPRICIOUSNESS.

A final questionnaire asked "How similar are the test questions you just completed to questions that you've seen on the SAT, ACT, or other standardized tests?" (1 = *not at all similar,* 7 = *extremely similar*). No participant expressed suspicion during postexperimental interviews that the personality and test portions of the experiment were related.

Our prediction that MS would improve performance on stereotype-affirming test items is based on the assumption that the content of those items reflects prevalent cultural stereotypes regarding social groups and institutions. We performed a pilot investigation to confirm this assumption. We had 46 (23 women and 23 men) participants drawn from the same subject pool as in Study 1 respond to the following items given the current state of American society: (a) Who is more likely to have a successful career in engineering and statistics? (Cindy/Charlie); (b) Who is more likely to make a mistake on his English class assignment? (William/Tyrone); and (c) Which of the following is our legal system better designed to combat? (chaos/self-governance). Both the items and each item's response options were presented in a randomized order. As expected, the majority of participants chose Charlie (39/46), $\chi^2(1) = 22.26$, $p < .001$; Tyrone (41/46), $\chi^2(1) = 28.17$, $p < .001$; and chaos (33/46), $\chi^2(1) = 22.26$, $p < .005$. Further analyses revealed no main effect or interactions for participant gender, χ^2s $< .22$, *ns*. These results strengthen our assumption that the stereotype-affirming test items were perceived as reflecting prevailing cultural stereotypes.

## Results and Discussion

Preliminary Analyses

Preliminary analysis revealed no hint that our experimental manipulations influenced test item typicality ratings (all Fs < 1, $p > .40$) or that these ratings mediated our primary analyses. We are therefore confident that the performance results are not due to the stereotype-relevant questions standing out as atypical. The grand mean for this item was 3.4 ($SD = 1.72$).

Performance

Responses on the three stereotype-relevant items and on the three neutral items showed relatively low internal
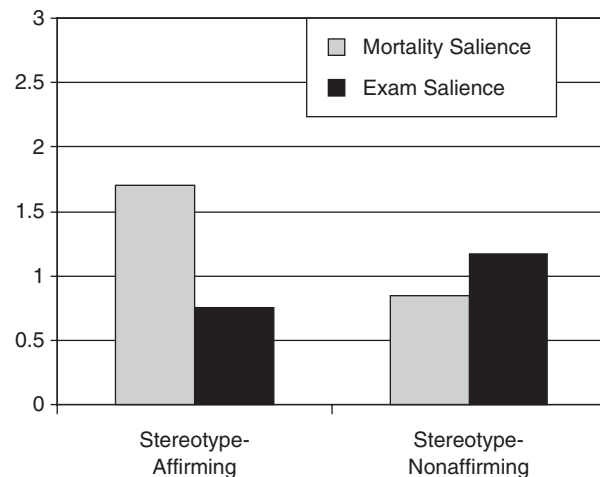
Figure 1 Performance on stereotype-relevant test items as a function of mortality salience and cultural loading.
NOTE: Number of questions answered correctly. Scale ranged from 0 to 3.
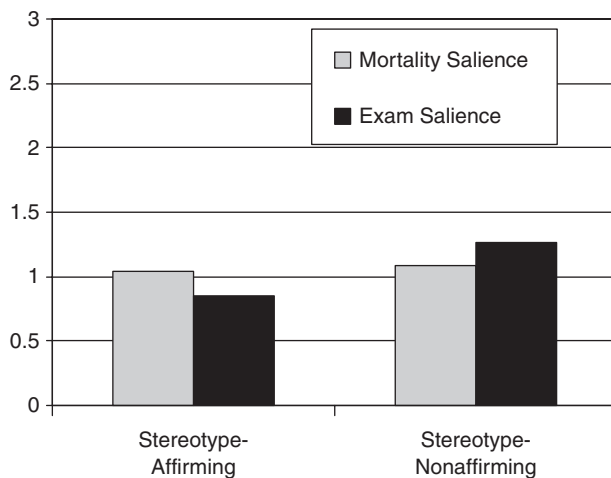
Figure 2 Performance on stereotype-neutral test items as a function of mortality salience and cultural loading.
NOTE: Number of questions answered correctly. Scale ranged from 0 to 3.

reliability coefficients (αs $= .58$ and $.53$, respectively). This was somewhat expected given that each composite was composed of only three items designed to assess distinct cognitive abilities. As an initial examination of the data, we summed correct responses (coded as 1; incorrect responses coded as 0) to the stereotype-relevant and neutral items separately and submitted those composite scores to a 2 (MS vs. exam) × 2 (cultural loading: stereotype affirming vs. nonaffirming) × 2 (question type: stereotype relevant vs. neutral) mixed ANOVA with

