THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2010, 63 (8), 1619–1630

Psychology Press
Taylor & Francis Group

Diagnosing and alleviating the impact of performance pressure on mathematical problem solving

Marci S. DeCaro, Kristin E. Rotar and Matthew S. Kendra
Miami University, Oxford, OH, USA

Sian L. Beilock
The University of Chicago, Chicago, IL, USA

High-pressure academic testing situations can lead people to perform below their actual ability levels by co-opting working memory (WM) resources needed for the task at hand (Beilock, 2008). In the current work we examine how performance pressure impacts WM and design an intervention to alleviate pressure's negative impact. Specifically, we explore the hypothesis that high-pressure situations trigger distracting thoughts and worries that rely heavily on verbal WM. Individuals performed verbally based and spatially based mathematics problems in a low-pressure or high-pressure testing situation. Results demonstrated that performance on problems that rely heavily on verbal WM resources was less accurate under high-pressure than under low-pressure tests. Performance on spatially based problems that do not rely heavily on verbal WM was not affected by pressure. Moreover, the more people reported worrying during test performance, the worse they performed on the verbally based (but not spatially based) maths problems. Asking some individuals to focus on the problem steps by talking aloud helped to keep pressure-induced worries at bay and eliminated pressure's negative impact on performance.

Keywords: Pressure; Working memory; Anxiety; Mathematics; Attention

There is a large body of research examining the cognitive processes underlying mathematical problem solving (see DeStefano & LeFevre, 2004, for a review). Less work, however, has addressed how these cognitive processes are impacted by the high-stakes testing situations in which mathematics performance frequently takes place. Students often desire to perform at their best in these important testing situations, but ironically pressure-filled environments often lead people to perform below their capabilities. "Choking under pressure" is the term used to describe the phenomenon by which high-pressure situations lead to less than optimal performance outcomes (see Beilock, 2008, for a review). Given the widespread reliance on tests in education settings, understanding how pressure-filled situations impact academic performance is crucial for test score interpretation. Such knowledge can also foster interventions that reduce unwanted performance decrements—

Correspondence should be addressed to Marci S. DeCaro, Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Place, Peabody #552, Nashville, TN 37203, USA. E-mail: marci.decaro@vanderbilt.edu

Marci Sammons DeCaro is now in the Department of Psychology and Human Development at Vanderbilt University. Research was supported by NSF (National Science Foundation) CAREER Grant DRL-0746970 to Sian Beilock.

DOI:10.1080/17470210903474286

helping test-takers demonstrate their actual abilities when it matters most.

Why does "choking under pressure" occur in important academic testing situations? One explanation is that high-pressure situations lead people to worry about their performance and its consequences (e.g., Beilock, 2008; Beilock & Carr, 2001). These worries are thought to compete for the working memory (WM) available for performance. WM is a short-term system involved in the control, regulation, and active maintenance of a limited amount of information immediately relevant to the task at hand (Miyake & Shah, 1999). If the ability of WM to maintain task focus is disrupted, performance may suffer.

Support for pressure's disruption of task-relevant WM may be most evident when examining how people perform on different types of mathematics problems across low- and high-pressure tests. Several studies have found that performance of mathematics problems that involve multiple intermediate computations, and thus tax WM resources, is hurt under high-pressure compared to low-pressure testing conditions. In contrast, the performance of simple single-step problems, or complex multistep problems that have been so extensively practised that their answers can be directly retrieved from long-term memory (and thus circumvent the usual demand on WM; Logan, 1988), is unaffected (Beilock & Carr, 2005; Beilock, Kulp, Holt, & Carr, 2004). Pressure appears to co-opt the WM resources needed for executing other cognitively demanding academic skills like problem solving, reasoning, and categorization as well (Gimmig, Huguet, Caverni, & Cury, 2006; Markman, Maddox, & Worthy, 2006).

Despite evidence that the WM available for executing the task at hand is compromised under pressure, the mechanism by which this co-option occurs has not been substantiated. Specifically, although it has been suggested that distracting thoughts and worries are at the root of this WM disruption, to our knowledge there is no direct evidence that this is the case. In the current experiment, we work towards providing such evidence. We also design an intervention to prevent these distracting thoughts and worries, if they do occur under pressure, from impinging on performance.

To carry out our objectives, we drew on Baddeley's (1986, 2003) multicomponent model of WM. According to this framework, a domain-general central executive controls and coordinates the information active in WM. Some of this information is represented and maintained in domain-specific short-term stores, such as the phonological loop for acoustic/verbal information and the visual–spatial sketchpad for visual images. A fourth component, a multimodal episodic buffer, serves to bind information from the phonological loop, the visual–spatial sketchpad, and long-term memory into a unitary episodic representation (Baddeley, 2000).

