

# From Poor Performance to Success Under Stress: Working Memory, Strategy Selection, and Mathematical Problem Solving Under Pressure

Sian L. Beilock  
The University of Chicago

Marci S. DeCaro  
Miami University

Two experiments demonstrate how individual differences in working memory (WM) impact the strategies used to solve complex math problems and how consequential testing situations alter strategy use. In Experiment 1, individuals performed multistep math problems under low- or high-pressure conditions and reported their problem-solving strategies. Under low-pressure conditions, the higher individuals' WM, the more likely they were to use computationally demanding algorithms (vs. simpler shortcuts) to solve the problems, and the more accurate their math performance. Under high-pressure conditions, higher WM individuals used simpler (and less efficacious) problem-solving strategies, and their performance accuracy suffered. Experiment 2 turned the tables by using a math task for which a simpler strategy was optimal (produced accurate performance in few problem steps). Now, under low-pressure conditions, the lower individuals' WM, the better their performance (the more likely they relied on a simple, but accurate, problem strategy). And, under pressure, higher WM individuals performed optimally by using the simpler strategies lower WM individuals employed. WM availability influences how individuals approach math problems, with the nature of the task performed and the performance environment dictating skill success or failure.

*Keywords:* working memory, math problem solving, strategy, pressure, stress

What determines successful performance on complex problem-solving tasks ranging from mathematical computations to analogical mappings? This question has been probed by individual difference researchers interested in understanding the role that working memory (WM) plays in skill execution (Conway et al., 2005; Engle, 2002), by social and cognitive psychologists examining the impact of negative or stressful environments on performance outcomes (Ashcraft & Kirk, 2001; Beilock, Kulp, Holt, & Carr, 2004; Schmader & Johns, 2003), and by developmental psychologists examining math problem-solving performance as a function of computational skill and arithmetic knowledge (Siegler, 1988a, 1988b; Siegler & Lemaire, 1997). Although these research areas have yielded important conclusions regarding skill learning and performance, these lines of work have—surprisingly—operated largely in isolation. For example, although there is a significant amount of research devoted to understanding the cognitive construct of WM and how it relates to higher level functions, such as general intellectual ability, less is known about how individual differences in WM shape the strategies individuals call upon to solve particular problems—especially when there are multiple routes to problem solution (Price, Catrambone, & Engle, 2007; although see Reder & Schunn, 1999; Schunn & Reder,

2001). Moreover, even less work has focused on the interplay of this type of cognitive control and environmentally induced stressors (Miyake & Shah, 1999).

In the current work, we identify key differences in the strategies individuals lower and higher in WM use to solve complex mathematical problems, and we explore how these strategies are impacted by pressure-filled performance situations. Not only does this work contribute to our understanding of how high-stress situations impact performance, but also it speaks to how individual differences in cognitive control shape one's approach to problem solving more generally. Such work is important for researchers interested in developing models of executive functioning that capture the complexity of real-world performance. Moreover, this work is imperative to the development of training regimens and performance strategies designed to maximize skill success and minimize failure—especially in those situations in which incentives for optimal performance are at their highest.

## Why WM?

Successful performance on tasks ranging from mathematical problem solving to sentence comprehension is thought to rely heavily on WM (Conway et al., 2005; DeStefano & LeFevre, 2004; Stevenson & Carlson, 2003). If the capacity of the WM system to oversee task-relevant information is disrupted, performance may suffer (Beilock & Carr, 2005; Trbovich & LeFevre, 2003). Because of the central role that WM plays in complex cognition, it is perhaps not surprising that diverse research areas ranging from social psychology (Schmader & Johns, 2003) to cognitive neuroscience (Kane & Engle, 2002) have emphasized the positive role of WM and attentional control in high-level performance. In fact, it has been suggested that “working memory is so

---

Sian L. Beilock, Department of Psychology, The University of Chicago; Marci S. DeCaro, Department of Psychology, Miami University.

This research was supported by Institute of Education Sciences Grant R305H050004 and National Science Foundation Grant BCS-0601148 to Sian L. Beilock.

Correspondence concerning this article should be addressed to Sian L. Beilock, Department of Psychology, 5848 South University Avenue, The University of Chicago, Chicago, IL 60637. E-mail: beilock@uchicago.edu

central to human cognition that it is hard to find activities where it is not involved" (Ericsson & Delaney, 1999, p. 259).

WM capacity can be thought of as a short-term memory system involved in the control, regulation, and active maintenance of a limited amount of information with immediate relevance to the task at hand (Miyake & Shah, 1999). WM is also an individual difference variable, meaning that some people have more of this construct and some people have less. Despite knowledge of how various aspects of WM (e.g., attentional control, inhibition) contribute to skilled performance (Engle, 2002; Kane & Engle, 2000), less is known about how individual differences in WM affect the problem-solving strategies individuals use to solve complex, multistep problems such as those found in domains like math (Price et al., 2007).

### Dual Processes in Problem Solving

There is a literature that we can look to, however, for insights concerning how WM might influence one's approach to math problem solving. Broadly speaking, dual-process theories propose that two distinct processes—associative and rule-based processes—support performance in reasoning and decision-making tasks such as the evaluation of persuasive arguments and judgments of logic (Evans, 2003; Sloman, 1996; Smith & DeCoster, 2000; Stanovich & West, 2000). *Associative processes* consist of similarity-based associations built up over repeated exposure to concurrent events. These processes are believed to operate relatively spontaneously and make few demands on WM resources (Logan, 1988; Petty & Cacioppo, 1986; Rydell, McConnell, Mackie, & Strain, 2006). *Rule-based processes*, on the other hand, rely on symbolically represented explicit knowledge as conventions to guide processing. The use of explicit rules to manipulate problems and derive solutions is believed to place heavy demands on WM (Carlson, 1997; Stevenson & Carlson, 2003).

Imagine that an individual is presented with the math-based question, *Does the answer to the problem  $(32 - 8) \div 4$  have a remainder?* Working through the step-by-step computations to derive an answer (i.e., subtracting 8 from 32 and then dividing this answer by 4 so as to conclude that the answer to the question is *no*) relies heavily on rule-based processing. Here, one is using explicit computations in a logically derived sequence to arrive at the problem solution. If one instead perceives that all of the numbers in the above equation are even, and although one has never seen this specific problem before, one has in many past math problem-solving experiences observed that remainders do not often result when even numbers are subtracted and divided, then one could conclude that the answer is likely to be *no* through more associative processing. Here, individuals are using a relatively spontaneous association between even numbers and a lack of remainders to respond. This is a far simpler strategy that is not reliant on the same types of rule-based processes as working through the step-by-step computations.

Associative and rule-based processes often converge on the same conclusion—as in the above math problem. However, this is not always the case. For example, if one were asked whether the answer to the problem  $(32 - 6) \div 4$  had a remainder, performing the step-by-step problem-solving computations would result in a *yes* response, whereas the use of the association between the division of even numbers and the low likelihood of a remainder

would result in a *no* response. Such conflicting situations are of special interest in the current work because they provide opportunities for understanding differences in the problem-solving strategies individuals use in particular situations.

### Individual Differences in Problem-Solving Processes

To the extent that rule-based (but not associative) processes depend heavily on WM, one might imagine that individual differences in WM would influence the process that is most readily and effectively utilized (De Neys, 2006; Evans, 2003; Stanovich & West, 2000). Specifically, individuals who come to the table with less WM capacity to begin with might show a tendency to rely less on rule-based processes and more on associative processes, precisely because these lower WM individuals have less of the cognitive capacity necessary to support rule-based computations. In contrast, higher WM individuals might rely more so on rule-based processes for the very reason that lower WM individuals do not. And, indeed, although not related to WM per se, in an examination of individual differences in subtraction and addition performance, Siegler (1988a) found that students could be classified into distinct groups on the basis of their knowledge of a particular problem (e.g., the ability to execute the correct problem-solving computation) and their retrieval threshold (i.e., the threshold set for directly retrieving a problem answer from memory versus explicitly computing an answer).

According to the distribution of associations model (Siegler, 1988b; Siegler & Shrager, 1984) and its successor, the adaptive strategy choice model (Siegler & Lemaire, 1997; Siegler & Shipley, 1995), individuals represent information about math problems (e.g.,  $13 - 3$ ) in an associative network that consists of a distribution of associations between specific problems and possible answers (Siegler, 1988b). These associations are built up over repeated prior exposure to math problems and corresponding answers. Whether a problem answer will be retrieved via its association with problem operands depends on where one sets a minimum confidence criterion (i.e., the minimum strength of association required for an answer to be spontaneously produced) and a maximum time limit before associatively based strategies are abandoned.

Siegler (1988a) suggested that one source of retrieval threshold difference may be previous experience in the use of complex, rule-based strategies to solve particular problems. Because rule-based strategy use is more resource intensive than is deriving answers associatively, individuals who are able to successfully execute such computations may set higher thresholds for retrieval, resulting in an increased use of rule-based computations. In contrast, if individuals do not have the resources necessary to accurately execute rule-based computations, they may be more likely to rely on associatively derived answers (i.e., set lower thresholds for retrieval). It is important to point out that Siegler's models do not apply solely to familiar problems or to problems one has seen in the past. Experience with related problems or component problem features should allow individuals to generalize problem-solving strategies they have relied on in the past to new problems they have never encountered (Siegler & Lemaire, 1997).

In terms of WM differences in math then, it may be that the difficulty experienced by those lower in WM in implementing complex, rule-based computations makes them more prone to rely

on associatively derived strategies that make few demands on WM and attentional control. Such a finding not only would demonstrate that individual differences in WM influence math problem-solving approaches, but also it would provide an explanation for a recently discovered phenomenon in our laboratory and others in which higher (but not lower) WM individuals are susceptible to skill failure under pressure (Beilock & Carr, 2005; Gimmig, Huguet, Caverni, & Cury, 2006).

### Performance Under Pressure

Pressure-induced performance decrements, or *choking under pressure*, has been defined as performing more poorly than expected given one's skill level in situations in which incentives for optimal performance are at a maximum (Baumeister, 1984; Beilock & Carr, 2001, 2005; Beilock et al., 2004; Lewis & Linder, 1997; Masters, 1992). In WM intensive tasks, such as mathematical problem solving, it has been demonstrated that high-pressure environments interfere with performance—and thus increase the likelihood of choking—by consuming (via worries about the situation and its consequences) the WM resources individuals need to perform at an optimal level (Beilock et al., 2004; Beilock, Rydell, & McConnell, 2007; Cadinu, Maass, Rosabianca, & Kiesner, 2005).