TABLE 1: Number of Correct Responses to Standardized Test Items Assessing Quantitative Reasoning, Sentence Completion, and Analogical Reasoning as a Function of Item Type (Stereotype Relevant vs. Neutral; Within Subjects), Cultural Loading (Stereotype Affirming vs. Nonaffirming), and Mortality Salience

	Stereotype-Relevant Item		Stereotype-Neutral Item	
	Stereotype Affirming	Stereotype Nonaffirming	Stereotype Affirming	Stereotype Nonaffirming
Quantitative reasoning				
Mortality salience	13 ($n = 25$)	8 ($n = 21$)	3 ($n = 25$)	4 ($n = 21$)
Exam salience	4 ($n = 20$)	8 ($n = 22$)	1 ($n = 20$)	5 ($n = 22$)
Sentence completion				
Mortality salience	12	2	6	4
Exam salience	4	5	5	6
Analogical reasoning				
Mortality salience	17	8	17	15
Exam salience	7	13	11	17

question type serving as a within-subjects factor.[2] A marginal Cultural Loading × Question Type interaction, $F(1, 84) = 3.55$, $p = .06$, was qualified by the predicted three-way interaction, $F(1, 84) = 4.18$, $p = .04$ (all other Fs < 1, ps > .41). To interpret this effect we conducted separate MS × Cultural Loading ANOVAs for each question type. Looking at stereotype-relevant items, we observed only the predicted two-way interaction, $F(1, 84) = 8.82$, $p = .004$. Pairwise comparisons (least significant difference) and the means presented in Figure 1 show that, as predicted, mortality salient participants answered more questions correctly when the correct answers affirmed prevailing cultural stereotypes as compared to exam salient participants responding to stereotype-affirming items ($F = 7.74$, $p = .007$) and mortality salient participants responding to stereotype-nonaffirming items ($F = 8.80$, $p = .004$). Mortality salient participants also tended to score lower on stereotype-nonaffirming items compared to exam salient participants, $F = 2.10$, $p < .15$. Within the exam salience condition, the difference between stereotype-affirming and stereotype-nonaffirming items did not reach significance ($F = 1.50$, $p = .23$). Looking only at performance on stereotype-neutral items, the same MS × Cultural Loading ANOVA revealed no significant effects (all Fs < 1.4, ps > .24; means are presented in Figure 2).

Because the internal reliability coefficients for the items were relatively small, we performed a series of hierarchical logistic regression analyses in which performance (1 = *correct response,* 0 = *incorrect response*) was predicted from MS and cultural loading (at Step 1) and their interaction (at Step 2). Table 1 presents the number of correct responses to the stereotype-relevant and stereotype-neutral items by condition.

For the quantitative reasoning item, we observed a main effect for MS ($B = 1.05$, $SE = .49$, $p = .04$), which was qualified by a marginally significant two-way interaction

($B = 1.74$, $SE = .97$, $p = .07$). Consistent with our predictions, mortality salient participants were more likely to correctly answer the stereotype-affirming item than were exam salient participants ($B = 1.99$, $SE = .76$, $p = .01$). Within the MS condition, participants were not significantly more likely to correctly answer the stereotype-affirming item than the stereotype-nonaffirming item ($B = .61$, $SE = .62$, $p = .33$), although the pattern of responses was in the predicted direction.

For the sentence completion item, we observed a main effect for stereotype relevance ($B = 1.02$, $SE = .51$, $p = .05$), which was qualified by the predicted two-way interaction ($B = 2.33$, $SE = 1.14$, $p = .04$). As predicted, mortality salient participants outperformed exam salient participants on the stereotype-affirming item ($B = 1.31$, $SE = .67$, $p = .05$) as well as mortality salient participants responding to the stereotype-nonaffirming item ($B = 2.17$, $SE = .84$, $p = .01$).

Finally, looking at the analogy item separately we again found the predicted two-way interaction ($B = 2.23$, $SE = .89$, $p = .01$). As expected, mortality salient participants outperformed exam salient participants on the stereotype-affirming item ($B = 1.37$, $SE = .63$, $p = .03$) as well as mortality salient participants responding to the stereotype-nonaffirming item ($B = 1.24$, $SE = .62$, $p = .05$).

Submitting responses to the stereotype-neutral items to the same logistic regression analyses revealed no significant main effects or interactions (see Table 1 for performance statistics by condition).

