AMERICAN PSYCHOLOGICAL ASSOCIATION SARMAC Journal of Applied Research in Memory and Cognition

ISSN: 2211-3681

https://doi.org/10.1037/mac0000023

EMPIRICAL ARTICLE

Performing Up to Par? Performance Pressure Increases Undergraduates' Cognitive Performance and Effort

Almaz Mesghina[1], Natalie Au Yeung[2], and Lindsey Engle Richland[2]
[1] Department of Psychology, Northwestern University, United States
[2] School of Education, University of California, Irvine, United States

Performance measures, including standardized test scores or cognitive tasks, are commonly conceptualized as stable measures, yet are often unreliable indices of skill. We examine two contextual factors, performance pressure and feedback, that may influence the extent to which individuals demonstrate their cognitive capacity by manipulating uncertainty and thereby changing the nature of participants' cognitive task engagement. We manipulate pressure prior to adults completing two cognitive tasks: a working memory (WM) and verbal reasoning task with some (Study 1) or no performance feedback (Study 2). Pressure increased demonstrated WM capacity, which could be explained by increased task-directed effort. The incentivizing effects were greater when feedback was provided. Those under pressure maintained their motivation to perform, which predicted performance gains, despite being more stressed and anxious than controls. Combined, this suggests that often relied upon cognitive performance indices may be malleable to contextual features and might not reflect true capacity or potential.

General Audience Summary

How can we encourage individuals to exert their full effort when performing? This is an important question for educators who rely on students' test scores to determine course placement, scholarship eligibility, and college and graduate school admittance. This is also important for researchers who use scores on cognitive tasks (e.g., working memory [WM] tasks) as diagnostic tools or to characterize differences between individuals in their cognitive capacity (i.e., categorizing people as high or low WM). In both cases, we assume participants are putting forth maximal effort. In reality, however, this is not always true. Research in laboratory and real-world settings finds that slight changes to performance contexts can make individuals more likely to put forth effort, which leads to more positive experiences during the task and improved performance. Much of the research focuses on the experience of uncertainty while performing: When reward for good performance on a task (e.g., money) is possible, but not certain, people become more motivated and exert more effort. Across two studies, we made some participants feel heightened uncertainty while completing two cognitive tasks. We heightened uncertainty in two ways: first, we added pressure. After participants tried the tasks, we told some of them that they then needed to perform above 90% on both tasks in order for them and another participant to get additional compensation. Second, we removed all feedback, meaning that participants would not be able to see how well they were doing. Even though pressure led to increased anxiety and stress, participants under pressure were more motivated than those not under pressure. With feedback, participants under pressure put forth more effort, which improved performance on a WM task. Our findings suggest that researchers and educators may underestimate individuals' true abilities and potential when focusing on scores without considering performance context.

Almaz Mesghina https://orcid.org/0000-0003-4001-5541
Lindsey Engle Richland https://orcid.org/0000-0003-1514-6013

This work was supported by the Institute of Education Sciences, U.S. Department of Education, under Grant Numbers R305B140048 and R305A170488 to the University of Chicago; and the National Science Foundation, under Grant Number 32027447 to the University of California, Irvine. The opinions expressed are those of the authors and do not necessarily represent views of the funders. We would like to thank our research assistants for their help with data collection and data entry.

The authors declare no conflict of interest.

Almaz Mesghina and Lindsey Engle Richland conceived of the overarching concept. Almaz Mesghina designed the studies and created all materials. Almaz Mesghina and Natalie Au Yeung oversaw the data collection and participant recruitment, including training research assistants in research protocols and data entry. Almaz Mesghina analyzed and interpreted the data and wrote the manuscript, with substantial revisions and feedback provided by Lindsey Engle Richland.