Recently, Beilock and Carr (2005) explored susceptibility to pressure-induced failure as a function of individual differences in WM. Individuals lower and higher in WM performed a difficult math task under both a low-pressure condition and a high-pressure testing condition (in which there were negative monetary and social consequences associated with poor performance). As one might expect, individuals higher in WM outperformed their lower capacity counterparts under low-pressure conditions. However, when placed in a high-pressure testing situation, those highest in WM were the ones most likely to fail. Higher WM individuals' performance fell to the level of the lower WM individuals when the pressure was on. Lower WM individuals' performance did not suffer under pressure.

Similar results regarding individual differences in WM and susceptibility to pressure-induced failure have been found using Raven's Standard Progressive Matrices as a test bed (Gimmig et al., 2006). In this task, individuals are presented with increasingly difficult patterns that contain one missing segment and are asked to choose which segment best completes the pattern from a number of available options. Consistent with Beilock and Carr (2005), Gimmig et al. found that the performance of higher (but not lower) WM individuals was impaired under pressure. Moreover, this performance decrement was limited to those problems that made the greatest demands on WM—exactly what one would expect if performance pressure taxes the resources higher WM individuals normally use to solve difficult problems.

This work sheds light on how variation in the performer contributes to skill success and failure. Nonetheless, it is still unclear how pressure changes the high-level performance that higher WM individuals normally exhibit or why individuals lower in WM are spared from pressure-induced decrements. We believe that considering these issues from a dual-process perspective can help provide answers to these questions. Specifically, if higher WM individuals are more likely to rely on rule-based computations (Evans, 2003; Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Stanovich

& West, 2000), then to the extent that pressure co-opts the WM resources needed to instantiate such problem-solving strategies, this may be exactly what makes these individuals susceptible to failure. In contrast, if lower WM individuals rely more heavily on associatively derived answers (i.e., simpler shortcuts that rely on previous associations between the components of a particular problem and the likelihood of a given answer and make few demands on WM and attentional control), then pressure-induced consumption of WM should not disrupt performance.

### Experiment 1

To address this issue, we turned to a math problem-solving task used in previous explorations of pressure-induced failure: Gauss's *modular arithmetic* task (see Bogomolny, 1996). This task involves judging the truth value of equations such as  $34 = 18 \pmod{4}$ . To solve such equations, the middle number is subtracted from the first number (i.e.,  $34 - 18$ ), and this difference is then divided by the last number (i.e.,  $16 \div 4$ ). If the dividend is a whole number (here, 4), the statement is *true*. Those problems with remainders are considered *false*. We chose this task because, although it is based on common subtraction and division procedures, there are associatively based shortcut strategies that can be used to derive the correct answer (some of the time) without requiring a multistep problem-solving algorithm. For example, if one concludes that problems with even numbers are likely to be true because they believe even numbers are associated less often with remainders in division, this shortcut will produce the correct answer on some trials (as in the previous example) but not in all trials (e.g.,  $52 = 16 \pmod{8}$ ). This type of shortcut strategy circumvents the need to maintain and manipulate intermediate problem steps on line in WM. However, because such shortcuts are not always appropriate, their use should result in less accurate problem-solving performance overall.

In Experiment 1, participants lower and higher in WM performed the modular arithmetic task under either low-pressure or high-pressure testing conditions. Following certain problems (unknown ahead of time to participants), individuals were told to write down their problem-solving strategies—that is, how they derived the answer to the previous problem. If rule-based processing is indeed a more accurate method for solving modular arithmetic, then the tendency to use a rule-based, multistep algorithm in contrast to a simpler shortcut should predict better math performance under low-pressure conditions. And, if individual differences in WM affect the problem-solving strategies individuals use, then those higher in WM should be more likely to rely on rule-based processing and perform at a higher level than those lower in WM—at least under low-pressure conditions. Finally, if pressure-induced consumption of WM denies individuals the resources necessary to compute demanding rule-based computations, then those individuals who rely most heavily on such processes to begin with (e.g., higher WM individuals) should be most likely to choke under pressure.

It is an open question as to how these pressure-induced performance decrements might occur. One possibility is that those higher in WM persist in the use of a complex, rule-based strategy despite the fact that the resources needed to support such computations are no longer available. As a result, mistakes are made that lead to poor performance. A second possibility is that pressure-induced

consumption of WM prompts higher WM individuals to rely on simpler associatively derived shortcut strategies (similar to those used by their lower WM counterparts). This latter point is intriguing because it suggests that how one approaches a difficult problem is not only dependent on stable individual differences but also on changes in response to environmental stressors. As a preview, this is exactly what was found.

### Method

#### Participants

Undergraduate students at a U.S. Midwestern university were randomly assigned to either a low-pressure condition ( $n = 48$ ) or a high-pressure condition ( $n = 44$ ). Ten additional participants were tested but not included in the experiment for the following reasons: (a) Four participants assigned to the high-pressure condition reported that they either did not believe the pressure manipulation or knew about it ahead of time; (b) 2 participants' scores on the two WM measures used in the current work (Reading Span [RSPAN] and Operation Span [OSPAN], see below) differed by more than 20 points, suggesting that these tasks were unable to capture consistent WM measures for these participants; and (c) 4 participants' accuracy on the practice problems (explained below) was less than 50% correct. This minimum accuracy criterion was implemented to ensure that individuals were performing above chance on the modular arithmetic task prior to the implementation of any experimental manipulations (see Beilock et al., 2007; DeCaro, Wieth, & Beilock, 2007).

WM scores were based on the average of participants' scores on two well-established measures of WM: Turner and Engle's (1989) OSPAN and a modified Daneman and Carpenter's (1980) RSPAN. The OSPAN involves solving a series of arithmetic equations while attempting to remember a list of unrelated words. Individuals are presented with one equation-word string at a time on a computer (e.g.,  $5 \times 2 - 2 = 8$ ; *DOG*) and are asked to verify aloud whether the equation is correct. Individuals then read the word aloud. At the end of the series, individuals are prompted to write down the sequence of words. The RSPAN involves reading a series of sentence-letter strings (e.g., *On warm sunny afternoons, I like to walk in the park; F*). In the RSPAN, individuals read the sentence aloud and are asked to verify whether the sentence makes sense. Individuals then read the letter aloud. At the end of the series they write down the sequence of letters. In both the OSPAN and RSPAN, each series consists of a random number of strings between two and five. Individuals are tested on three series of each length (12 series total). OSPAN and RSPAN scores (range: 0–42) consist of the total number of words or letters recalled from perfectly recalled trials. Span scores averaged across the two WM tests ranged from 5.5 to 39.5 ( $M = 16.07$ ,  $SE = 0.78$ ).

#### Procedure

Individuals first filled out a consent form and were then set up in front of a monitor controlled by a standard laboratory computer. Participants were introduced to modular arithmetic through a series of written instructions presented on the computer screen. Specifically, individuals were informed that they would be judging the truth value of modular arithmetic problems presented on the

computer as quickly and as accurately as possible by mentally computing the answers (i.e., without the use of paper). Participants were then provided with two example problems and answers that detailed an algorithm that could be used to solve the modular arithmetic problems with 100% accuracy. These examples illustrated that there were two steps one could take to determine the validity of the modular arithmetic problems, such as  $38 = 19 \pmod{4}$ : First, the middle number is subtracted from the first number (i.e.,  $38 - 19 = 19$ ). Second, this difference is divided by the mod number (i.e.,  $19 \div 4$ ). If the dividend is a whole number, the statement is *true*. If the dividend is not a whole number (i.e., as in the current example), the statement is *false*. Participants were told that when they had derived an answer to a problem, to press the corresponding *T* or *F* keys on the standard keyboard in front of them.

Problems were of the form  $45 = 27 \pmod{9}$ , with the value of each of the initial two numbers less than 100 and the value of the last number (i.e., the mod number) ranging from 2 to 9. Each problem was presented once across the entire experiment. Half of the modular arithmetic equations presented to participants were true, and the rest were false. Additionally, each true problem had a false correlate that only differed as a function of the number involved in the mod statement. For example, if the *true* problem,  $51 = 19 \pmod{4}$ , was presented, then a *false* correlate problem,  $51 = 19 \pmod{3}$ , was also presented at some point in the same problem block. This pairing was designed to equate the true and false problems as much as possible in terms of the specific numbers used in each equation (Beilock & Carr, 2005; Beilock et al., 2004).

All participants initially performed 12 practice problems to familiarize them with the math task. All practice problems were presented in a different random order to each participant. Each trial began with a 500-ms fixation point at screen center, which was immediately replaced by a math problem present until response. After response, the word *Correct* or the word *Incorrect* appeared for 1,000 ms, providing feedback. The screen then went blank for a 1,000-ms intertrial interval. Following the practice trials, individuals performed the experimental problem block preceded by instructions specific to the condition to which they had been assigned.

*Low-pressure condition.* Participants were simply told to work as quickly and as accurately as possible. Problems were presented in the same manner as the practice trials with the exception that individuals did not receive performance feedback. In addition, prior to the experimental problem block, individuals were told that the experimenters were interested in capturing the steps participants went through to derive answers to the problems and that, after certain problems, they would be prompted to write down how they solved the previous problem. Individuals were informed that there were no right or wrong processes to use in solving the problems and that the experimenters were interested in how individuals perform this type of math task, regardless of the processes used. In addition, participants were explicitly told that they could take as much time as they needed to write down how they solved the problem. These instructions were designed to encourage participants to report how they solved the problem without modification. Participants performed two practice problems, each followed by a strategy report, in order to ensure that they understood what was being asked of them.



Individuals performed 24 problems in the experimental block, presented in a different order for each participant. Following 8 of these problems (which were selected randomly, with the constraint that only 1 problem was selected out of every 3 problems in a sequence), participants were prompted by the computer to write down the processes they used to solve the previous problem. Specifically, appearing on the computer screen were the following instructions:

*Using the paper in front of you, please write down how you solved the previous problem. Please write down your memory of all the steps and processes you went through and any strategies you may have used in solving the previous problem.*

After writing down their problem-solving processes, participants pressed the space bar on the keyboard to continue. Participants were not aware of the specific problems after which they would be prompted for strategy reports. Following completion of the experimental block, participants were thanked and debriefed.