Performance Speed

Reaction times (in milliseconds) for stereotype-relevant and neutral items were separately averaged (and converted to seconds) and submitted to a 2 (MS) × 2 (cultural loading) × 2 (question type) ANOVA with question type serving as a within-subjects factor. The three-way
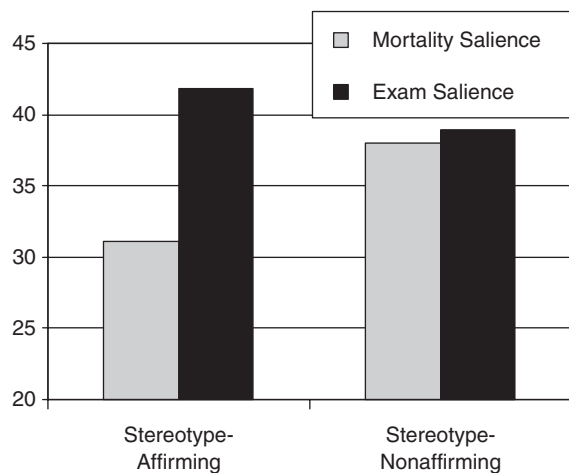
Figure 3 Performance speed (in seconds) on stereotype-relevant test items as a function of mortality salience and cultural loading.
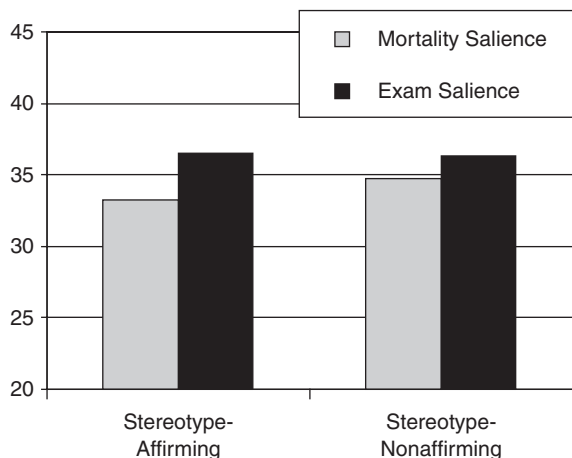
Figure 4 Performance speed (in seconds) on stereotype-neutral test items as a function of mortality salience and cultural loading.

interaction was marginally significant, $F(1, 84) = 3.42$, $p = .06$. We then conducted separate MS × Cultural Loading ANOVAs for each question type. Looking at the stereotype-relevant items, we observed an MS main effect such that mortality salient participants were overall faster ($M = 36.44$, $SD = 8.94$) than exam salient participants ($M = 38.45$, $SD = 10.70$), $F(1, 84) = 6.80$, $p = .01$. This was qualified by the predicted two-way interaction ($F = 4.49$, $p = .04$; see Figure 3 for means). As predicted, mortality salient participants responded faster to stereotype-affirming items as compared to exam salient participants responding to stereotype-affirming

items ($F = 11.71$, $p = .001$) and mortality salient participants responding to stereotype-nonaffirming items ($F = 4.19$, $p = .04$); no other pairwise comparisons attained significance ($Fs < 1$, $ps > .35$). Submitting stereotype-neutral item performance speed to the same two-way ANOVA revealed no significant effects (all $Fs < 1$, $ps > .44$; see means in Figure 4).

The results of Study 1 confirmed our hypothesis that MS would improve performance on standardized test items when, incidental to the problem structure, the correct answers affirmed prevalent cultural stereotypes. As in prior TMT research, the effects of MS were found to be different from making salient an upcoming, uncertain, and aversive event (an upcoming test) and influenced responses specifically to worldview-relevant stimuli without generally influencing performance on worldview-neutral items. The latter finding furthermore suggests that the observed effects are not due to MS simply creating a cognitive load (in line with, e.g., Greenberg, Simon, Porteus, et al., 1995). However, these findings substantially extend prior TMT research by showing that MS-induced worldview adherence influences the speed and accuracy with which people respond to cognitive ability measures that are subtly loaded with cultural content. These findings suggest that a previously undocumented motivational factor in culturally loaded test performance is people's internal motivation to adhere to their cultural worldview, which leads them to approach worldview-affirming items more comfortably and enthusiastically than worldview-nonaffirming items. We sought to extend this finding in Study 2 by examining whether MS-induced worldview adherence influences performance on tests that are culturally loaded by the salience of stereotypic expectations for the performance of one's own group.

STUDY 2

According to TMT, stereotypes function in part to sustain a meaningful and stable understanding of the world (Allport, 1954) that functions ultimately to hold mortality concerns at bay. As discussed, research supports this claim by showing that MS heightens both a reliance on stereotypes when describing others and a preference for stereotype-affirming outgroup members. But one implication of this research that has not yet received empirical attention is that terror management processes may also underlie commitment to *self*-relevant group stereotypes, even if they are associated with underachievement. Some aforementioned research by Greenberg, Simon, Porteus, et al. (1995) shows that MS heightens reluctance to act in ways that violate culturally sacred symbols, but no research has yet

addressed whether MS similarly increases reluctance to act in ways that contradict a self-relevant stereotype. Combining and extending prior work, we designed Study 2 to test whether women who were primed with mortality and subsequently stereotyped to perform poorly on an academic test would perform poorly when they had the opportunity to excel.