Correspondence concerning this article should be addressed to Almaz Mesghina, Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208, United States. Email: mesghina@northwestern.edu

Keywords: performance pressure, feedback, uncertainty, working memory, effort

Supplemental materials: https://doi.org/10.1037/mac0000023.supp

Is task performance a meaningful measure of individuals' cognitive capabilities? The distinction between learning, capacity, and performance is key to unpacking the factors underpinning scores on any behavioral task (Soderstrom & Bjork, 2015), since in many behavioral studies, performance is the only source of measurement, yet assessment of internal capacity is the intended aim. Studies across motor and cognitive skills assessments show that performance can be an unreliable index of ability for many reasons. Sometimes performance scores provid e misleading data about whether training has been effective (e.g., Lee & Magill, 1983). For example, research in applied memory shows high performance in a blocked practice study condition may in fact signal poor long-term skill acquisition, while lower performance during a spaced study intervention may be misleading, yielding higher long-term performance (Dempster, 1988). Similarly, cross-cultural research has revealed that sociocultural context and participant expectations can systematically impact task performance (e.g., Peña, 2007). Despite using an identical task, or a linguistically accurate translation, there remain considerable cultural differences in expectations, language semantic and pragmatic equivalencies, and related factors when a task is administered across communities, suggesting that task performance here may not be a meaningful index of participants' cognitive skills, but rather reflects the sociocultural context.

In this article, we draw attention to the role of *uncertainty* in the performance context as another contextual factor rarely considered in memory and cognition research yet we posit is critical to ensure that performance on a task provides a meaningful index of participants' skills and knowledge. We argue that uncertainty of one's performance during a task, as well as uncertainty of success on an important measure (e.g., a high stakes test), can have important yet underrecognized impacts on performance.

Uncertainty, Arousal, and Optimal Performance

Uncertainty is theorized to relate to performance through arousal following an inverted-U shape, where moderate, transient arousal best promotes performance, and too little or too much arousal can compromise performance (Sapolsky, 2015; Yerkes & Dodson, 1908). Given the possibility of underperforming, optimizing arousal is key to assessing individuals' true cognitive potential. Researchers studying declarative and episodic memory have found that learning and memory improve when individuals perform with the potential for reward (see Miendlarzewska et al., 2016). For example, engineering undergraduates learned more from an educational game when there was uncertainty of reward (rolling dice) than when reward values were certain (constant point values; Ozcelik et al., 2013).

Indeed, the uncertainty of reward, not the reward itself, is arousing (Miendlarzewska et al., 2016), which heightens individuals' motivation, dopaminergic activation, and goal-directed attention, leading to improved memory and learning (see Howard-Jones & Jay, 2016; Miendlarzewska et al., 2016). This anticipation of reward and concurrent uncertainty can promote further emotional engagement with the content (Howard-Jones & Demetriou, 2009) and yields improvements in memory and cognitive performance in an inverted-U fashion (Cheng et al., 2020). Accordingly, calls to "gamify" educational and task performance contexts, wherein uncertainty of reward is embedded, have increased (Luria et al., 2020).

One explanation for the benefits of gamification may be affective, such that uncertainty creates enjoyment which is arousing and increases effort. Similarly, pressure, in the form of performance-contingent rewards or raising the stakes, has been shown to incentivize performance on a variety of tasks, including standardized tests (e.g., graduate record exam [GRE]; Attali, 2016; Schlosser et al., 2019) and working memory (WM) tasks (e.g., Heitz et al., 2008; Jimura et al., 2010). Yet, pressure can also threaten performance by inducing verbally rehearsed worries that co-opt those very WM resources (see Beilock, 2008). When the stakes for task performance are elevated, the uncertainty of meeting a performance criterion become increasingly impactful on limited cognitive resrources like WM, which are involved in both verbally rehearsed worry and attention to detail within a task and can compromise performance (see Ashcraft & Kirk, 2001; Maloney et al., 2014).

Effort

Despite the seeming contradiction, both lines of inquiry highlight the role of effort and implicate the importance of uncertainty. Pressure may incentivize cognitive performance by motivating individuals to devote more of their limited cognitive resources to a task than they otherwise would (i.e., exert more effort; Bonner & Sprinkle, 2002; Botvinick & Braver, 2015). This may be particularly true when individuals are uncertain whether they can attain a required outcome for an incentive. Additionally, increasing effort may mitigate any threatening consequences of worries on performance. Though worries are thought to consume WM and harm cognitive performance (Ashcraft & Kirk, 2001), highly anxious individuals might compensate for their worries by exerting more task-directed effort that is productive, rather than competitive, with task success (Hardy et al., 2007; Putwain & Symes, 2018; see Eysenck et al., 2007).