*High-pressure condition.* Individuals assigned to the high-pressure condition took part in the same procedure as those in the low-pressure condition, with some exceptions. Prior to the experimental block of problems, participants were given a pressure scenario that used several sources of pressure commonly experienced in real-world testing situations: Monetary incentives (e.g., future scholarships, educational opportunities), peer pressure, and social evaluation (e.g., admissions committees, parents, teachers, and peers). This scenario has been previously established to be highly effective in producing feelings of performance pressure and anxiety in participants (Beilock & Carr, 2001, 2005; Beilock et al., 2004; Markman, Maddox, & Worthy, 2006). These feelings do not differ as a function of math ability, and thus these factors are not confounded with response to pressure (Beilock & Carr, 2005; Beilock et al., 2004). This is also true with respect to WM capacity: As is stated below, posttest reports of perceived pressure and state anxiety did not differ as a function of individual differences in WM.

Specifically, the pressure manipulation involved informing individuals that the computer used a formula that equally takes into account reaction time and accuracy in computing a modular arithmetic score. Participants were told that if they could improve their modular arithmetic score by 20% relative to the preceding practice trials, they would receive \$10. Participants were also informed that receiving the monetary award was a "team effort." Specifically, participants were told that they had been randomly paired with another individual, and to receive their \$10, not only did the participant presently in the experiment have to improve in the next set of problems, but also the individual they were paired with had to improve. Next, participants were informed that this individual, "their partner," had already completed the experiment and had improved by the required amount. If the participant presently in the experiment improved by 20%, both the participant and his or her partner would receive \$10. However, if the present participant did not improve by the required amount, neither the participant nor his or her partner would receive the money. Finally, participants were told that their performance would be videotaped during the test situation so that local math teachers, students, and professors in the area could examine their performance on this new type of math task. Participants were additionally told that

their strategy reports, and the time involved in reporting such strategies, was not considered part of the time involved in solving the math problems and thus they should take as much time as they needed to write down their problem-solving strategies.

The experimenter set up the video camera on a tripod directly to the right of participants, approximately 1 m away. The field of view of the video camera included both the participant and the computer screen. Participants then completed the experimental block of problems. Afterward, the experimenter turned off the camera and faced it away from the participants.

Following the high-pressure problem block, individuals in the high-pressure condition filled out questionnaires that assessed their feelings of anxiety and performance pressure. These questionnaires were designed to ensure that feelings of pressure and anxiety induced by the high-pressure situation did not differ as a function of WM. Individuals first filled out the State Form of the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970). The STAI is a widely used measure of state anxiety consisting of 20 questions designed to assess participants' feelings at a particular moment in time. Individuals assign values to statements such as, *I feel calm* and *I feel at ease* on a 4-point scale ranging from 1 (*not at all*) to 4 (*very much so*). The State Form of the STAI has been used in a number of studies investigating the impact of anxiety on complex task performance (e.g., Tohill & Holyoak, 2000).

Following the STAI, participants answered questions related to their perceptions of performance under pressure (Beilock et al., 2004). Individuals were asked to rate on two 7-point scales (a) how important they felt it was for them to perform at a high level in the posttest (1 = *not at all important to me*; 7 = *extremely important to me*) and (b) how much pressure they felt to perform at a high level in the posttest (1 = *very little performance pressure*; 7 = *extreme performance pressure*).

Reports regarding the importance of performing at a high level in the high-pressure condition did not significantly differ as a function of WM ( $n = 44$ ;  $r = .26$ ,  $p = .10$ ;  $M = 4.95$ ,  $SE = 0.18$ ). Although one might be concerned that this relation is approaching significance, it is important to point out that importance reports have not differentiated pressure conditions in previous work. For example, using the same high-pressure scenario as the current work, Beilock et al. (2004) found that both low-pressure participants ( $M = 4.63$ ,  $SE = 0.21$ ) and high-pressure participants ( $M = 5.03$ ,  $SE = 0.19$ ) reported that it was at least "moderately important" to perform at a high level.

Most important, reports of perceived performance pressure ( $n = 44$ ;  $r = .12$ ,  $p = .44$ ;  $M = 4.93$ ,  $SE = 0.22$ ) and state anxiety ( $n = 44$ ;  $r = .13$ ,  $p = .94$ ;  $M = 49.21$ ,  $SE = 1.24$ ) that did reliably discriminate low- and high-pressure performance in previous work (see Beilock et al., 2004) did not significantly relate to WM in current work. These reports were nearly identical to those taken after high-pressure performance in Beilock et al.'s (2004) work (pressure:  $M = 5.08$ ,  $SE = 0.21$ ; anxiety:  $M = 42.68$ ,  $SE = 1.87$ ) and were substantially higher than low-pressure participants' reports in Beilock et al. (pressure:  $M = 3.95$ ,  $SE = 0.24$ ; anxiety:  $M = 32.08$ ,  $SE = 1.20$ ). Thus, given the above pattern of self-report data, it would be difficult to explain any of the WM differences in the current work in terms of those variables shown in previous work (e.g., perceived pressure, state anxiety) to reli-

ably differentiate individuals performing under low- and high-pressure conditions.

### Results

#### WM and Math Performance

We began by examining modular arithmetic problem accuracy as a function of individual differences in WM capacity and pressure condition for the 24-problem experimental block of interest. To do this, we regressed modular arithmetic accuracy on WM span score (i.e., average of the RSPAN and OSPAN), pressure condition (dummy coded), and their interaction. This regression resulted in no main effect of WM ( $\beta = .09$ ,  $t(88) = .87$ ,  $p = .39$ ), no main effect of pressure condition ( $\beta = .26$ ,  $t(88) = 1.05$ ,  $p = .30$ ), and a significant Working Memory  $\times$  Pressure interaction ( $\beta = -.47$ ),  $t(88) = -1.95$ ,  $p = .05$ . As can be seen in Figure 1, under low-pressure conditions, the higher the individuals' WM, the higher their modular arithmetic accuracy ( $r = .39$ ,  $p < .02$ ). In contrast, under high-pressure conditions, there was no relation between WM span and math accuracy ( $r = -.10$ ,  $p = .52$ ). Thus, more WM capacity benefits modular arithmetic task accuracy when performing in a low-pressure situation. In contrast, there is no relation between WM and accuracy when performing under high-pressure testing conditions.

Analysis of reaction times (RTs) for trials on which responses were correct neither contradicted the conclusions drawn from the accuracy data nor did they indicate that the accuracy effects could be attributed to a speed-accuracy trade-off. Specifically, regressing modular arithmetic RTs ( $M = 5,638$  ms,  $SE = 232$  ms) on WM span score, pressure condition, and their interaction produced no significant effect

of WM score ( $\beta = -.08$ ),  $t(88) = -.76$ ,  $p = .45$ , no significant effect of pressure ( $\beta = .02$ ),  $t(88) = 0.10$ ,  $p = .92$ , and no Pressure  $\times$  Working Memory interaction ( $\beta = -.36$ ),  $t(88) = -1.51$ ,  $p = .13$ . The lack of an interaction involving WM and pressure condition suggests that the above-reported differences in accuracy were not produced by a trade-off in RT. Moreover, the lack of RT effects is consistent with previous modular arithmetic work in which performance decrements under stress are consistently seen in accuracy (see Beilock & Carr, 2005; Beilock et al., 2004).

#### Strategy Selection and Math Performance

The analyses conducted thus far show how individual differences in the performer and variations in the performance environment combine to impact skill execution. Nonetheless, from these data alone, it is unclear as to how pressure changes the high-level performance that higher WM individuals normally exhibit or why individuals lower in WM are spared from pressure-induced failure. We turn to the problem-solving strategy reports to address these issues.

Strategies were coded independently by two experimenters unaware of participants' WM scores or the specific pressure condition to which participants were assigned ( $\alpha = .87$ ). All disagreements between the two raters were discussed, and if a consensus could not be reached, a third judge rated the strategy. Each written strategy was classified into one of the following three categories:

1. A WM intensive rule-based algorithm that involved a series of step-by-step computations. Examples include the following: "I subtracted the numbers in the 10s spot first, then the numbers in the 1s spot and divided;" "I subtracted the smaller number from the larger number, figured out the answer and found it did not divide equally, and I pressed F;" and "Looked at first number, took second number and subtracted it from first. Basically just imagined that instead of equal sign it was a minus sign. Took result and divided from third number."
2. Estimation or guessing based on previous associations with specific problem operands as a means to circumvent the steps necessary to explicitly compute a problem solution. Examples include the following: "Both numbers were even, so the result of subtracting would be even and probably divisible by the mod number;" "I knew the relation between the two numbers already as being multiples of 3 and only one multiple of 3 apart;" and "I know from seeing all the numbers group together that there was no possible way I could have gotten a number without a remainder."
3. A statement that did not make sense or in which not enough detail was given to code. Examples include the following: "Math in head." This last category occurred relatively infrequently (i.e., 8% of all responses).

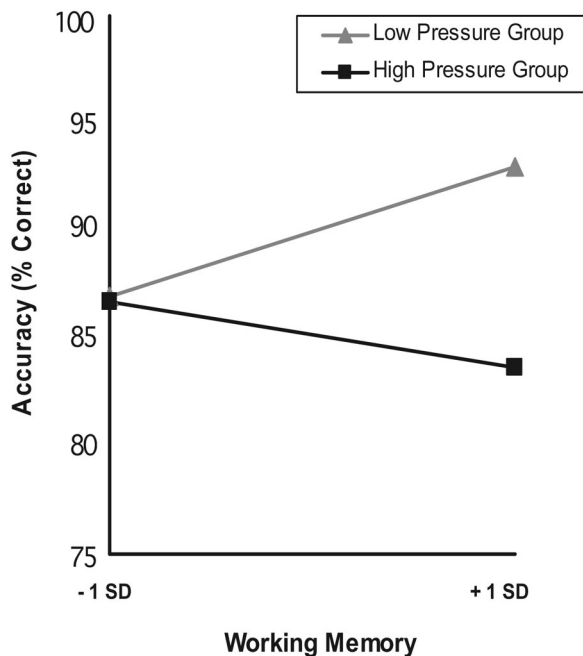


Figure 1. Mean modular arithmetic problem accuracy (% correct) as a function of individual differences in working memory and pressure condition. Nonstandardized coefficients are plotted at  $\pm 1$  standard deviation.