To test this hypothesis, we manipulated MS by priming women with either mortality or an aversive control topic (dental pain) and then assessed their performance in the domain of mental spatial rotation. We manipulated stereotype threat by telling half of the participants that women are stereotyped as generally less competent than men in mental spatial rotation. We then had participants take both an easy and a difficult spatial rotation test to test two specific predictions. Research shows that increasing the salience of a negative cultural stereotype impairs performance specifically on complex or demanding tasks at least in part because of the additional cognitive load resulting from stereotype-activated apprehensions (O'Brien & Crandall, 2003; see Schmader, Johns, & Barquissau, 2008). We therefore predicted that, independent of making mortality salient, stereotype threat would impair performance on a difficult spatial rotation test.

But what about performance on an easy test? Here examinees have the opportunity to excel, but even the implicit realization that one is acting contrary to stereotypic expectancies may undermine the meaning conferred by a self-relevant group stereotype. Insofar as that stereotype is part of the prevailing worldview that serves a terror management function, MS should introduce a motivation to exemplify self-relevant stereotypic expectancies (even negative ones) and thus a reluctance to excel beyond them. More specifically, we hypothesize that women who are primed with mortality and subsequently stereotyped to perform poorly on an academic test would perform poorly even when it is easy to do well.

Method

Eighty-four female undergraduates were tested in a classroom setting in small groups. At least 4 male participants were present at all sessions to minimize suspicion that recruitment was gender specific (see, e.g., Schmader et al., 2004). Results from male participants were not analyzed because the final sample size was very small ($n = 19$) and because we were interested in underperformance among women. A Euro-American male experimenter first administered a packet of questionnaires containing the MS manipulation. Participants in the MS condition completed the same open-ended items pertaining to one's death used

in Study 1. Participants in the aversive control condition completed parallel questions regarding dental pain. The PANAS-X served as the necessary delay and distraction.

Stereotype Threat

Once all participants had completed the packets, the experimenter explained that for the second experiment they would complete a test of mental spatial rotation, or the ability to rotate objects "in one's head." In line with previous research (e.g., Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995), the experimenter told participants in the stereotype threat condition that the task was sensitive to gender differences and that women were stereotyped as generally less competent than men in spatial rotation ability (specifically, "One thing we'll look at is how men and women differ in their performance on the test, and how true the stereotype, or the generally held belief is, that women have more trouble with spatial rotation tasks"). In the no-threat condition, the experimenter explained that the test was being piloted for future research. Stereotype threat effects have been successfully replicated in this domain using similar methods (e.g., Martens, Johns, Greenberg, & Schimel, 2006).

The experimenter urged all participants to give a strong effort when taking the test. To encourage effort on the easy test he explained that test scores would be determined by the number of correct responses as well as the total number of items attempted.

Easy and Difficult Tests

Each participant received an easy and a difficult test, counterbalanced in order. The tests were adapted from Vandenberg and Kuse's (1978) Mental Rotation Test. For both versions, each item depicted a target shape and four comparison shapes: two rotated versions of the target figure and two distracters. Participants were instructed to mentally rotate the comparison shapes to determine which two matched the target shape, and they were told that correct responses required identification of both comparison shapes. The format and instructions for the easy and difficult versions were identical; the tests differed only in the number of items and the type of shape used. The difficult test had 24 items and used three-dimensional shapes developed by Shepard and Metzler (1971; see Figure 5). Because they are three-dimensional and complex, these shapes are relatively difficult to mentally rotate. The easy test had 72 items and used more distinctive shapes (adapted from Tarr & Pinker, 1990) rotated in only two dimensions. We used more items for the easy test to avert a ceiling effect on performance.
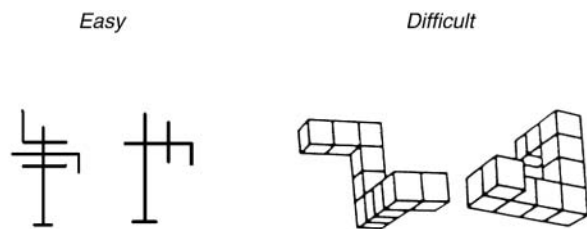
Figure 5 Sample shapes used in the easy and difficult spatial rotation tests (Study 2).

After completing a practice page to ensure that they understood the task, participants were given 6 min to work on each test. Finally, participants completed two test difficulty manipulation check questions: "How difficult was the first test you took?" and "How difficult was the second test you took?" (1 = *not at all difficult*, 9 = *very difficult*). Finally, participants were asked "To what extent do you believe that females are inferior to males on spatial rotation ability?" (1 = *not at all*, 9 = *a lot*).