Behaviorally, increased effort exertion often manifests as greater time-on-task (Bonner & Sprinkle, 2002). For example, adults completing mathematics problems for a financial reward maintained visual attention on key problem-solving areas to a greater extent than those performing without reward (Castro et al., 2018). Pressured participants also reattempted problems more frequently than controls (Castro et al., 2018). Similarly, Konheim-Kalkstein and van den Broek (2008) found that adults demonstrated greater reading comprehension and spent more time reading when performing for monetary reward versus no reward. Attali (2016) increased pressure within subjects by raising the performance stakes—participants completed the real GRE, followed by an experimental section of the GRE quantitative or verbal reasoning sections. Test takers devoted significantly less time to the experimental sections, and this difference in effort across real and experimental formats accounted for much of the gap in scores (Attali, 2016). Indeed, researchers in educational assessment recommend using response times as an index for effort, with exceedingly quick response times indicating unmotivated, noneffortful performance (see Wise & DeMars, 2010; Wise & Kong, 2005).

Feedback

Feedback, like pressure, can alter the extent to which one demonstrates their true potential, and has a direct effect on how participants may experience the uncertainty of a task. Performance feedback reduces uncertainty and reveals gaps between one's desired and actual performance (Hattie & Timperley, 2007) and thereby may be motivating or defeating depending on the gaps' sizes. Consistent with inverted-U models, one way to optimize task performance is to provide motivating feedback. This can incentivize goal-directed effort, which is the most effective way to reduce gaps without abandoning the task (Hattie & Timperley, 2007). Feedback can facilitate metacognitive awareness and motivational approaches to learning and performance, which in turn influence performance (Hattie, 2012; Pekrun et al., 2014).

Feedback may be particularly influential when measuring WM capacity. For example, after participants completed one WM task, but prior to another, Acklin (2012) provided them with falsified negative or positive performance feedback, which hindered and boosted performance, respectively (also see Hodges & Spielberger, 1969). Others find that feedback improves WM performance (Adam & Vogel, 2016) yet hinders metacognitive awareness of one's performance (Adam & Vogel, 2018) relative to no feedback. Receiving feedback during tests or WM tasks can substantially change one's demonstrated performance and confidence.

Implications for Cognitive Behavioral Studies

While experimental and survey research has recognized the necessity of considering baseline effort while analyzing performance, the literature traditionally relies upon attention and manipulation checks (e.g., Oppenheimer et al., 2009) or outlier removal to exclude individuals who do not appear to exert sufficient effort. Yet, these measures are imperfect (see Curran, 2016) and there remains much variability in effort for those deemed sufficiently attentive. For example, participants completing a WM task reported off-task thoughts 56% of the time they were probed. Given that inattentiveness was related to poorer WM performance (Adam & Vogel, 2017), it is crucial to maximize opportunities for participants to exert effort when performing in order to produce valid measurements of cognitive or learning outcomes.

The Present Studies

We report two studies demonstrating the crucial role of social context and implications for participants' uncertainty in measurements of what are generally believed to be stable skills—WM and verbal reasoning. We specifically highlight the roles of effort and affective changes. We manipulate the stakes of the performance context through a pressure manipulation because with higher stakes, the uncertainty of one's ability to meet a threshold becomes increasingly salient.

WM may play a particularly important role in this phenomenon, as it is a resource that is integral to task performance across many domains, regardless of whether it is the intended target. Furthermore, WM has been posited as a mechanism through which pressure and social context can impact performance on other tasks: though pressure can deplete WM resources and threaten performance by inducing task-irrelevant worries (Ashcraft & Kirk, 2001), having greater available WM resources can also reduce the frequency of such distractions (Randall et al., 2014) and their effect on performance (Kane & McVay, 2012). Thus, WM is often treated as a stable individual difference ("high" and "low" WM individuals; e.g., Beilock, 2008) yet is malleable to contextual features of the task used to measure it. Our primary objective is to test how pressure and feedback in the performance context impact individuals' demonstrated performance on cognitive tasks and to examine if these effects operate through a WM pathway specifically. We included a measure of WM—the operation span task (OSPAN; Unsworth et al., 2005)—and a more multifaceted, yet challenging, performance measure—items from the GRE verbal reasoning test. We included the verbal reasoning task to distinguish specific WM-related performance changes from variability on this more general cognitive task. Moreover, verbal reasoning as assessed by the GRE is also often treated as a stable individual difference measure: here, an index of graduate school potential (Educational Testing Service [ETS], n.d.).