The percentage of strategy use in the first two categories was computed for each participant by dividing the number of strategy reports in that category by the total number of codable strategies. Because the proportion of algorithm and shortcut strategy reports represent inverses of each other, we only present the analyses for the proportion of rule-based algorithm use below.

We began by examining the relation between modular arithmetic performance and rule-based algorithm use. If rule-based processing leads to superior performance on our math task, then greater algorithm use should relate to better math performance. However, this may only hold under low-pressure conditions. To the extent that high-stakes situations reduce available WM resources (Beilock et al., 2004; Schmader & Johns, 2003), the use of an explicit rule-based algorithm to solve an equation under pressure may actually lead to worse performance than if one were to use a less demanding shortcut strategy. That is, a shortcut strategy that produces the right answer some of the time might lead to better overall performance than might an explicit rule-based algorithm if the computation of that algorithm is impaired due to insufficient WM resources.

To address the above issues, we turned to the performance measure shown in the previous set of analyses to be sensitive to fluctuations in performance pressure: modular arithmetic accuracy. We then regressed modular arithmetic accuracy on reported algorithm use, pressure condition (dummy coded), and their interaction. This regression resulted in no significant effect of pressure condition ( $\beta = .37$ ,  $t(88) = 1.41$ ,  $p = .16$ , no significant effect of algorithm use ( $\beta = -.01$ ,  $t(88) = -0.08$ ,  $p = .94$ , and a significant Pressure Condition  $\times$  Algorithm Use interaction ( $\beta = -.59$ ,  $t(88) = -2.26$ ,  $p < .03$ ). As can be seen in Figure 2, the greater the proportion of algorithm use, the relatively better the modular arithmetic performance under low-pressure conditions. In contrast, under high-pressure testing conditions, the greater the proportion of algorithm use, the relatively worse the performance. Thus, greater use of a rule-based algorithm does lead to relatively more accurate performance—as long as environmentally induced stressors do not impact the resources necessary to support such complex and working memory intensive computations.

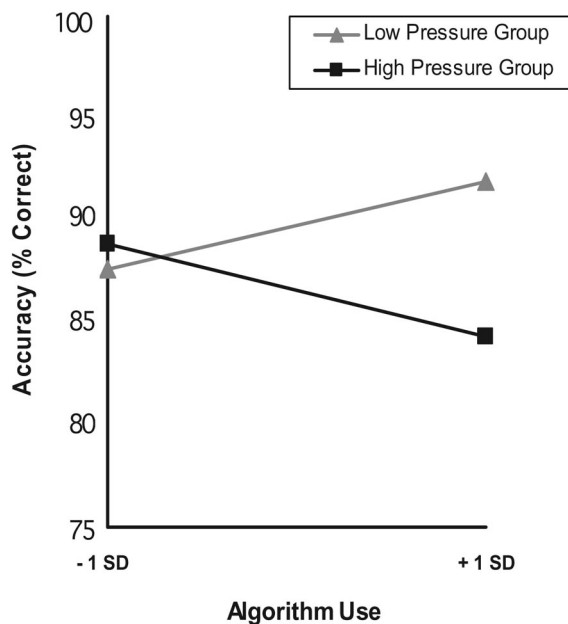


Figure 2. Mean modular arithmetic problem accuracy (% correct) as a function of the tendency to use rule-based algorithmic computations and pressure condition. Nonstandardized coefficients are plotted at  $\pm 1$  standard deviation.

### Strategy Selection and WM

We demonstrated in our initial set of analyses that the higher one's WM, the better his or her math performance, but only under low-pressure conditions. Our second set of analyses revealed that the greater the proportion of explicit rule-based algorithm use, the better one's performance, but again, only in low-pressure situations. If higher WM individuals' superior performance—at least under low-pressure conditions—can be explained by a greater use of rule-based processes, then this should be apparent in their strategy reports. Moreover, these strategy reports should also lend insight into why higher WM individuals' performance is less accurate in high-pressure than it is in low-pressure conditions.

To explore these ideas, we next regressed proportion of algorithm use on WM span score, pressure condition (dummy coded), and their interaction. This regression resulted in no significant effect of WM ( $\beta = .10$ ,  $t(88) = 0.98$ ,  $p = .33$ , an effect of pressure condition that approached significance ( $\beta = .43$ ,  $t(88) = 1.76$ ,  $p = .08$ , which was qualified by a significant Working Memory  $\times$  Pressure interaction ( $\beta = -.50$ ,  $t(88) = -2.05$ ,  $p < .05$ ). As seen in Figure 3, under low-pressure conditions, the higher individuals' WM, the greater their rule-based algorithm use ( $r = .31$ ,  $p < .04$ ). In contrast, under high-pressure conditions, there was no relation between WM span score and algorithm use ( $r = -.11$ ,  $p = .46$ ). Higher WM individuals' use of the rule-based algorithm was no different than their lower WM counterparts when performing under pressure. This evidence of simpler processing under pressure, together with the above analyses directly linking algorithm use to performance accuracy, provides a mechanistic account for why higher WM individuals under high-pressure show poorer math performance compared with higher WM individuals under low-pressure and why lower WM individuals do not show a differential pattern of performance under low- and high-pressure conditions.

### Discussion

Despite known differences in how various aspects of executive control (e.g., attentional control, inhibition) contribute to performance (Engle, 2002), less is known about how these differences manifest themselves in terms of the problem-solving strategies lower versus higher WM individuals use to solve complex, multistep problems such as mathematical computations (Price et al., 2007). The findings of Experiment 1 demonstrate that individual differences in WM influence how people approach difficult problems and that consequential testing situations can alter these approaches. Using dual-process theories of reasoning as a jumping board, we hypothesized that under low-pressure situations, the higher one's WM, the more likely he or she would be to rely on accurate (but computationally demanding) rule-based processing compared with simpler, associatively derived approaches. However, we also suggested that when pressure consumes the resources on which higher WM individuals normally rely to support explicit algorithm use, they might respond by using simpler (and less accurate) strategies. This is exactly what we found.

Thus, pressure does not simply serve to disrupt complex problem solving per se. Rather, when faced with high-pressure circumstances, individuals with characteristically higher capacity approach problems as if their capacity was lower to begin with, opting for simpler problem-solving strategies that alleviate the

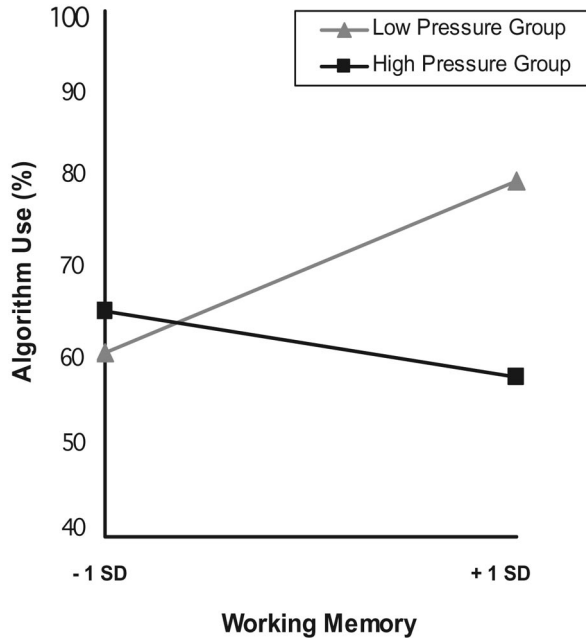


Figure 3. Proportion of rule-based algorithm use (%) as a function of individual differences in working memory and pressure condition. Non-standardized coefficients are plotted at  $\pm 1$  standard deviation.

burden on WM. Is this a poor response to a stressful situation? Because higher algorithm use under pressure led to relatively poorer performance than did lower algorithm use under pressure (see Figure 2), this strategy could be characterized as somewhat adaptive. Nonetheless, lower algorithm use under pressure still resulted in worse performance overall than did algorithm use when pressure was absent. Thus, the reliance on a simpler associative strategy seems limited in terms of performance success. Moreover, when considering strategy efficacy, it is also important to take the knowledge of the performer into consideration. As mentioned previously, in Siegler's (1988a) examination of students' arithmetic abilities, it was found that students differed in their problem knowledge as well as in their thresholds for relying on associatively based retrieval strategies. To the extent that a student has high problem knowledge and favors accuracy above all else, then a simpler association that only produces the right answer some of the time would not be more optimal (in terms of producing the correct problem answer) than would the use of a step-by-step problem-solving algorithm.

At this point, one might wonder exactly how often an associatively derived strategy would produce a correct answer on our modular arithmetic problems. However, determining a comprehensive base rate for associative strategy success proved difficult. First, it was hard to determine whether certain associatively derived strategies would always lead to the correct answer (e.g., "I know from seeing all the numbers grouped together that there was no possible way I could have gotten a number without a remainder"). Second, participants reported multiple association strategies that could not be applied to all math problems. For example, the strategy, "I rationalized that since the first number was even and the second number was odd, the answer would be an odd number that could not be divided by the mod number,

which was even," could not be applied to problems with all even or all odd numbers. Nonetheless, we did calculate an associatively derived base rate for strategies whose success rates could be definitively identified. Out of the 24 experimental problems presented to participants, a strategy of judging modular arithmetic problems in which the first two numbers consisted of one even number and one odd number and paired with an odd mod number (e.g.,  $82 = 55 \pmod{3}$ ) to be true or paired with an even mod number (e.g.,  $76 = 27 \pmod{6}$ ) to be false could be applied to 14 of the 24 experimental problems. A strategy of judging modular arithmetic problems in which the first two numbers were both odd and paired with an odd mod number (e.g.,  $63 = 27 \pmod{9}$ ) to be true or paired with an even mod number (e.g.,  $53 = 35 \pmod{6}$ ) to be false could be applied to 8 of the problems. Third, a strategy of judging equations in which the first two numbers were even and the mod number was even (e.g.,  $92 = 36 \pmod{6}$ ) to be true or paired with an odd mod number (e.g.,  $92 = 36 \pmod{7}$ ) to be false could be applied to 2 problems. In total, a combination of these strategies would result in success on our modular arithmetic problems 66% of the time. It should also be noted that some participants actually used the opposite assumption with respect to the last two strategies (e.g., "If two even numbers were to be subtracted, the division could only be true if the mod was an odd number," or "If two odd numbers were to be subtracted, the division could only be true if the mod was an even number"). This specific combination resulted in an 80% success rate on our modular arithmetic problems. Finally, the use of a multiple or factor approach in which one ignores even-odd and instead judges problems in which the mod number is not a factor of the other two numbers to be false (e.g.,  $76 = 27 \pmod{7}$ ) resulted in a success rate of 58%. Thus, while associatively derived shortcut strategies may circumvent the need to maintain and manipulate intermediate problem steps on line, they clearly do not produce the correct answer for all problems in our modular arithmetic task.