Results and Discussion

Manipulation Check Questions

We submitted test difficulty ratings to a 2 (difficult vs. easy test; within) × 2 (test order; between) mixed ANOVA. Only the predicted test version effect emerged (all other $Fs < 1$). As expected, participants found the difficult test to be more difficult ($M = 7.10$, $SD = 1.48$) than the easy test ($M = 2.34$, $SD = 1.30$), $F(1, 82) = 408.11$, $p < .01$. A 2 (MS vs. pain) × 2 (stereotype threat vs. no threat) × 2 (test version order) ANOVA on the stereotype endorsement question revealed no significant main effects or interactions (all $Fs < 1.6$, $ps > .2$). The overall mean for this item was 3.6 (on a 9-point scale).[3]

Performance

Spatial rotation performance was analyzed using the number of items correctly answered for each test. Because the tests had different numbers of items, we used z-transformed scores for the repeated measures analysis including test version and nontransformed scores when analyzing easy and difficult test performances separately. We submitted transformed test scores to a 2 (MS) × 2 (stereotype threat) × 2 (easy vs. difficult test) × 2 (test version order) ANOVA with test difficulty serving as a within-subjects factor. We observed only the predicted three-way interaction among MS, stereotype threat, and test version, $F(1, 76) = 5.10$, $p = .03$ (all other $Fs < 1.10$, $ps > .30$). We performed separate ANOVAs on each test.
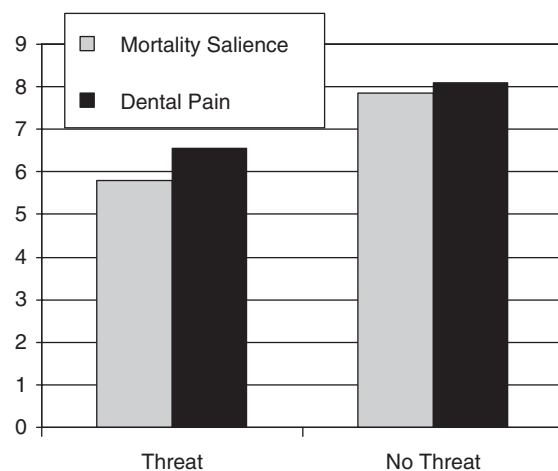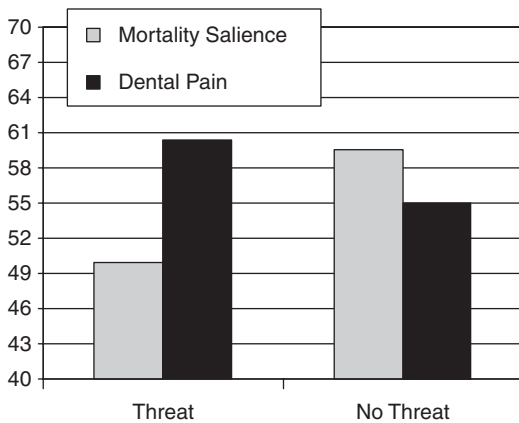

Figure 6 Performance on the difficult spatial rotation test as a function of mortality salience and stereotype threat (Study 2).
NOTE: Number of questions answered correctly. Scale ranged from 0 to 24.

Difficult test. We expected stereotype threat to impair difficult test performance. A 2 (MS) × 2 (stereotype threat) × 2 (test order) ANOVA revealed only the main effect for stereotype threat, $F(1, 76) = 4.22$, $p = .04$ (all other $Fs < 1$, $ps > .44$; see Figure 6). As predicted, women under stereotype threat correctly answered fewer questions ($M = 6.18$, $SD = 3.16$) than women in the no-threat condition ($M = 7.98$, $SD = 4.14$).

Easy test. Submitting easy test scores to the same analysis revealed only the predicted MS × Threat interaction, $F(1, 76) = 5.31$, $p = .02$ (all other $Fs < 1$, $ps > .36$; see Figure 7). Mortality salient and pain salient women performed equally well on the easy spatial rotation test in the no-threat condition. When exposed to the negative stereotype, however, mortality salient women performed significantly worse than pain salient women, $F(1, 76) = 5.57$, $p = .02$. Furthermore, within the MS condition, women under stereotype threat performed worse than women not under threat, $F(1, 76) = 4.16$, $p = .05$. There was also an unexpected but nonsignificant trend among pain salient women to perform better on the easy test under stereotype threat, $F(1, 76) = 1.83$, $p = .18$.[4]

In sum, women performed relatively poorly on a challenging test of spatial rotation when a negative group-based performance stereotype was made salient, thus replicating the basic stereotype threat effect (e.g., Martens et al., 2006). More pertinent for our current purposes, women who were stereotyped as deficient in spatial rotation performed poorly on an *easy* test when mortality was salient, but not otherwise. Our

Figure 7 Performance on the easy spatial rotation test as a function of mortality salience and stereotype threat (Study 2).
NOTE: Number of questions answered correctly. Scale ranged from 0 to 72.

interpretation of this finding is that when taking the easy test, female participants realized they could readily excel, thereby disconfirming the cultural gender stereotype; hence, there was no stereotype effect in the control prime condition. However, when mortality concerns were made salient along with the gender stereotype, participants underperformed because they became reluctant to act in a way that could threaten the culture's stereotypic view of their group. The results of this study are the first to show that MS leads people to act in accordance with self-relevant cultural stereotypes and, furthermore, that this effect holds even when the self-relevant stereotype prescribes underperformance on an ego-relevant task.