We reasoned that if pressure induces distracting thoughts and worries, these would most heavily tax the phonological aspect of working memory thought to support inner speech and thinking in the service of complex cognitive activities (Hayes, Hirsch, & Mathews, 2008; Miyake & Shah, 1999; Rapee, 1993). If so, then mathematics problems that depend relatively equally on central executive resources but differ in their reliance on verbal WM might show differential performance patterns under pressure. Specifically, signs of "choking under pressure" should occur most strongly for problems that rely on verbal WM resources in addition to general executive components.

To examine this idea, we borrowed a paradigm used by Trbovich and LeFevre (2003), in which they established that the orientation of a mathematics problem impacts the type of WM resources recruited to solve the problem. Trbovich and LeFevre demonstrated that, although all arithmetic problems involve central executive resources, problems oriented horizontally (see Figure 1a) require verbally maintaining intermediate steps in memory (e.g., reminding oneself to "carry the 1" in a multiple-digit addition problem). On the other hand, problems oriented vertically rely more on visuospatial resources, as individuals tend to solve such problems in a spatial mental workspace similar to how they would be solved on paper (e.g., visualizing themselves "carrying the 1" as if it were written on top of a column of numbers to

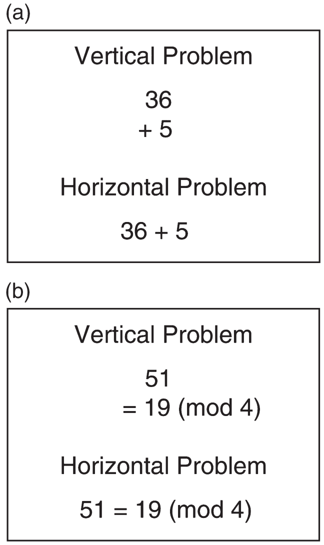

Figure 1. *(a) Example of vertically and horizontally oriented arithmetic problems. (b) Example of vertically and horizontally oriented modular arithmetic problems.*

be added). Specifically, Trbovich and LeFevre found that performance was worse when horizontally oriented addition problems were performed concurrently with a verbal secondary task (maintaining a series of nonwords in memory) than when they were performed with a visuospatial secondary task (remembering a visually presented star pattern). The opposite pattern was found for vertically oriented problems.

In the current study, we drew upon the finding that a problem's orientation can lead to the differential recruitment of verbal versus visuospatial WM resources to examine how performance pressure impacts WM. We varied the orientation of mathematical problems commonly used in the pressure literature—Gauss's (1801) modular arithmetic (Beilock et al., 2004)—and explored how individuals perform these problems under either a low-pressure or a high-pressure test.

The object of modular arithmetic is to calculate whether statements such as "51 = 19 (mod 4)" are true or false. To do this, the problem's second number is subtracted from the first number (i.e., 51 − 19), and this difference is divided by the last number (i.e., 32 ÷ 4). If the dividend is a whole number (here, 8), then the statement is true. Because previous research has demonstrated that the primary WM demand in this task occurs within the subtraction procedure (Beilock et al., 2004), horizontal and vertical orientation were altered in this portion of the modular arithmetic problem (Figure 1b). Modular arithmetic involves common arithmetic procedures combined in a relatively novel way. This enabled us to examine performance decrements on the types of subtraction and division problems commonly encountered on real-world mathematics tests while simultaneously controlling individuals' learning history.

If, in high-pressure situations, verbal WM resources are devoted to thoughts and worries, then performance on problems that depend more heavily on these resources (e.g., horizontally oriented problems) should be more negatively impacted than problems that do not (e.g., vertically oriented problems). Such findings would be in line with previous work in our lab, in which we demonstrated that women who were reminded about gender differences in mathematical ability (a stereotype threat manipulation thought to create worries about performance; Cadinu, Maass, Rosabianca, & Kiesner, 2005) showed a decrement on horizontal but not on vertical problems, whereas women who were not given this information performed equally well on both problem types (Beilock, Rydell, & McConnell, 2007).

Beyond linking verbal WM consumption to poor performance under pressure, in the current work we also investigated the direct relation between reported worries and performance. After completing mathematics problems under a low- or high-pressure testing condition, we asked participants to write down everything they remembered thinking about during the mathematics task. To the extent that distracting thoughts and worries co-opt verbal WM resources, these reported worries should predict performance on horizontal (i.e., verbally based), but not vertical problems. In other words, the more an individual reports worrying, the worse his or her verbal-based mathematics performance should be.