Admittedly, because Experiment 1's strategy reports were open-ended, this not only made it hard to determine a comprehensive base rate of success for associatively derived strategies, but also coding answers into either algorithm or simpler shortcuts left open the possibility that we were missing more nuanced differences in strategy use as a function of WM. To address this, in Experiment 2 we turned to a task in which problem-solving strategies were apparent from the answers given (i.e., participants' answers consisted of the formula they used to derive their solution). If, under pressure, higher WM individuals adopt problem-solving strategies that are of a similar type to those used by their lower WM counterparts, this will be directly evident from the answers reported.

## Experiment 2

Individuals performed Luchins's (1942) water jug task either under low-pressure or high-pressure testing conditions in Experiment 2. The goal of the water jug task is to derive a mathematical formula resulting in a specified "goal" quantity of water with jugs of various capacities by using the simplest strategy possible (see Figure 4). The first three problems can only be solved by using a difficult, WM demanding rule-based algorithm involving several subtraction and multiplication steps (i.e.,  $B - A - 2 \times C$ ). The last three problems are solvable by this same difficult formula rule or by a much simpler formula (e.g.,  $A - C$ ). Of interest is whether



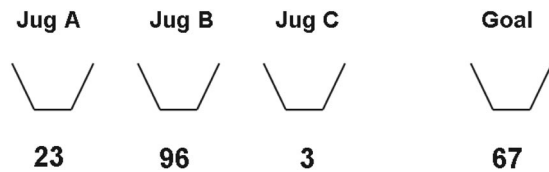


Figure 4. Water jug display. Participants derived a formula to obtain a “goal” quantity of water by using jugs of various capacities. The first three problems were only solvable by the formula  $B - A - 2 \times C$  (i.e., Fill Jug B, pour out enough to fill Jug A, then pour the remaining into Jug C twice, leaving the goal quantity in Jug B). The last three problems were solvable by this same difficult formula in addition to a much simpler formula (e.g.,  $A - C$ ). Individuals were informed that the water supply was unlimited and that not all jugs needed to be used.

Individuals recognize this simpler shortcut when available or whether they continue to use the same demanding formula. When individuals persist in using the difficult strategy in lieu of the simpler one, they exhibit mental set (McDaniel & Schlagler, 1990). Because the use of less effort to produce a correct answer is generally better, and individuals were instructed to solve the problems using the simplest strategy possible, the tendency to break mental set is considered success on this task (Gasper, 2003; cf. Fantino, Jaworski, Case, & Stolarz-Fantino, 2003). It is important to point out that although one might assume that continued use of the same demanding formula on every problem would render instantiation of that formula effortless, this is not necessarily the case. This is because recognizing that the complex formula will result in the goal quantity of water still involves performing a series of step-by-step computations to derive an end state.

One can also consider performance on Luchins’s (1942) water jug task from the standpoint of a dual-process framework. The use of the necessary step-by-step computations to arrive at the solution  $B - A - 2 \times C$  involves the explicit manipulation of a series of operands—a rule-based process. In contrast, immediately recognizing that  $A - C$  equals the goal quantity relies on very different processes—the retrieval of previously learned associations between these operands and the problem answer.

If lower WM individuals lack the ability to accurately compute complex solutions, they may be more likely to rely on associatively based strategies to recognize the simpler problem answer (e.g.,  $23 - 3 = 20$ ). In contrast, higher WM individuals may compute the complex solution—at least under low-pressure conditions—simply because they have the resources to do so. Indeed, as mentioned in the introduction, it has been demonstrated that those individuals least able to carry out accurate math problem computations often are the most likely to rely on associatively based retrieval strategies to circumvent these demanding processes (Siegler, 1988a). Thus, under low-pressure conditions, the lower individuals’ WM, the more likely they may be to find the simple solution and break mental set. Although such an idea might seem surprising given the emphasis placed on the positive role of WM and attentional control (Conway et al., 2005), it is consistent with the speculation that higher WM individuals use cognitively demanding strategies to solve problems that could be solved in more efficient ways (Barrett, Tugade, & Engle, 2004). And, if pressure prompts higher WM individuals to use the shortcut strategies that those with lower WM normally use, then individual differences in WM should not predict shortcut use under pressure because

everyone should rely on the simple shortcut in this more demanding situation.

### Method

#### Participants

One cannot look at the tendency to adopt a shortcut strategy (i.e., break mental set) following difficult problem performance if individuals do not perform the problems correctly to begin with (i.e., if they never achieve mental set). Thus, only individuals who correctly performed the first three problems (i.e., learned the complex, rule-based formula) served as study participants. Similar criteria have been used in previous research to ensure that only individuals who achieve mental set serve as participants (Gasper, 2003). This resulted in 45 low-pressure participants and 46 high-pressure participants (undergraduate students at the same U.S. Midwestern university used in Experiment 1). Seven additional participants were tested but not included in the experiment for the same reasons as those in Experiment 1: (a) Four participants in the pressure condition were excluded because they reported during the experiment that they either did not believe the pressure manipulation or knew about it ahead of time, and (b) 3 additional participants were excluded because the scores of their two WM measures differed by more than 20 points, suggesting that the RSPAN and the OSPAN were unable to capture consistent WM measures for these participants. Again, WM scores were based on the average of participants’ scores on the OSPAN and the RSPAN. Span scores averaged across the two WM tests ranged from 5.0 to 39.5 ( $M = 17.08$ ,  $SE = .75$ ).<sup>1</sup>

#### Procedure

Participants first signed informed consent and then were introduced to a computerized version of Luchins’s (1942) water jug task. On the computer screen, individuals were shown four water jug graphics labeled as *Jug A*, *Jug B*, *Jug C*, and *Goal* (see Figure 4). Individuals were instructed that the purpose of the task was to obtain a specified goal quantity of water with jugs of various capacities by using the simplest strategy possible. The water supply was unlimited.

Each problem appeared individually on the computer screen. Participants were instructed to solve the problems in their head and to use only their answer sheet to write down their final response—the formula used to obtain the goal quantity of water. Once participants had recorded their answer, they pressed the space bar

<sup>1</sup> Thirty low-pressure participants and 40 high-pressure participants did not perform the first three problems correctly and were not included in Experiment 2. It should be noted that span scores for these individuals ( $M = 14.22$ ,  $SE = .77$ ) were lower than for those who were included in the experiment ( $M = 17.08$ ,  $SE = .75$ ),  $F(1, 159) = 6.90$ ,  $p < .01$ . This is not surprising given that greater WM should aid in the implementation of the complex computations required to solve the first three problems. However, on average, those who did not achieve mental set had lower WM scores than those who did, which suggests that lower WM individuals were more likely to be excluded from the analyses of Experiment 2. Nonetheless, the criteria of having to solve the first three problems correctly did not appear to significantly alter the WM scores of Experiment 2, given that the span means did not differ from those found in Experiment 1 ( $M = 16.07$ ,  $SE = .78$ ;  $F < 1$ ).

on the keyboard to advance to the next problem. Individuals were provided with a paper packet to record their answers. One page was devoted to each problem so that participants could not readily refer back to previous answers. In addition, individuals were told that they did not necessarily have to use all the jugs to solve the problems and that there may be more than one way to solve a problem. During instructions, participants practiced the procedure, and the experimenter answered any questions to ensure they understood the task. All individuals only saw one example problem and answer: Jug A = 29, Jug B = 11, and Goal = 7. The example problem with answer only used two of the jugs so as to limit the similarity of the example to the experimental problem set. Participants worked out the answer (i.e.,  $A - 2 \times B$  or  $29 - 11 - 11$ ), and the experimenter checked to ensure they had computed it correctly.

Individuals performed six experimental problems (see Table 1) preceded by instructions specific to the low-pressure or high-pressure condition to which they had been assigned (see below). The first three experimental problems could only be solved with the equation  $B - A - 2 \times C$ . In the last three problems, two problem solutions were possible: Individuals could continue to use the difficult formula (i.e.,  $B - A - 2 \times C$ ), or alternatively, they could use a much simpler shortcut (i.e.,  $A - C$  or  $A + C$ ). Notably, the difficult equation requires the maintenance and manipulation of intermediate steps on line in WM.

*Low-pressure condition.* Prior to experimental problem performance, individuals were simply informed to solve the problems as quickly and as accurately as possible by using the simplest equation possible. Following completion of the water jug task, individuals were thanked and debriefed.

*High-pressure condition.* Prior to the experimental water jug problems, participants were given a pressure scenario similar to that used in Experiment 1. Specifically, individuals were told that the problems they were about to perform had been given to other university students the previous year and that the experimenters had derived average scores for these problems on the basis of students' performance. Participants were informed that if they could perform 20% better than the average student from the previous year (in terms of both RT and accuracy), they would receive \$10. Participants were then given the exact same "team effort" and video camera dialogue and manipulation used in Experiment 1.

Following completion of the experimental water jug problems, individuals in the high-pressure condition filled out the same anxiety (STAI; Spielberger et al., 1970), importance, and perfor-

mance pressure questionnaires (Beilock et al., 2004) used in Experiment 1 to ensure that lower versus higher WM individuals did not differ in their feelings of anxiety and performance pressure while performing the water jug task under the high-pressure condition.

Indeed, reports regarding the importance of performing at a high level ( $n = 46$ ;  $r = -.13$ ,  $p = .40$ ;  $M = 4.43$ ,  $SE = 0.24$ ), as well as reports of perceived performance pressure ( $n = 46$ ;  $r = -.06$ ,  $p = .67$ ;  $M = 4.93$ ,  $SE = 0.19$ ), and state anxiety ( $n = 45$ ;  $r = .01$ ,  $p = .97$ ;  $M = 50.91$ ,  $SE = 1.48$ ) did not differ as a function of WM in the high-pressure condition.<sup>2</sup> Thus, as in Experiment 1, it would be difficult to explain any observed WM differences in performance under pressure by general differences in perceived pressure or anxiety.