We are, of course, not claiming that the aversive effects of stereotype threat on intellectual performance are due entirely to a motivated tendency to exemplify self-relevant stereotypes. Indeed, research indicates that for difficult tasks, these effects are primarily the result of cognitive load and arousal (Schmader et al., 2008). However, evidence indicates that stereotype threat effects are moderated by salient social identities (Marx & Stapel, 2006; Shih, Pittinsky, & Ambady, 1999), so we think it is plausible that stereotype threat effects are sometimes affected by a reluctance to excel beyond a negative stereotypic expectation that is consistent with a group identity that defines one's role within the broader social context. The current study suggests that when, as a result of a death reminder, motivation to uphold the worldview is particularly high, salient negative stereotypes will lead to poor performance even when performance is not likely to be hampered by cognitive load.

GENERAL DISCUSSION

The present studies found that MS-induced concern with affirming and upholding the prevailing cultural worldview improves performance when test items support one's worldview and impairs performance when poor performance is prescribed by the self-relevant stereotypes of that worldview. These findings suggest that reminders of death can thereby contribute to relative performance advantages for members of the majority group on tests supportive of the cultural mainstream and in contexts in which negative stereotypes of other groups are salient. This has practical significance because reminders of death can be activated by life events involving friends and family, reading a newspaper or other experiences on the way to a test site, course content, or by test items themselves, such as reading comprehension stories focused on deadly diseases, war, and natural disasters.

Along with reminders of death, other factors that increase motivation to uphold the prevailing worldview may also aid the culturally advantaged group members and hamper culturally disadvantaged group members. It may be that any threats to the individual's worldview or self-worth, stressors in the home or at school, or societal threats of external attacks or economic downturn also increase embracing of the mainstream worldview (e.g., Landau et al., 2004; Sales, 1972). In the broadest sense, then, the current work suggests that any conditions that increase the need for validation of the mainstream worldview will make culturally loaded tests and stereotype-activating testing environments especially disadvantageous to marginalized and stigmatized groups within a culture.

A great deal of scholarship concerns the valid interpretation of test scores based on items reflecting dissimilar cultural content. Controversy surrounds whether differential familiarity with the item content of standardized tests (e.g., the Wechsler Intelligence Scale for Children) illegitimately discriminates against certain groups (e.g., minority populations, children from low socioeconomic backgrounds; Horn & Goldsmith, 1981) and undermines the tests' validity as a measure of academic aptitude. The current findings add to the weight of evidence that these are legitimate concerns by showing that, when motivation to adhere to the cultural worldview is high, culturally loaded test items supporting the mainstream worldview clearly advantage examinees, and test environments that make negative stereotypes salient clearly disadvantage those targeted by such stereotypes.

Developing measures and testing environments that minimize such biases is thus a worthy albeit difficult goal. Some have sought culture-free tests to minimize factors that may unfairly advantage or disadvantage the

performance of some groups by making the content material familiar to all groups (Oakland, 1982), but others say that culture-free measures of intellectual capabilities are impossible because intelligent behavior is defined on the basis of what behaviors the culture judges to be of value. Still others have sought to reduce bias in cognitive ability assessment by making culture-fair tests that equalize the exposure of different groups to different racial, cultural, and socioeconomic socialization experiences. For example, since the 1970s test publishers have sought to reduce sex bias in standardized assessments by adjusting the content of test items to equalize gender-specific linguistic markers and portrayals of both sexes engaged in a variety of activities (e.g., Stern, 1972). Others have sought to minimize features of testing procedures that bias test performance of different groups (e.g., Fuchs & Fuchs, 1986; Sattler, 1988). In light of evidence that scores on the SAT and ACT are influenced by examinees' socioeconomic status, race, and other factors not related to their ability, a growing number of colleges and universities are easing reliance on standardized test scores in admissions decisions (Rimer, 2008).

Although the current research supports these efforts, it also suggests that factors that influence people's internal motivation to support the prevailing worldview should be considered as well, especially because purely unbiased tests and test environments are so hard to achieve. The question then becomes how to avoid increasing students' motivation to adhere to the prevailing worldview. Safe schools, homes, and neighborhood environments, and a peaceful, thriving society would all help, but these are of course hard things to accomplish.