We manipulated pressure using a common performance-contingent social-evaluative paradigm, where one's performance determines another's receipt of a prize (Beilock et al., 2004). Before and after administering the pressure prompt, we assessed participants' performance on the WM and verbal reasoning tasks and measured their self-reported stress, anxiety, and motivation. In addition to measuring the impact of pressure on demonstrated ability, we examined differences in metacognitive judgments operationalized as participants' confidence in their performance following the tasks (Beilock et al., 2004). Finally, our mediator of interest was participants' exerted effort on the WM task, which we measured using response time during the WM task. Consistent with prior work using time-on-task to measure effort (e.g., Attali, 2016), we anticipated that slower response times would indicate greater exerted effort (Bonner & Sprinkle, 2002). We manipulated feedback on the WM task across studies. In Study 1, we provided standard administration of the OSPAN, including performance feedback after every test trial. To understand the effects of feedback on WM performance, we removed the feedback in Study 2.

Study 1

We test the effects of a pressure induction on participants' affect, metacognition, and performance. Specifically, we assessed adults' performance on a WM task and a verbal reasoning test, their confidence in their performance on the tasks, and their changes in self-reported stress, anxiety, and motivation after pressure. Furthermore, we examined changes in effort on the WM task under pressure, and whether effort mediated the pressure–performance relation.

Participants

One hundred fifty-seven participants were recruited from online research study databases or flyers across two data collection sites:

universities in Chicago, Illinois and Irvine, California. Recruitment materials advertised a cover story stating that the purpose of the study was to validate new thinking and reasoning tasks. We conducted a power analysis to determine the minimum sample size required. We used Johns et al. (2008; Study 3; $d = 0.26$) as a guide, as our original study design had a similar repeated measures design with three conditions (see Procedure section for detail on omitted condition). At $\alpha = 0.05$ and power = 0.80, we estimated the minimum sample size to be 126. With a larger sample and fewer conditions, we believe we are sufficiently powered to explore the intended main effects while also controlling for additional factors like baseline performance.

Participants were dropped from analyses for misunderstanding the task instructions ($n = 2$), experimenter error ($n = 1$), computer malfunction ($n = 5$), suspecting the other participant was a confederate ($n = 3$; pressure condition only), or failing to complete the task due to too much stress ($n = 1$), for a total of 145 participants ($n_{\text{Chicago}} = 95$; $n_{\text{Irvine}} = 50$; $M_{\text{age}} = 21.60$ years, $SD_{\text{age}} = 4.39$ years; 89 women). Participants were randomized within site to either pressure ($n = 96$) or control ($n = 49$) conditions. Within the pressure condition, half of the participants ($n = 48$) were randomly assigned to receive a positive reappraisal message, prompting them to positively reframe any feelings of stress or anxiety as advantageous for performance. The other half ($n = 48$) did not receive an additional prompt. Though similar positive reappraisal prompts have been shown to improve performance and reduce negative emotions under pressure (Johns et al., 2008), we did not find this to be the case.[1] Therefore, we collapsed across reappraisal and no reappraisal conditions to create one pressure condition ($n = 96$). Informed consent was obtained for all participants.

Procedure

The study procedure was comprised of two blocks. In Block 1, we assessed participants' baseline performance on the two tasks and their baseline affect during the tasks. Then, in Block 2, participants received pressure or control manipulations and again completed the two tasks and reported their affect. Study materials were presented using E-Prime 3.0 for Windows. Afterward, participants provided demographic information and reported their confidence in their performance.

Pressure Protocol

Modeled after prior research (e.g., Beilock et al., 2004; see Dickerson & Kemeny, 2004), the pressure manipulation was comprised of a social component (peer pressure due to shared consequences) and two evaluative components (live progress tracking by an authority and peer evaluation). Upon entering the lab, participants in the pressure condition were greeted by an experimenter and escorted to a waiting area, where a confederate was seated. The two remained seated in the waiting area for 3 min while the experimenter prepared the study materials. Then, the experimenter returned and provided instructions verbally to both confederate and participant. To bolster the later pressure manipulation, the experimenter told both individuals that they would collaborate on the second half of the task. Then, the confederate and participant were escorted to different testing rooms for the remainder of the study.