Finally, drawing on the idea that verbal WM consumption is directly linked to poor mathematics performance under pressure (at least on problems that rely heavily on verbal WM), we designed an intervention to help alleviate pressure's negative effects. Specifically, we asked some participants to say the steps of the problems out loud while solving. Talking aloud was intended to direct phonological resources to the steps of the problem, keeping distracting thoughts at bay. We predicted that if verbal WM stores critical to the mathematics task are co-opted by distracting thoughts, then helping individuals keep their verbal resources focused on the task may serve to counteract pressure-induced performance decrements. Thus, individuals who talk themselves through the problem steps should not show poor mathematics performance under pressure.

Method

Participants

Participants ($N = 78$) were college-level students ranging in age from 18 to 22 years ($M = 18.96$, $SD = 0.92$) who had taken no more than two university mathematics courses and reported no previous exposure to modular arithmetic. Participants were assigned to one of four cells in a 2 (*pressure test*: low pressure, high pressure) × 2 (*talk-aloud condition*: no talk aloud; talk aloud) design.

Modular arithmetic task

Participants were tested one at a time. After giving informed consent, individuals were introduced to the modular arithmetic task. Specifically, they were told they would see problems on the computer screen (see Figure 1b) and that their goal was to determine whether each problem was true or false. To do this, they were instructed to mentally subtract the first two numbers and then divide that result by the "mod" number in parentheses. If the division yielded a whole number, the problem was said to be true. If not, it was false.

Participants were instructed to complete the problems as quickly and accurately as possible and to indicate their answer by pressing the "T" or "F" keys on the computer keyboard (corresponding to "true" or "false," respectively). Each trial began with a 500-ms fixation point in the centre of the screen, which was immediately replaced by a problem present until response. After response, feedback was given (i.e., the phrase "Your response to the problem was CORRECT!" or "Your response to the problem was INCORRECT!" appeared for 1,500 ms). Then the screen went blank for a 1,500-ms intertrial interval.

Participants performed several initial trials to familiarize themselves with the task. Then, after the experimenter clarified any questions, all participants performed a practice block of 32 problems. Half of these problems were oriented horizontally and half vertically (Figure 1b). Within these orientations, half of the problems were low in subtraction demand, and half were high. *Low-demand* problems were composed of single digit numbers without a borrow operation—for example, $8 = 3$ (mod 3). *High-demand* problems consisted of double-digit numbers with a borrow operation during subtraction—for example, $33 = 15$ (mod 3). Larger numbers and borrow operations involve longer sequences of steps and require maintaining more intermediate products, which place greater demands on WM (Ashcraft, 1992). If pressure exerts its impact by co-opting the WM available for performance, then the performance of those problems higher in subtraction demand should be most susceptible to poor performance under pressure.

Following practice, participants in the talk-aloud and no-talk-aloud groups were given instructions for their particular condition (see below). Then participants took the low-pressure or high-pressure test. Each test consisted of 32 problems, of which half were presented in a vertical orientation and half in a horizontal orientation. In addition, half of the problems in each orientation were lower in subtraction demand and half higher.

Pressure tests

Low-pressure test participants were told that they should work through the problems as quickly and accurately as they could. The experimenter then left the room. High-pressure test participants were given a pressure scenario intended to elicit commonly experienced pressures, such as monetary reward, peer pressure, and social evaluation. Specifically, participants were told that if both they and a "partner" could improve their reaction time and accuracy by 20%, relative to the practice block, then each could earn $5. They were also told that their partner had already improved by the required amount, leaving the present participant to earn the money for both parties. Participants were also videotaped by the experimenter and were told that the footage would be watched by students, professors, and maths teachers to examine how people perform this new type of maths task. The video camera was set up directly to the side of participants, such that the camera could record both the participant and the computer screen. Following the test block, the experimenter turned the camera off and pointed it away from the participant.

Each modular arithmetic problem was presented once across the entire experiment. In addition, the problems within each orientation were counterbalanced across participants (i.e., horizontal problems given to one participant were presented as vertical problems to another participant). Finally, half of the modular arithmetic problems presented within each subtraction demand level were true, and half were false. Each true problem had a false correlate that differed as a function of the number involved in the mod statement. For example, if the "true" problem "$51 = 19 \pmod 4$" was given, then a "false" correlate problem such as "$51 = 19 \pmod 3$" was also presented at some point within the same problem block. This pairing was designed to equate the numbers within the true and false problems as much as possible.

Talk-aloud conditions

No-talk-aloud condition. Following the practice block, participants in the no-talk-aloud condition were informed that they would be performing another set of problems and were asked to try to work as quickly and accurately as they could.