Another approach would be to bolster the individual's psychological resources. Terror management research shows that anything that bolsters faith in people's worldview and sense of self-worth reduces the need to bolster the prevailing worldview (for a review, see Greenberg et al., 2008). Supporting this idea in the context of test performance, research shows that self-affirmations reduce the effects of stereotype threat in the lab and can improve minority group performance in the classroom (Good, Aronson, & Inzlicht, 2003, Martens et al., 2006).

A final point perhaps worth noting is that these results have implications for test validity beyond the academic realm. Certainly tests of personality, vocational skill, and psychopathology are culturally loaded, and our analysis suggests that the individual's motivation to identify with and uphold their cultural worldview may further undermine the validity of those tests as well.

NOTES

1. For both studies, we assessed whether mortality salience (MS) affected mood by performing MANOVAs and ANOVAs on the various subscales of the Positive and Negative Affect Schedule–Expanded Form (PANAS-X) and ANOVAs on the aggregate positive and negative affect scores using our primary predictors. Consistent with previous terror management theory research demonstrating that MS does not engender self-reported affect, these analyses revealed no effects. To ensure that the MS effects were not mediated by affect, we conducted ANCOVAs with the affect subscale scores as covariates, and our primary predicted effects remained significant.

2. These analyses were originally conducted with gender as a between-subjects factor. Because we observed no main effects or interactions involving gender, we do not further report gender effects.

3. This is consistent with research by Steele, Spencer, and Aronson (2002) and Wheeler and Petty (2001) showing that stereotype threat effects are not affected by whether the individual endorses the stereotype but rather whether he or she is simply aware of it.

4. The trend for women in the control prime condition to perform better on an easy test while under stereotype threat is consistent with O'Brien and Crandall's (2003) finding that stereotype threat (in the absence of MS) can improve performance on an easy task.

REFERENCES

Allport, G. W. (1954). *The nature of prejudice.* Reading, MA: Addison-Wesley.

Arndt, J., Cook, A., & Routledge, C. (2004). The blueprint of terror management: Understanding the cognitive architecture of psychological defense against the awareness of death. In J. Greenberg, S. Koole, & T. Pyszczynski (Eds.), *Handbook of experimental existential psychology* (pp. 35-53). New York: Guilford.

Arndt, J., Greenberg, J., Solomon, S., Pyszczynski, T., & Simon, L. (1997). Suppression, accessibility of death-related thoughts, and cultural worldview defense: Exploring the psychodynamics of terror management. *Journal of Personality and Social Psychology, 73,* 5-18.

Fuchs, D., & Fuchs, L. S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research, 56,* 243-262.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43,* 127-171.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24,* 645-662.

Greenberg, J., Pyszczynski, T., Solomon, S., Rosenblatt, A., Veeder, M., Kirkland, S., et al. (1990). Evidence for terror management II: The effects of mortality salience on reactions to those who threaten or bolster the cultural worldview. *Journal of Personality and Social Psychology, 58,* 308-318.

Greenberg, J., Pyszczynski, T., Solomon, S., Simon, L., & Breus, M. (1994). Role of consciousness and accessibility of death-related thoughts in mortality salience effects. *Journal of Personality and Social Psychology, 67,* 627-637.

Greenberg, J., Simon, L., Harmon-Jones, E., Solomon, S., Pyszczynski, T., & Chatel, D. (1995). Testing alternative explanations for mortality effects: Terror management, value accessibility, or worrisome thoughts? *European Journal of Social Psychology, 12,* 417-433.

Greenberg, J., Simon, L., Porteus, J., Pyszczynski, T., & Solomon, S. (1995). Evidence of a terror management function of cultural icons: The effects of mortality salience on the inappropriate use of cherished cultural symbols. *Personality and Social Psychology Bulletin, 21,* 1221-1228.

Greenberg, J., Solomon, S., & Arndt, J. (2008). A basic but uniquely human motivation: Terror management. In J. Y. Shah & W. L.

Gardner (Eds.), *Handbook of motivation science.* (pp. 114-134). New York: Guilford.

Hamdi, N., Knirk, F., & Michael, W. B. (1982). Differences between American and Arabic children in performance on measures of pictorial depth perception: Implications for valid interpretation of test scores based on items reflecting dissimilar cultural content. *Educational and Psychological Measurement, 42,* 285-296.

Heine, S. J., Proulx, T., & Vohs, K. D. (2006). The meaning maintenance model: On the coherence of social motivations. *Personality and Social Psychology Review, 10,* 88-110.

Helms, J. E. (2006). Fairness is not validity or cultural bias in racial-group assessment: A quantitative perspective. *American Psychologist, 61,* 845-859.

Horn, J., & Goldsmith, H. (1981). Reader be cautious: A review of bias in mental testing. *American Journal of Education, 89,* 305-329.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology, 33,* 1-27.