Participants first completed a 13-item Partner Questionnaire created by the researchers and intended only to bolster the confederate's role in the experiment (e.g., "When working with others, I tend to take the lead on projects"; see Supplemental Materials for complete items). Then, the researcher introduced the tasks to the participant and informed them that the researcher would monitor their progress and accuracy via an iPad application, which was shown to the participant (the "application" was, in fact, a spreadsheet). Afterward, the participant began Block 1, during which they completed an abbreviated version of the WM task, a verbal reasoning test, and were prompted to report their affective state periodically. Both tasks and all instructions were presented on a computer. Task order was counterbalanced across participants.

In Block 2, participants received pressure prompts followed by the tasks (see Table 1 for full condition manipulations). Pressure participants read a prompt stating that they had been paired with another participant (presumably the confederate) who was also completing the tasks, and that they must perform above 90% on the tasks, otherwise they and the other participant would not receive an additional $5. Participants were reminded that the researchers could monitor their task progress and accuracy. Social-evaluative threats like these have reliably been shown to both incentivize performance (e.g., Boksem et al., 2006) and induce choking under pressure (e.g., Beilock et al., 2004; see Dickerson & Kemeny, 2004). This prompt was followed by a 190 s pause, during which the screen read "Please wait a few minutes while the task configures." This pause was intended to provide adequate time for the participants to reflect on the pressure prompt and for any feelings to intensify prior to completing the tasks (Sheppes & Meiran, 2008; see Footnote 1). After the pause, half of the participants received the positive reappraisal prompt, which is not a focal point of the present study (see Footnote 1). Then, participants completed the full-length WM task and another verbal reasoning test. Task order was again counterbalanced. To prevent forgetting, brief reminders of the

[1] We manipulated a positive reappraisal prompt for participants in the pressure conditions of both Studies 1 and 2. After participants received the pressure manipulation in Block 2 of the tasks, half of the participants received the positive reappraisal prompt, which informed participants of the adaptive features of stress and anxiety, and encouraged them to view these feelings as assets for their performance on the upcoming tasks (modeled after Crum et al., 2013; Jamieson et al., 2010). The other half of participants in the pressure condition received a control prompt, which asked participants to simply reflect on any feelings of stress or anxiety they were experiencing prior to the upcoming task. Both prompts were designed to make reference to the words "stress" and "anxiety" at similar frequencies. We included positive reappraisal prompts to address the theoretical possibility that the pressure manipulation could threaten performance by increasing stress and anxiety that consumes limited cognitive resources, yet, we did not find this to be the case: in Studies 1 and 2, the pressure and reappraisal conditions had no differences in their changes in self-reported anxiety, Study 1: $t(94) = -0.66$, $p = .51$ and Study 2: $t(135) = 0.17$, $p = .87$; stress, Study 1: $t(94) = -0.76$, $p = .45$ and Study 2: $t(135) = 0.65$, $p = .52$; or motivation, Study 1: $t(94) = 0.22$, $p = .82$ and Study 2: $t(135) = 1.62$, $p = .11$. Moreover, there were no difference in their overall average performance on the WM task, Study 1: $F(1, 93) = 0.69$, $p = .41$; Study 2: $F(1, 134) = 1.23$, $p = .27$, or the verbal reasoning task, Study 1: $F(1, 93) = 1.14$, $p = .29$; Study 2: $F(1, 134) = 0.34$, $p = .56$, conditional on Block 1 performance. The conditions had equal confidence in their performance on the tasks, conditional on their actual performance, Study 1: $F(1, 93) = 1.25$, $p = .27$; Study 2: $F(1, 134) = 0.51$, $p = .48$. In Study 2, we also administered a manipulation check, asking participants to report how often they tried to "Think about how your stress could actually help your performance" during the tasks (1 = *Never*, 5 = *Sometimes*, 9 = *Always*). There was no difference between pressure and reappraisal conditions, $t(104) = 1.30$, $p = .20$.