Talk-aloud condition. Participants in the talk-aloud condition were told that we were interested in what they say to themselves as they perform the problems, so they should talk aloud as they work on the problems. They were told to specifically focus on the steps they took to solve the problems, talking themselves through these steps out loud, into a microphone set up in front of them. Participants were told that what they said would be recorded, but that their identity would remain anonymous. Participants were also told that we were interested in how individuals solve these types of problems and that there are no right or wrong things to say (Ericsson & Simon, 1984). Following the instructions, the experimenter clarified any questions, started the recording device, and stood in the back of the room, while participants completed several problems where they practised talking aloud. During this practice, if a participant stopped talking for an extended period of time, he or she was reminded to keep talking out loud (Ericsson & Simon, 1984).

At the end of the experiment, participants in the talk-aloud condition completed two items designed to ensure that they complied with the talk-aloud instructions. Participants reported (a) whether they followed the instructions to talk out loud as best they could and (b) whether they forgot to talk out loud during the experiment. Both reports were made on scales ranging from 1 (strongly disagree) to 9 (strongly agree). The second item was reverse coded, and these items were averaged. Only individuals who reported compliance with the talk-aloud instructions at the midpoint (i.e., 5) or above on this measure were included in the experiment, as one cannot ascertain the potential benefit of talking aloud if an individual did not do so.

Questionnaires

Immediately following the mathematics task, individuals completed several questionnaires, beginning with a retrospective verbal report intended to elicit the thoughts they had during the last block of problems. This questionnaire stated: "We all have several thoughts that run through our mind at any given time. Please describe everything that you remember thinking about as you performed the last set of modular arithmetic problems." Then participants reported how much pressure they felt during the last block of problems, ranging from 1 (very little performance pressure) to 7 (extreme performance pressure). Next, individuals completed a demographic sheet detailing previous mathematics experience. Finally, participants were fully debriefed, and those in the high-pressure condition were given the monetary reward regardless of performance.

Results

Perceptions of pressure

We began by examining reports of felt performance pressure in a 2 (*pressure test*: low, high) × 2 (*talk-aloud condition*: no talk aloud, talk aloud) analysis of variance (ANOVA). Only a main effect of pressure test was significant, $F(1, 74) = 6.70$, $p < .05$, $\eta_p^2 = .08$. Participants given the high-pressure test ($M = 4.81$, $SE = 0.21$) reported feeling more performance pressure than participants given the low-pressure test ($M = 4.06$, $SE = 0.20$). Our manipulation increased participants' perceptions of pressure, and this did not differ as a function of talk-aloud condition.

Modular arithmetic performance

We next examined mathematical problem solving accuracy in the low- and high-pressure tests for participants in the no-talk-aloud and talk-aloud conditions. In order to ensure that all participants demonstrated reasonable performance on the modular arithmetic task prior to the introduction of our experimental manipulations, only participants whose problem-solving accuracy in the practice block was greater than chance (i.e., 50% correct) were included in the experiment.

A 2 (*pressure test*: low, high) × 2 (*talk-aloud condition*: no talk aloud, talk aloud) × 2 (*problem subtraction demand*: low, high) × 2 (*problem orientation*: horizontal, vertical) ANOVA on modular arithmetic accuracy revealed a main effect of problem subtraction demand, $F(1, 74) = 68.02$, $p < .001$, $\eta_p^2 = .48$, Pressure Test × Problem Orientation interaction, $F(1, 74) = 4.57$, $p < .04$, $\eta_p^2 = .06$, and a significant four-way interaction, $F(1, 74) = 9.68$, $p < .01$, $\eta_p^2 = .12$. To interpret this interaction, we examined accuracy separately for low-demand and high-demand problems.

In terms of low-demand problem accuracy, a Pressure Test × Talk-Aloud Condition × Problem Orientation ANOVA revealed no significant main effects or interactions. In contrast, for high-demand problems a Pressure Test × Problem Orientation interaction, $F(1, 74) = 5.28$, $p < .03$, $\eta_p^2 = .07$, and a significant three-way interaction obtained, $F(1, 74) = 6.00$, $p < .02$, $\eta_p^2 = .08$.

As shown in Figure 2, for high-demand problems, vertically oriented problem accuracy did not differ as a function of pressure or talk-aloud conditions, Fs < 1. In contrast, horizontally oriented problem accuracy did differ, as revealed by main effects of pressure test, $F(1, 74) = 6.02$,