## Results

### Problem-Solving Strategies

Of central interest were the strategies that individuals used to solve the last three water jug problems. To reiterate, such problems were solvable by either the demanding algorithm used to solve the first three problems or by recognizing that a simpler, single-step shortcut exists. Moreover, as mentioned above, only those individuals who solved the first three problems correctly (only solvable with the complex algorithm) were retained as participants. One cannot look at the tendency to adopt a shortcut strategy (i.e., break mental set) following difficult problem performance if individuals do not perform the difficult problems correctly to begin with (i.e., if they never achieve mental set). Moreover, equating success on the first three problems across pressure condition and individual differences in WM allowed us to test whether variation in WM and the testing environment impacts the strategies individuals rely on to solve the last three problems when the complex algorithm is part of all individuals' performance repertoire.

Because we were interested in examining the propensity of shortcut strategy use when available, we began by regressing the number of shortcut strategies used in the last three problems on WM span score (i.e., average of RSPAN and OSPAN), pressure condition (dummy coded), and their interaction. This regression resulted in no significant effect of WM ( $\beta = -.09$ ),  $t(87) = -0.85$ ,  $p = .40$ , no significant effect of pressure condition ( $\beta = -.42$ ),  $t(87) = -1.53$ ,  $p = .13$ , and a significant Working Memory  $\times$  Pressure interaction ( $\beta = .546$ ),  $t(87) = 2.01$ ,  $p < .05$ . As can be seen in Figure 5, under low-pressure conditions, the higher individuals' WM, the less likely they were to recognize the shortcut strategy ( $r = -.32$ ,  $p < .04$ ). In contrast, under high-pressure conditions, there was no relation between WM span score and shortcut strategy use ( $r = .11$ ,  $p = .46$ ). Thus, more WM capacity leads to a lower likelihood of recognizing a shortcut strategy in a low-stress situation. In contrast, under pressure, higher WM individuals' shortcut strategy use equaled the level of their lower WM counterparts.

It is important to note that when individuals were not using the shortcut, they were using the complex algorithm that produced the correct answer almost all of the time. In fact, out of 91 participants,

<sup>2</sup> One participant did not fill out the STAI.

Table 1  
Water Jug Problems

Problem	Jug			Goal
	A	B	C	
1	23	96	3	67
2	11	48	6	25
3	20	59	4	31
4	23	49	3	20
5	15	39	3	18
6	14	36	8	6

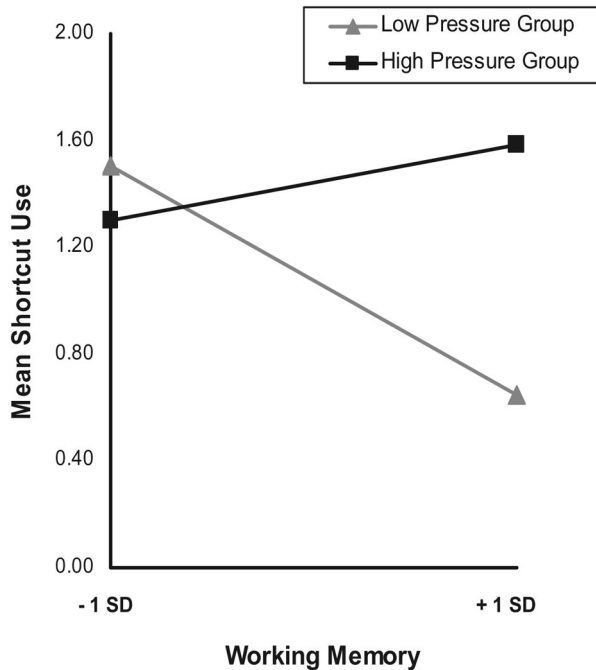


Figure 5. Mean shortcut strategy use in the last three problems as a function of individual differences in working memory and pressure condition. Nonstandardized coefficients are plotted at  $\pm 1$  standard deviation.

only 7 individuals reported one answer on the last three problems that did not produce the goal quantity of water (i.e., <3% of all answers). Thus, for a majority of the trials on which participants were not using the simplest solution possible, they were still coming up with a solution that worked roughly 97% of the time.

#### Problem-Solving RTs

The time to solve each problem was defined as the time from problem onset until participants recorded their answer and pressed the space bar to continue on to the next problem. We began by comparing RTs in the first three and the last three water jug problems for those participants with successful water jug solutions—regardless of what strategy they used (i.e., WM intensive algorithm vs. simpler strategy). Overall, problem-solving RTs were rather long (first three problems:  $M = 54.74$  s,  $SE = 2.96$  s; last three problems:  $M = 27.15$  s,  $SE = 2.86$  s) and a few RTs were exceptionally long (e.g., 234 s). In order to ensure that such outliers were not unduly influencing our results, we excluded participants with RTs more than three standard deviations above the mean of the first three and last three problems—three in total.<sup>3</sup>

In order to capture the within-subjects comparison of water jug RTs from the first three problems to last three problems, we performed a 2 (time: first three problems, last three problems)  $\times$  2 (condition: low pressure, high pressure)  $\times$  2 (WM: lower, higher) repeated measures analysis of variance (ANOVA) with lower versus higher WM span determined by a median split of the average of the RSPAN and OSPAN scores.<sup>4</sup> A significant three-way interaction obtained,  $F(1, 77) = 7.48$ ,  $p < .01$ ,  $MSE = 25.43 \times 10^7$ .

A 2 (time: first three problems, last three problems)  $\times$  2 (condition: low pressure, high pressure) ANOVA for the lower WM individuals resulted in only a main effect of time,  $F(1, 38) = 45.36$ ,  $p < .001$ ,  $MSE = 31.48 \times 10^7$ , in which the first three problems ( $M = 51.77$  s,  $SE = 3.81$  s) were solved more slowly than the last three problems ( $M = 24.12$  s,  $SE = 2.38$  s). There was no main effect of pressure ( $F < 1$ ) and no Pressure  $\times$  Time interaction,  $F(1, 38) = 1.1$ ,  $p = .30$ .

A similar ANOVA for the higher WM individuals resulted in a significant Time  $\times$  Pressure interaction,  $F(1, 39) = 9.69$ ,  $p < .01$ ,  $MSE = 19.54 \times 10^7$ . There was no significant difference in higher WM individuals' RTs for the first three problems as a function of pressure condition,  $F(1, 39) = 3.33$ ,  $p = .08$ ,  $MSE = 45.74 \times 10^7$ , although the RTs under the high-pressure condition ( $M = 59.86$  s,  $SE = 5.40$  s) were somewhat slower than RTs under the low-pressure condition ( $M = 47.67$  s,  $SE = 4.01$  s). In contrast, there was a significant difference in higher WM individuals' RTs for the last three problems as a function of pressure condition,  $F(1, 39) = 4.17$ ,  $p < .05$ ,  $MSE = 12.13 \times 10^7$  (low pressure:  $M = 28.79$  s,  $SE = 2.14$  s; high pressure:  $M = 21.77$  s,  $SE = 2.72$  s). Thus, pressure led to somewhat slower performance for the higher WM individuals on the first three problems that could only be solved with the difficult WM intensive algorithm but led to faster performance on the last three problems that could be solved via either the difficult algorithm or a simpler shortcut. This faster performance under pressure is consistent with the notion that higher WM individuals relied on the simpler solution under stress. And in fact, if one looks at lower WM individuals' performance on the last three problems, they were relatively fast under both the low-pressure condition ( $M = 24.22$  s, 3.08 s) and the high-pressure condition ( $M = 23.99$  s, 3.84 s), suggesting that they persisted in using the simpler shortcut regardless of the testing situation.

We also looked separately at RTs for correct solutions for the first three problems and the last three problems as a function of pressure condition and WM. We first regressed average problem-solving RTs for the first three problems on WM span score, pressure condition (dummy coded), and their interaction. This regression resulted in no main effect of WM ( $\beta = -.02$ ),  $t(77) = -.14$ ,  $p = .89$ , and a main effect of pressure condition that approached significance ( $\beta = -.55$ ),  $t(77) = -1.88$ ,  $p = .07$ , in which individuals tended to take longer to correctly solve the water jug problems in the high-pressure condition ( $M = 53.88$  s,  $SE =$

<sup>3</sup> Because there were so few problems in each time period (i.e., three), participants missing RTs either because they got one wrong ( $n = 7$ ) or because their problem-solving RT was three standard deviations above the group mean ( $n = 3$ ) were excluded from this analysis so as to ensure that our RT measure was based on the same number of observations per participant (i.e., we did not want some participants with only one RT observation and others with three). However, including these participants would not have changed the reported pattern of results.

<sup>4</sup> WM was treated as a median split (lower WM group:  $M = 11.38$ ,  $SE = .46$ ; higher WM group:  $M = 22.72$ ,  $SE = .86$ ) in this analysis so as to perform a repeated measures ANOVA that preserved the within-subject comparison of RTs for the first and last three problems. Nonetheless, regressing the RT difference from the first three to the last three problems on WM (as a continuous variable), pressure condition, and their interaction, resulted in the same significant pattern of performance as that reported in the text.