Jost, J. T., & Hunyady, O. (2002). The psychology of system justification and the palliative function of ideology. In W. Stroebe & M. Hewstone (Eds.) *European review of social psychology* (pp. 111-153). Hove, UK: Psychology Press.

Landau, M. J., Solomon, S., Greenberg, J., Cohen, F., Pyszczynski, T., Arndt, J., et al. (2004). Deliver us from evil: The effects of mortality salience and reminders of 9/11 on support for President George W. Bush. *Personality and Social Psychology Bulletin, 30,* 1136-1150.

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.

Marx, D. M., & Stapel, D. (2006). It's all in the timing: Measuring emotional reactions to stereotype threat before and after taking a test. *European Journal of Social Psychology, 36,* 687-698.

McGregor, I., Zanna, M. P., Holmes, J. G., & Spencer, S. J. (2001). Compensatory conviction in the face of personal uncertainty: Going to extremes and being oneself. *Journal of Personality and Social Psychology, 80,* 472-488.

Milton, G. A. (1959). Sex difference in problem solving as a function of role appropriateness of problem content. *Psychological Reports, 5,* 705-708.

Oakland, T. (1982). Nonbiased assessment of minority group children. In J. T. Nersworth (Ed.), *Assessment in special education* (pp. 147-162). Rockville, MD: Aspen Systems.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29,* 782-789.

Proulx, T., & Heine, S. J. (in press). The case of the transmogrifying experimenter: Reaffirmation of moral schemas following implicit change detection. *Psychological Science.*

Renkema, L. J., Stapel, D., Maringer, M., & van Yperen, N. W. (2008). Terror management and stereotyping: Why do people stereotype when mortality is salient? *Personality and Social Psychology Bulletin, 34,* 553-564.

Reynolds, C. R. (1982). The problem of bias in psychological assessment. In C. Reynolds & T. Gutkin (Eds.), *The handbook of school psychology* (pp. 178-208). New York: Wiley.

Reynolds, C. R., & Brown, R. T. (Eds.). (1984). *Perspectives on bias in mental testing.* New York: Plenum.

Rimer, S. (September 22, 2008). College panel calls for less focus on SATs. *New York Times,* p. A14.

Rosenblatt, A., Greenberg, J., Solomon, S., Pyszczynski, T., & Lyon, D. (1989). Evidence for terror management theory I: The effects of mortality salience on reactions to those who violate or uphold cultural values. *Journal of Personality and Social Psychology, 57,* 681-690.

Rosser, P. (1989). *The SAT gender gap: Identifying the causes.* Washington, DC: Center for Women Policy Studies.

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist, 63,* 215-227.

Sales, S. M. (1972). Economic threat as a determinant of conversion rates in authoritarian and nonauthoritarian churches. *Journal of Personality and Social Psychology, 23,* 420-428.

Sattler, J. M. (1988). *Assessment of children.* San Diego, CA: Author.

Schimel, J., Hayes, J., Williams, T., & Jahrig, J. (2007). Is death really the worm at the core? Converging evidence that worldview threat increases death-thought accessibility. *Journal of Personality and Social Psychology, 92,* 789-803.

Schimel, J., Simon, L., Greenberg, J., Pyszczynski, T., Solomon, S., Waxmonsky, J. et al. (1999). Stereotypes and terror management: Evidence that mortality salience enhances stereotypic thinking and preferences. *Journal of Personality and Social Psychology, 77,* 905-926.

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs 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.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701-703.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10,* 80-83.

Solomon, S., Greenberg, J., & Pyszczynski, T. (1991). A terror management theory of social behavior: The psychological functions of self-esteem and cultural worldviews. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 93-159). San Diego, CA: Academic Press.

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.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613-629.

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.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 379-440). San Diego, CA: Academic Press.

Stern, M. (1972). *Changing sexist practices in the classroom* (Ed.).Washington, DC: Women's Rights Committee of the American Federation of Teachers, AFL-CIO.

Sternberg, R. J. (1980). Intelligence and test bias: Art and science. *Behavioral and Brain Sciences, 3,* 353-354.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science, 1,* 253-256.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47,* 599-604.

van den Bos, K. (2001). Uncertainty management: The influence of uncertainty salience on reactions to perceived procedural fairness. *Journal of Personality and Social Psychology, 80,* 931-941.

Wason, P. C. (1969). Regression in reasoning? *British Journal of Psychology, 60,* 471-480.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* London: Batsford.

Watson, D., & Clark, L. A. (1992). Affects separable and inseparable: On the hierarchical arrangement of the negative affects. *Journal of Personality and Social Psychology, 62,* 489-505.

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127,* 797-826.

Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology, 89,* 697-716.

Zimbardo, P., Ebbesen, E., & Maslach, C. (1977). *Influencing attitudes and changing behavior* (2nd ed.). Reading, MA: Addison-Wesley.