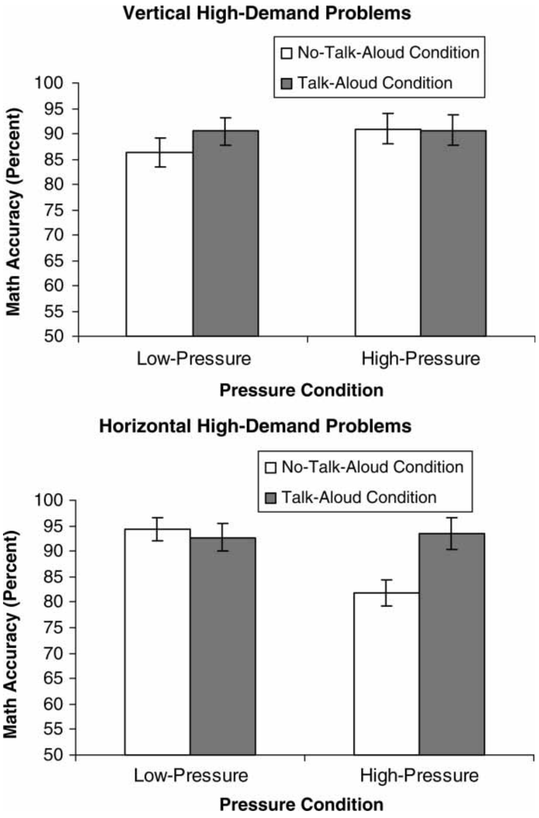

Figure 2. *Accuracy of vertical and horizontal high-demand problems as a function of pressure and talk-aloud condition.*

$p < .02$, $\eta_p^2 = .08$, talk-aloud condition, $F(1, 74) = 4.27$, $p < .05$, $\eta_p^2 = .06$, and a Pressure Test × Talk-Aloud Condition interaction, $F(1, 74) = 7.57$, $p < .01$, $\eta_p^2 = .09$. When participants did not talk aloud, horizontally oriented problems were performed significantly less accurately in the high-pressure than in the low-pressure test, $t(37) = 3.28$, $p < .01$. In contrast, participants who talked aloud under pressure were inoculated against pressure's negative effects. These individuals performed horizontal problems no differently from no-talk-aloud and talk-aloud participants in the low-pressure test, $t(35) = 0.28$, *ns*, and $t(37) = -0.25$, *ns*, respectively, and significantly better than no-talk-aloud participants in the high-pressure test, $t(34) = -2.91$, $p < .01$.

We next examined problem-solving reaction time (RT). A 2 (pressure test) × 2 (talk-aloud condition) × 2 (problem subtraction demand) × 2 (problem orientation) mixed ANOVA with log-transformed RTs revealed main effects of problem subtraction demand, pressure condition, and talk-aloud condition. Low-demand problems ($M = 3.33$, $SE = 0.02$) were solved more quickly than high-demand problems ($M = 3.85$, $SE = 0.02$), $F(1, 74) = 1{,}972.72$, $p < .001$, $\eta_p^2 = .96$. Participants given high pressure ($M = 3.55$, $SE = 0.02$) solved the problems more quickly than those given low pressure ($M = 3.63$, $SE = 0.02$), $F(1, 74) = 6.66$, $p < .02$, $\eta_p^2 = .08$. Finally, participants who did not talk aloud ($M = 3.52$, $SE = 0.02$) solved the problems more quickly than those who talked aloud ($M = 3.66$, $SE = 0.02$), $F(1, 74) = 25.80$, $p < .001$, $\eta_p^2 = .26$. No interactions obtained, suggesting that these RT results do not qualify the interactions seen above in accuracy.

It should be noted that, because talking aloud increased RT across all conditions, any benefit of talking aloud might come at a cost of slower problem-solving time. One might wonder whether the time cost associated with talking aloud reduces the practical benefit of this intervention. However, as shown above, individuals solved problems more quickly in high-pressure than in low-pressure testing conditions. Thus, even though talking aloud slowed performance overall, participants who talked aloud during a high-pressure test were not solving significantly more slowly than individuals given a low-pressure test (see Table 1). Moreover, talking aloud under pressure had the added benefit of improving accuracy.

Because of the slower RT in talk-aloud conditions, the benefit of directing the focus of attention towards problem steps is not decoupled from a potential benefit of taking more time to solve the problems in the talk-aloud condition. However, the lack of any main effects or interactions with problem orientation indicates that our key findings cannot be solely attributed to a speed–accuracy trade-off—any changes in RT due to pressure or talk-aloud conditions occurred uniformly across horizontal and vertical problems (see Table 1). This is in contrast to the increase in accuracy in the talk-aloud condition, which was localized to the problems that depend most heavily on verbal WM (i.e., high-demand horizontal problems).