4.17 s) as compared with those in the low-pressure condition ( $M = 51.72$  s,  $SE = 3.15$  s). This main effect was qualified, however, by a Working Memory  $\times$  Pressure interaction ( $\beta = .645$ ),  $t(77) = 2.21$ ,  $p < .03$ . Under low-pressure conditions, the higher the WM, the faster the individual solved the problem ( $r = -.31$ ,  $p < .05$ ;  $-1$  standard deviation from the WM mean,  $M = 57.87$  s, and  $+1$  standard deviation from the WM mean,  $M = 44.95$  s). In contrast, there was no significant relationship between problem-solving RT and WM under high-pressure conditions ( $r = .19$ ,  $p = .20$ ). Thus, under normal low-pressure conditions, the higher individuals' WM, the less time they spent solving the first three water jug problems. Under pressure, problem-solving time did not differ as a function of WM. This mirrors the findings of Experiment 1 and of previous work (Beilock & Carr, 2005; Gimmig et al., 2006), demonstrating that the performance advantage of those greater in WM is eliminated in high-stress situations.<sup>5</sup>

Because individuals higher in WM solved the water jug problems faster—at least under low-pressure conditions—this brings up the possibility that the use of the simpler shortcut solution on our task may not have been any more efficient than producing the complex algorithm for these participants. To explore this issue, we looked at the relationship between problem-solving RT and WM in the last three problems that could be solved either by a WM intensive algorithm or by the simpler shortcut. We also added shortcut strategy use as a factor. Specifically, we regressed problem-solving RT on shortcut strategy use, WM, pressure condition (dummy coded), and their interaction. This analysis only produced a main effect of shortcut strategy use ( $\beta = -.75$ ),  $t(76) = -2.76$ ,  $p < .01$ , in which the higher the shortcut strategy use, the faster the RT. No other main effects or interactions were significant (all  $ps > .26$ ). To the extent that the use of the shortcut is not only simpler (in terms of the number of problem-solving steps) but also most efficient (in terms of problem-solving time), then there should be a relation between shortcut use and RT such that increased shortcut use is related to faster RT. And this relation should not be dependent on WM. This is exactly what was found. Not only does this analysis demonstrate that the use of the shortcut was the quickest way to solve the problems (i.e., even more so than implementing the demanding formula used previously), but also it suggests that individuals did not merely arrive at the shortcut after spending time searching for the multistep algorithm or computing other possible solutions. If this were the case, it seems unlikely that one would observe a negative relation between problem-solving time and shortcut use. Thus, less WM availability—whether manifested by the capacity an individual brings to the table to begin with or the product of a high-pressure situation—serves to draw individuals in the water jug task to the efficient simple equation.<sup>6</sup>

### Discussion

Experiment 1 demonstrated that pressure prompted higher WM individuals to use the simpler (and less efficacious) problem-solving strategies of the type typically used by those lower in WM. Experiment 2 directly examined this mechanism of performance under pressure through the use of a math task in which problem-solving strategies were apparent from the answers given. Results demonstrated that under low-pressure conditions, individuals lowest in WM were most likely to recognize the shortcut strategy when available. Under pressure, higher WM individuals recog-

nized the shortcut strategy at a level equal to their lower WM counterparts.

When individuals persist in using the difficult strategy in lieu of the simpler one on this task, they exhibit mental set (McDaniel & Schlager, 1990). Because, all things being equal, the use of less effort (in terms of number of both problem-solving steps and time) to produce a correct answer is generally better, the tendency to use the simple formula is preferable on this task. Thus, unlike the modular arithmetic task used in Experiment 1, the use of a simpler strategy when available denotes successful performance in Luchins's (1942) water jug task. Although WM capacity has been shown to be positively related to several higher level cognitive functions such as general intellectual ability, reasoning, and analytic skill, and is touted as one of the most powerful predictive constructs in psychology (Conway et al., 2005), the findings of Experiment 2 demonstrate that for certain task types, the relation between individual differences in WM and performance can be not only absent but also reversed.

### General Discussion

The current work demonstrates that individual differences in WM affect how individuals approach math problem-solving tasks and that these approaches can differ depending on the demands of the environment under which performance takes place. Although one might assume that higher WM capacity should provide greater facility with on line computations and also support more efficacious strategy use (Schunn & Reder, 2001), we demonstrate that those higher in WM are, ironically, less apt to use simple problem-solving strategies than are their lower capacity counterparts—even when such strategies produce the desired outcome. How can this be explained?

<sup>5</sup> One might wonder whether the fact that higher WM was associated with faster RTs under low-pressure is indicative of higher WM individuals learning the complex algorithm to a higher level than their lower WM counterparts. This is despite the fact that, regardless of WM, only those who solved the first three problems correctly were retained in the analyses. Nonetheless, if this possibility were true, it could explain why, under low-pressure conditions, the higher the individuals' WM, the less likely they were to abandon the complex algorithm for the simpler strategy. To support such an idea, one would expect to see a significant relation between problem-solving RT in the first three problems and shortcut strategy use. That is, the shorter the RT in the first three problems, the less likely one should be to rely on the simpler strategy in the last three problems. However, such a relation did not occur ( $r = .20$ , *ns*).

<sup>6</sup> Experiment 2 represents an ideal test bed for looking at the relation between problem-solving RT and strategy use as both the complex algorithm and simpler shortcut in the water jug task can lead to a correct problem-solving solution. This is not the case in Experiment 1. Unlike a correctly computed rule-based algorithm, associatively derived shortcuts do not always lead to the correct answer. Thus, only looking at correct RTs in Experiment 1 proves problematic for assessing the relation between shortcut use and problem-solving time. Although one might propose that such an analysis could be performed by taking into account incorrect problem-solving RTs as well, this would entail making inferences about performance on incorrect trials. The multiple correct solution paths of Experiment 2 allow us to demonstrate that increased shortcut use is indeed related to shorter problem-solving RTs in a way that the modular arithmetic task used in Experiment 1 does not.



Using dual-process theories of reasoning (for a review, see Evans, 2003; Smith & DeCoster, 2000; Stanovich & West, 2000) along with associative models of arithmetic (Siegler, 1988b; Siegler & Lemaire, 1997; Siegler & Shipley, 1995; Siegler & Shrager, 1984), we hypothesized that because lower WM individuals have less capacity to compute demanding problem computations on line, they may be more likely to rely on associatively derived answers that make few demands on attentional control rather than complex rule-based computations. In contrast, because those higher in WM are able to support demanding computations, these individuals may be more likely to rely on rule-based algorithms to derive problem solutions. And, to the extent that performance pressure impacts those resources on which higher WM individuals usually rely to compute rule-based algorithms, the problem-solving strategies these individuals use may be different under low-pressure and high-pressure situations. This is exactly what we found.

In Experiment 1, individuals higher in WM outperformed their lower capacity counterparts on the modular arithmetic task under low-pressure conditions. Yet, higher WM individuals' performance equaled the level of lower WM individuals when under pressure. Such a finding is consistent with work by Kane and Engle (2002), who found that adding a demanding secondary task to the performance of a proactive interference memory task essentially made higher WM individuals look like their lower WM counterparts (see also Rosen & Engle, 1997). It is also consistent with work in our lab and in others that has examined the detrimental impact of pressure on performance (Beilock & Carr, 2005; Gimmig et al., 2006).

Our findings, however, go beyond mere performance outcomes. Reports of problem-solving strategies revealed that higher WM individuals' tendency to rely on computationally demanding rule-based problem-solving strategies was just what made them susceptible to fail under pressure. Under low-pressure conditions, higher WM individuals were more likely to report implementing complex, multistep problem-solving algorithms. Lower WM individuals more often reported that they used a simpler strategy involving a reliance on previous associations with problem operands to derive their answer (e.g., "Both numbers were even, so the result of subtracting would be even and probably divisible by the mod number"). This type of shortcut produced the correct answer more often than chance yet did not result in a performance level equal to those higher in WM. However, because such shortcut strategies circumvent the need to maintain and manipulate intermediate problem steps on line, they are relatively immune to performance pressure's negative impact. And, when under pressure, higher WM individuals used simpler strategies as well.

The findings from Experiment 1 are consistent with work examining the relation between individual differences in general fluid intelligence (gF) and performance on a challenging WM task (i.e., the three-back task). J. R. Gray, Chabris, and Braver (2003) had individuals view a series of words or faces (with a new stimulus item appearing every few seconds) and instructed them to indicate as quickly and as accurately as possible whether each new stimulus matched the stimulus seen three items previously. Gray et al. found that susceptibility to "lure" trials (i.e., two-back, four-back, or five-back) correlated with gF such that individuals low in gF were more likely to fall prey to lure trials than were those higher in gF. Moreover, neural activity in the lateral prefrontal

cortex, thought to be involved in effortful reasoning, mediated the relation between gF and performance on lure trials. Thus, in light of the current work, higher gF individuals could be characterized as using a high-effort, rule-based reasoning strategy to correctly reject lures, whereas lower gF individuals appeared more likely to rely on associations between the lure items and recency or familiarity to make their judgments.

We turned the tables in Experiment 2 by presenting individuals with a problem-solving task in which a simpler process rather than a demanding algorithm produces optimal performance (Luchins, 1942). Not only does this type of task allow for the unconfounding of the typical relation between successful performance and WM capacity, but also it is advantageous because individuals' problem-solving strategies are apparent from the answers given (i.e., participants' problem-solving answers consisted of the formula they used to derive their solution). Under low-pressure conditions, lower WM individuals outperformed their higher WM counterparts, recognizing the simpler shortcut solution to solve the water jug problems. But when put in a high-pressure situation thought to tax the resources on which higher WM individuals normally rely (Beilock & Carr, 2005), higher WM individuals used the simpler strategy just like those lower in WM.

Although one might assume that higher WM capacity should allow greater facility with on line computations and also support more efficacious strategy use, in the current work we demonstrate that those higher in WM are, ironically, less apt to use simple performance strategies than are their lower capacity counterparts—even when such strategies produce optimal performance. Given a scientific literature that emphasizes the positive aspects of WM and attentional control (Miyake & Shah, 1999), one might suppose that those higher in WM should always outperform their lower WM counterparts. How can this be explained?

To the extent that higher WM individuals are especially good both at focusing their attention on select task properties and at ignoring others, these individuals may actually be worse at detecting alternate problem solutions. Rationale for this idea comes from Conway, Cowan, and Bunting's (2001) investigation of the performance of individuals lower and higher in WM in a dichotic listening paradigm. Individuals were told to listen to a message in one ear and to ignore a message in the other ear (in which their name was sometimes mentioned). Lower WM individuals were more likely to notice their name in the unattended ear than were higher WM individuals, suggesting that lower WM individuals were allocating attention to information both focal and disparate to the task at hand. By analogy to the current work, higher WM individuals may be especially good at focusing their attention on certain task properties and at ignoring others, whereas lower WM individuals may not be able to allocate attentional resources solely to one task approach. As a result, low WM individuals may actually be more likely to recognize alternative problem solutions.

Added support for this idea comes from recent work by Ricks, Turley-Ames, and Wiley (in press) that examined verbal problem-solving performance as a function of WM and domain-relevant knowledge. Ricks et al. found that for individuals high in domain-relevant knowledge that went down a wrong problem solution path, performance was actually poorer the higher they were in WM. Ricks et al. suggested that because high WM individuals are essentially too good at maintaining attention to the wrong information, they are less likely to abandon the wrong solution path to

find the correct one. Thus, higher WM individuals may at times have difficulty identifying the most efficient performance strategies—at least until situation-induced pressures limit the WM resources on which these individuals normally rely.