To further demonstrate that our key accuracy findings cannot simply be attributed to speed–

Table 1. *High-demand problem reaction time*

		Pressure condition		
Problem	*Talk-aloud condition*	*Low pressure*	*High pressure*	Mean
Vertical	No talk aloud	3.79 (0.03)	3.75 (0.03)	*3.77* (0.02)
	Talk aloud	3.96 (0.03)	3.85 (0.04)	*3.91* (0.02)
	Mean	*3.88* (0.02)	*3.80* (0.02)	
Horizontal	No talk aloud	3.84 (0.03)	3.73 (0.03)	*3.78* (0.02)
	Talk aloud	3.96 (0.03)	3.88 (0.04)	*3.92* (0.02)
	Mean	*3.90* (0.02)	*3.80* (0.03)	

Note: Mean reaction times (log-transformed). Standard errors in parentheses.

accuracy trade-offs, we next examined high-demand problem accuracy while controlling for variation in RT. A 2 (pressure test) × 2 (talk-aloud condition) analysis of covariance (ANCOVA) for high-demand horizontal problems, using high-demand horizontal problem RT as a covariate, revealed main effects of pressure test, $F(1, 73) = 8.42$, $p < .001$, $\eta_p^2 = .10$, and talk-aloud condition, $F(1, 73) = 6.96$, $p = .01$, $\eta_p^2 = .09$, and a Pressure Test × Talk-Aloud Condition interaction, $F(1, 73) = 8.13$, $p < .01$, $\eta_p^2 = .10$. In contrast, the same analysis with vertical problem accuracy and RT showed no significant effects. Controlling for RT does not change the pattern or significance of the accuracy data reported above.

Retrospective verbal reports

Responses on the retrospective verbal report questionnaire were divided into five categories:

1. Worries/negative thoughts and thoughts regarding monitoring performance (e.g., “I kept thinking about how much I hate math”, “I felt nervous because I didn’t want to let my partner down by not improving”, “When I tried to hurry, sometimes I made mistakes”).
2. Thoughts related to performing the task itself (e.g., “Subtracting then dividing”, “I would first subtract the tens digits then the ones digits”).
3. General thoughts related to the situation or task (e.g., “I was thinking, I can’t remember the last time I had to do subtraction and division in my head”, “What I’d do with the $5”).
4. General distress/tension (e.g., “Is this over yet?”).
5. Thoughts unrelated to the experiment (e.g., “What am I doing this weekend?”).

Two experimenters unaware of the experimental conditions independently coded the thought data. Interjudge agreement was very high ($\alpha = .91$), so one judge’s coding was used for all responses. Individuals who did not fill out the questionnaire were excluded from all analyses.

As can be seen in Table 2, across the pressure tests and talk-aloud conditions, individuals reported similar total numbers of thoughts ($M = 3.72$, $SE = 0.22$), $F\text{s} < 1$. In terms of specific thought-categories, a Pressure Test × Talk-Aloud Condition multivariate analysis of variance (MANOVA) on the number of thoughts in each category revealed that (a) participants in the high-pressure test reported more thoughts involving worrying and their performance than did participants who took the low-pressure test, $F(1, 74) = 6.08, p < .02$, $\eta_p^2 = .08$. This suggests that worries/negative thoughts and performance monitoring increase under pressure. Conversely, (b) high-pressure participants reported fewer thoughts related to performing the task than did low-pressure participants, $F(1, 74) = 17.47$, $p < .001$, $\eta_p^2 = .19$. Third, (c) high-pressure participants reported more general thoughts about the performance situation than did low-pressure participants, $F(1, 74) = 3.97$, $p = .05$, $\eta_p^2 = .05$, perhaps because this category included general thoughts about the pressure situation itself (e.g., “What I’d do with the $5”). In addition, (d) no-talk-aloud participants reported more thoughts related to general distress than did talk-aloud participants, $F(1, 74) = 4.26$, $p < .05$, $\eta_p^2 = .05$. Finally, (e) no-talk-aloud participants also reported a greater number of thoughts that were unrelated to the performance situation than did talk-aloud participants, $F(1, 74) = 5.62, p < .02$, $\eta_p^2 = .07$. These last two effects may be because talking through the problem steps prevented individuals’ attention from wandering off-task overall. No other main effects or Pressure Test × Talk-Aloud Condition interactions were found.

In addition to analysing the frequency of reported thoughts by condition, we also looked at how these reported thoughts related to mathematics performance. To do this, we correlated the frequency of each type of reported thought with accuracy on the horizontal and vertical high-demand problems, collapsing across pressure tests and conditions. If particular thoughts (e.g., worries) compromise verbal WM resources needed for the task at hand, problems heavily reliant on these resources should be impacted regardless of what situation one finds oneself in.

Worries/negative thoughts and performance monitoring were significantly negatively related

Table 2. *Number of thoughts reported in each of the five categories*

		Pressure condition		
Thought type	*Talk-aloud condition*	*Low pressure*	*High pressure*	*Mean*
1. Worries/negative thoughts & monitoring performance	No talk aloud	0.85 (.34)	1.74 (.34)	*1.29* (.24)
	Talk aloud	1.09 (.32)	1.88 (.36)	*1.49* (.24)
	Mean	*0.97* (0.23)	*1.81* (0.25)	
2. Performing the task	No talk aloud	1.55 (0.23)	0.90 (0.23)	*1.22* (0.16)
	Talk aloud	2.05 (0.22)	0.77 (0.25)	*1.41* (0.16)
	Mean	*1.80* (0.16)	*0.83* (0.17)	
3. General thoughts related to the situation or task	No talk aloud	0.40 (0.22)	0.84 (0.23)	*0.62* (0.16)
	Talk aloud	0.32 (0.21)	0.77 (0.24)	*0.54* (0.16)
	Mean	*0.36* (0.15)	*0.80* (0.16)	
4. General distress/tension	No talk aloud	0.65 (0.15)	0.26 (0.15)	*0.46* (0.11)
	Talk aloud	0.23 (0.14)	0.06 (0.16)	*0.14* (0.11)
	Mean	*0.44* (0.10)	*0.16* (0.11)	
5. Thoughts unrelated to the experiment	No talk aloud	0.50 (0.14)	0.16 (0.14)	*0.33* (0.10)
	Talk aloud	0.00 (0.13)	0.00 (0.15)	*0.00* (0.10)
	Mean	*0.25* (0.09)	*0.08* (0.10)	

Note: Standard errors in parentheses.

to high-demand horizontal problem accuracy ($r = -.37$, $p = .001$) but not to high-demand vertical problem accuracy ($r = .05$, *ns*; see Table 3).[1] These correlation coefficients were significantly different from each other ($z = -2.57$, $p < .05$; Meng, Rosenthal, & Rubin, 1992). The more individuals reported worrying, the worse their accuracy—but only for the problems that rely most heavily on verbal WM resources. Interestingly, reported worries were also negatively related to thoughts focused on performing the task ($r = -.26$, $p < .03$): A lower number of worries corresponded with more task-related thoughts. And, the more task-related thoughts people had, the better their high-demand horizontal accuracy ($r = .24$, $p < .04$). This relationship was not found for vertical problems ($r = .07$, *ns*).

In sum, the more worries and negative thoughts that individuals reported, the worse they performed on the horizontally oriented problems thought to require extensive verbal WM resources. Additionally, the more one reported verbalizing the steps of performance, the better one performed these problems. No relationship was found between reported thoughts and vertical problem accuracy, probably because vertical problems do not require verbal WM resources in the same way as do horizontally oriented problems (Beilock et al., 2007; Trbovich & LeFevre, 2003).

It is worth noting that high-pressure participants who talked aloud did not actually report that they worried less than high-pressure participants who did not talk aloud, as indicated by a lack of a Pressure Test × Talk Aloud interaction when looking at number of thoughts in the first questionnaire category. Why might this be the case? It seems likely that asking individuals to report thought content the way we did captured the most salient types of thoughts that individuals had during performance, but did not fully capture the intensity of these thoughts (Beilock et al., 2007). For example, an individual could have the same worry repeatedly across the problem block (e.g., thinking "I really hate mathematics" several times during the course of problem

---

[1] The same correlation analyses with RT revealed no significant effects.

Table 3. *Correlations between horizontal and vertical high-demand problem accuracy and number of each thought type reported on the retrospective thought questionnaire*

	Problem accuracy	
Thought type	*Horizontal*	*Vertical*
Worries/negative thoughts and monitoring performance	−.37*	.05
Performing task	.24*	.07
General situation-related thoughts	.03	.05
General distress/tension	.02	−.01
Unrelated to the experiment	.11	.00

*$p < .05$.

solving) but only report this as one negative thought on our questionnaire (e.g., "I kept thinking about how much I hate mathematics"). If high-pressure participants who talked aloud had negative thoughts less frequently, because they verbalized the problem steps, then this could explain their lack of "choking under pressure". Unfortunately, our report method may not have been able to capture these nuances (see Cadinu et al., 2005, for similar reporting issues when a spontaneous thought report method was used). Nonetheless, the fact that we were able to document increased worrying under pressure using retrospective reports and a relation between this worrying and horizontally oriented problem performance, and that participants who talked aloud performed better under pressure than those that did not, lends overall support to our conclusions that high-pressure situations harm test performance by creating negative thoughts and worries. Getting individuals to focus on the problem steps helps thwart these negative performance effects.