Our work also seems similar to research in the expertise literature showing that novices or individuals with intermediate problem knowledge sometimes outperform their expert counterparts. For example, in Chase and Simon's (1973) classic study of expert and novice chess players, experts recalled more chess pieces than did their novice counterparts when chess boards were presented in actual game-play configurations. However, novices somewhat outperformed experts when chess boards were randomized. To the extent that chess experts were attempting to represent nonexistent meaningful patterns in the randomized boards that novices were not, this may have disrupted their ability to remember randomized boards for which such normality did not exist. Furthermore, in X-ray diagnosis, although expert radiologists have better memory for the atypical features of X-rays that they have seen, they are worse than are novices at recognizing X-rays without such abnormalities (Myles-Worsley, Johnston, & Simons, 1988). The knowledge that experts use to help them diagnose abnormalities seems to hinder their ability to encode normal X-rays. In Experiment 2 of the current work, we show that the advantages higher WM individuals usually possess as a result of their ability to compute demanding algorithms may impede performance in situations in which such computations are not needed.

#### *Multiple Routes to Skill Failure*

We have conceptualized performance pressure as either harming or aiding performance via the consumption of WM resources that could otherwise be allocated to math task execution. However, it should be noted that there is a large body of work exploring pressure-induced failure in well-learned skills that operate largely outside of WM. And, indeed, this work suggests that a very different mechanism underlies skill failure. According to explicit monitoring theories (Baumeister, 1984; Beilock & Carr, 2001; R. Gray, 2004; Lewis & Linder, 1997; Masters, 1992), performance pressure increases self-consciousness about performing correctly, which in turn induces individuals to increase the attention they devote to controlling step-by-step performance in order to ensure a positive outcome. Unfortunately, increased attention to proceduralized task control can backfire, disrupting what should have been fluent, automatic execution (Beilock & Carr, 2001; Kimble & Perlmutter, 1970; Langer & Imber, 1979). Support for explicit monitoring theories comes mainly from complex sensorimotor skills such as golf putting, soccer dribbling, and baseball batting that become proceduralized with extended practice. Such skills are not harmed when WM capacity is reduced, for example, by a memory load or concurrent task, but they are hurt by directing explicit attention to automatic processes that normally proceed outside WM's control (Beilock, Carr, MacMahon, & Starkes, 2002; R. Gray, 2004; Jackson, Ashford, & Norsworthy, 2006).

Because the problem-solving tasks in the current work are not based on highly practiced proceduralized knowledge structures in the same way that a well-learned golf putting or soccer dribbling task might be (Beilock & Carr, 2001; Beilock et al., 2002; R. Gray, 2004; Jackson, Ashford, & Norsworthy, 2006; Lewis & Linder, 1997), pressure-induced attention to execution should not impact

performance—after all, there are no multistep procedures to disrupt. In contrast, pressure-induced consumption of WM resources does impact performance in the current experiments. Of course, whether such an impact is harmful depends on whether the problem-solving task being performed optimally requires explicit rule-based computations (i.e., the demanding modular arithmetic computations of Experiment 1) or does not (i.e., the simple shortcut strategies for the water jug task in Experiment 2). Such ideas underscore the importance of considering the multiple determinants of the task, environment, and performer in capturing the success and failure of complex skills (see also Markman, Maddox, & Worthy, 2006).

#### *Implications for High-Stakes Testing*

Our work demonstrates that the advantages individuals higher in WM have on the types of demanding math problems used in Experiment 1 and those that high-stakes tests often embody (Sternberg, 2004) are just what make them susceptible to failure when pressure is added. One might wonder how this could be the case, given that high-stakes testing has been used to gauge students' ability for many decades. Here we show how important testing situations limit the efficacy of these evaluations. These results align with recent concerns regarding the ability of admissions tests to elicit optimal performance in underrepresented groups, especially high-achieving racial minorities and women in the math and sciences (Atkinson, 2001; Beilock et al., 2007; Steele, 1997; Sternberg & Williams, 1997). Such individuals feel added pressure to perform at a high level, often in an effort to overcome well-known and widely held stereotypes regarding the intelligence or academic skill of the social groups to which they belong (Beilock & Carr, 2005; Beilock, Jellison, Rydell, McConnell & Carr, 2006; Beilock et al., 2007; Sternberg & Williams, 1997). The finding that the more important a test is, the more likely the best performers will take up the strategies of the worst demonstrates just how perilous a strong reliance on test scores may be, especially for those most in need (and deserving) of high-level performance to advance in academics and beyond. Ironically, the conditions under which admissions tests are conducted may impact the very constructs they are attempting to measure.

#### *Conclusions*

In two experiments, we capitalize on variations in the skill execution environment and the demands of the task being performed to identify key differences in the strategies that individuals lower and higher in WM use to solve difficult math problems and how consequential testing situations impact such strategy use. Despite the well-established link between WM and performance, less is known about how individual differences in executive functioning manifest themselves in terms of the problem-solving strategies lower versus higher WM individuals use to solve complex problems (Price et al., 2007). Moreover, even less work has examined how environmental factors such as stressful situations impact problem solving across tasks with different computational demands. We demonstrate that the availability of WM resources influences how individuals approach math problems, and the demands of the task being performed dictate whether such approaches will result in skill success or failure. Such research not

only lends insight into the development of training and performance strategies designed to reveal optimal performance when it is most important, but also it highlights the importance of accounting for the complexities of the real world in the development of comprehensive theories of executive functioning.

## References

- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*, 224–237.
- Atkinson, R. C. (2001, February 18). *Standardized tests and access to American universities*. The 2001 Robert Atwell Distinguished Lecture, 83rd Meeting, American Council on Education, Washington, DC.
- Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*, 553–573.
- Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, *46*, 610–620.
- 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., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied*, *8*, 6–16.
- Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality & Social Psychology Bulletin*, *32*, 1059–1071.
- 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.
- Bogomolny, A. (1996). *Modular arithmetic*. Retrieved March 1, 2000, from <http://www.cut-the-knot.com/blue/Modulo.shtml>
- 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.
- Carlson, R. A. (1997). *Experienced cognition*. Mahwah, NJ: Erlbaum.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.
- Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*, 331–335.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.
- Daneman, M., & Carpenter, P. A. (1980). Individual-differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.
- DeCaro, M. S., Wieth, M., & Beilock, S. L. (2007). Methodologies for examining problem-solving success and failure. *Methods*, *42*, 58–67.
- De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, *17*, 428–433.
- DeStefano, D., & LeFevre, J. A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16*, 353–386.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.
- Ericsson, K. A., & Delaney, P. R. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 257–297). New York: University Press.
- Evans, J., St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Science*, *7*, 454–459.
- Fantino, E., Jaworski, B. A., Case, D. A., Stolarz-Fantino, S. (2003). Rules and problem solving: Another look. *American Journal of Psychology*, *116*, 613–632.
- Gaspar, K. (2003). When necessity is the mother of invention: Mood and problem solving. *Journal of Experimental Social Psychology*, *39*, 248–262.
- Gimmig, D., Huguet, P., Caverni, J. P., & Cury, F. (2006). Choking under pressure and working memory capacity: When performance pressure reduces fluid intelligence. *Psychonomic Bulletin & Review*, *13*, 1005–1010.
- Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*, 316–322.
- Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*, 42–54.
- Jackson, R. C., Ashford, K. J., & Norsworthy, G. (2006). Attentional focus, dispositional reinvestment, and skilled motor performance under pressure. *Journal of Sport & Exercise Psychology*, *28*, 49–68.
- Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 336–358.
- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*, 637–671.
- Kimble, G. A., & Perlmutter, L. C. (1970). The problem of volition. *Psychological Review*, *77*, 361–384.
- Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*, 26–52.
- Langer, E. J., & Imber, L. (1979). When practice makes imperfect: Debilitating effects of overlearning. *Journal of Personality and Social Psychology*, *37*, 2014–2024.
- Lewis, B., & Linder, D. (1997). Thinking about choking? Attentional processes and paradoxical performance. *Personality & Social Psychology Bulletin*, *23*, 937–944.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.
- Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, *54*, 1–95.
- Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science*, *17*, 944–948.
- Masters, R. S. W. (1992). Knowledge, knerves and know-how: The role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. *British Journal of Psychology*, *83*, 343–358.
- McDaniel, M. A., & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction*, *7*, 129–159.
- Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: University Press.

- Myles-Worsley, M., Johnston, W., & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 553–557.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). New York: Academic.
- Price, J., Catrambone, R., & Engle, R. W. (2007). When capacity matters: The role of working memory in problem solving. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems* (pp. 49–76). NY: Erlbaum.
- Reder, L. M., & Schunn, C. D. (1999). Bringing together the psychometric and strategy worlds: Predicting adaptivity in a dynamic task. In D. Gopher & A. Koriat (Eds.), *Cognitive regulation of performance: Interaction of theory and application. Attention and performance XVII* (pp. 315–342). Cambridge, MA: MIT Press.
- Ricks, T. R., Turley-Ames, K. J., & Wiley, J. (in press). Effects of working memory capacity on mental set due to domain knowledge. *Memory & Cognition*.
- Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*, 213–227.
- Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science*, *17*, 954–958.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440–452.
- Schunn, C. D., & Reder, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General*, *130*, 59–76.
- Siegler, R. S. (1988a). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development*, *59*, 833–851.
- Siegler, R. S. (1988b). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*, 258–275.
- Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, *126*, 71–92.
- Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Erlbaum.
- Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.
- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality & Social Psychology Review*, *4*, 108–131.
- Spielberger, C. C., Gorsuch, R. L., & Lushene, R. (1970). *State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychology Press.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–726.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.
- Sternberg, R. J. (2004). Theory-based university admissions testing for the new millennium. *Educational Psychologist*, *39*, 185–198.
- Sternberg, R. J., & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists? A case study. *American Psychologist*, *52*, 630–641.
- Stevenson, L. M., & Carlson, R. A. (2003). Information acquisition strategies and the cognitive structure of arithmetic. *Memory & Cognition*, *31*, 1249–1259.
- Tohill, J. M., & Holyoak, K. (2000). The impact of anxiety on analogical reasoning. *Thinking and Reasoning*, *6*, 27–40.
- Trbovich, P. L., & LeFevre, J. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition*, *31*, 738–745.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154.

Received March 8, 2007

Revision received July 17, 2007

Accepted July 18, 2007 ■

### E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!