Discussion

In the current work, we provide evidence that verbal WM resources are compromised in consequential testing situations. Performance pressure disrupted accuracy on problems that relied most heavily on verbal WM resources (i.e., WM-demanding horizontally oriented problems but not WM-demanding vertically oriented problems). When we asked our participants to talk through the problem steps out loud, this poor horizontal problem accuracy under pressure was eliminated. In other words, asking participants to explicitly direct their verbal WM resources towards the steps required to solve the problems helped them excel above those who did not talk aloud during the pressure test. Moreover, across all participants, as the number of reported worries increased, performance on the WM-demanding horizontal problems decreased. Together, these results not only provide evidence for the negative impact of distracting thoughts under pressure but also demonstrate how such knowledge can be used to develop methods to eliminate pressure's negative effects.

Our work may be especially relevant to individuals higher in WM capacity, as past research has shown that the performance of these individuals is, ironically, more impacted by pressure than the performance of their lower WM counterparts (Beilock & Carr, 2005; Beilock & DeCaro, 2007). The current findings align with the suggestion that situation-induced worries compete for the WM resources that individuals with higher WM capacity normally rely on for their superior performance (Beilock, 2008). Thus, it may be that a talk-aloud intervention would benefit higher WM individuals the most.

The current findings also demonstrate that certain performance situations increase the presence of specific types of thoughts. For example, our high-pressure test increased the presence of worries/negative thoughts compared to our low-pressure test. An increase in worries/negative thoughts directly corresponded to a decline in verbal-based mathematical performance. There may, however, be other effects on the WM system in situations that increase other types of thoughts. For example, as opposed to worries (which are thought to result in verbal ruminations), more general anxious arousal (thought to draw attention to physiological symptoms of anxiety) has been shown to impact visuospatial WM (Shackman et al., 2006; see Eysenck & Calvo, 1992; Ikeda, Iwanaga, & Seiwa, 1996). In addition, thoughts unrelated to the experiment at all (e.g., "What am I doing this weekend?) have been shown to tax the central executive component of WM (Teasdale et al., 1995). Although in the current work neither general distress nor unrelated thoughts were related to performance, there was a relatively low base-rate of these types of thoughts. Different types of performance situations may vary in how they increase the salience of particular thoughts or emotions, impacting the WM system in multiple ways.

Much like the high-pressure testing situation in the current work, other performance situations have also been shown to increase worries and negative thoughts about performance, For example, in stereotype threat situations, individuals faced with a negative stereotype about their social group begin to worry about conforming to this stereotype, hampering performance on WM-demanding skills like mathematics (Beilock et al., 2007; Cadinu et al., 2005; Schmader & Johns, 2003). Similarly, test-anxious individuals tend to worry more in testing situations, compromising verbal WM resources (Ikeda et al., 1996). By documenting the verbal nature of WM disruption in the current study, we demonstrate a similarity between the performance pressure literature and these other types of pressure-inducing situations (cf., Beilock et al., 2007), while also offering new insight into ways to ameliorate the negative impact of these stressful situations.

Original manuscript received 21 August 2009
Accepted revision received 17 October 2009
First published online 5 February 2010

REFERENCES

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44,* 75–106.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4,* 417–423.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews*, *4,* 829–839.

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science*, *17,* 339–343.

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, *130,* 701–725.

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science*, *16,* 101–105.

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33,* 983–998.

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.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spill over. *Journal of Experimental Psychology: General*, *136,* 256–276.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science*, *16,* 572–578.

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16,* 353–386.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis*. Cambridge, MA: MIT Press.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The Processing Efficiency Theory. *Cognition and Emotion*, *6,* 409–434.

Gimmig, D., Huguet, P., Caverni, J. P., & Cury, F. (2006). Choking under pressure and working memory capacity: When performance pressure reduces fluid intelligence (Gf). *Psychonomic Bulletin & Review*, *13,* 1005–1010.

Hayes, S., Hirsch, C., & Mathews, A. (2008). Restriction of working memory capacity during worry. *Journal of Abnormal Psychology*, *117,* 712–717.

Ikeda, M., Iwanaga, M., & Seiwa, H. (1996). Test anxiety and working memory system. *Perceptual and Motor Skills*, *82,* 1223–1231.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95,* 492–527.

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science*, *17,* 944–948.

Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111,* 172–175.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: University Press.

Rapee, R. M. (1993). The utilization of working memory by worry. *Behavior Research and Therapy*, *31,* 617–620.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85,* 440–452.

Shackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D. A., Lavric, A., & Davidson, R. J. (2006). Anxiety selectively disrupts visuospatial working memory. *Emotion*, *6,* 40–61.

Teasdale, J. D., Dritschel, B. H., Taylor, M. J., Proctor, L., Lloyd, C. A., Nimmo-Smith, I., et al. (1995). Stimulus-independent thought depends on central executive resources. *Memory & Cognition*, *23,* 551–559.

Trbovich, P. L., & LeFevre, J. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition*, *31,* 738–745.