Table 1
Full Condition Manipulations and Reminder Prompts

Message length	Pressure	Control
Full prompt	You have been randomly paired with another participant who is completing the same tasks. If you both score above 90% on the two reasoning tasks, you will both receive an additional $5 for participating in our study. However, if even one of you scores below 90% on either task, neither of you will receive the $5. Thus, it is important that you make as few mistakes as possible on the following two reasoning tasks.	Please pay close attention to the instructions provided on the screen and on your response packet. Remember, when you are done reading the instructions on the screen, you must click the mouse to continue to the next screen.
Reminder	Remember, you must keep a score of 90% or higher in order to ensure that you and the other participant receive the additional $5.	Remember, when you are done reading the instructions on the screen, you must click the mouse to continue to the next screen.

Note. Full prompts are provided at the beginning of Block 2. Reminders are provided immediately before each of the two tasks in Block 2.

pressure prompt were provided twice, each time occurring after the instructions of a task but before the test trials began.

Upon completion of Block 2, the experimenter returned, provided the participant with a demographic questionnaire, and assessed their confidence in their performance overall. Afterward, all participants were debriefed and received the additional $5, regardless of their performance.

Control Protocol

We used an identical procedure for the control group, with three exceptions. First, there was no confederate in the waiting area. Second, the control group was not administered the Partner Questionnaire. Finally, participants in the control were not informed of a performance criterion nor were they informed that the researchers would monitor their progress and accuracy. Rather, they were simply asked to continue paying attention (see Table 1).

Measures

Working Memory

WM capacity was assessed twice for each participant. In Block 1, we used an abbreviated version of the OSPAN to assess baseline WM capacity (Foster et al., 2015). The WM test trials are comprised of a maintenance period, in which participants must maintain letters in memory while attending to distracting information (whether an arithmetic solution is true or false), and a recall period, in which participants must recall the order of the presented letters. Participants completed five test trials which varied in length from three to seven letters to be memorized, for a total possible score of 25 letters correctly memorized in order. Test trials of differing set sizes were presented in random order, and all participants completed all set sizes. In Block 2, we assessed WM capacity with the full-length version of the OSPAN (Unsworth et al., 2005), which is three times longer than the abbreviated OSPAN, for a total possible score of 75 letters correctly memorized in order. WM capacity was measured as the proportion correct out of 25 (Block 1) or 75 (Block 2). Importantly, participants were unaware as to which parts of the task were distracting. As far as those under pressure knew, performance on both the maintenance and recall periods was evaluated as part of the 90% criteria for reward.

Performance feedback was provided on a trial-by-trial basis for all test trials per the standard administration of the OSPAN (see Unsworth et al., 2005, for details). Upon completion of each test trial, participants saw a white screen which summarized their accuracy on the true/false arithmetic items and the letters they just recalled. For example, a person who failed to recall two letters in the correct order and incorrectly answered three true/false items during a test trial of set size seven would see: “You got 5 out of 7 letters correct. You made 3 math error(s) for this set of trials.” Feedback for each trial was displayed for 2 s, after which the next test trial would begin. Thus, participants received this performance feedback five times during the Block 1 WM task and 15 times during Block 2.

Effort

Response time is a commonly used measure of exerted effort under pressure (e.g., Attali, 2016; Castro et al., 2018). The reaction times during maintenance periods of the OSPAN test trials—when participants held letters in mind while completing the true/false arithmetic calculations—were extracted and averaged as a measure of participants’ effort. Prior work has shown that the facilitative effects of pressure on WM are greatest during these maintenance periods of WM tasks (e.g., Zedelius et al., 2011). Effort on the WM task during Blocks 1 and 2 was measured as the average amount of time (in seconds) participants spent evaluating each arithmetic solution, from the onset of the solution displays to the time the participant advanced to the next screen.

Verbal Reasoning Test

Performance on a verbal reasoning test was also assessed at Block 1 and Block 2. We used sentence equivalence questions from the verbal reasoning section of a GRE practice software (Magoosh: https://gre.magoosh.com) to contrast changes in available WM capacity specifically with performance changes on tasks assessing multiple facets of cognition under pressure. For each forced-choice verbal reasoning question, participants were provided a sentence containing two blanks and were asked to choose the two appropriate words (of six possible choices) to fill the blank. These verbal reasoning questions consist of recall and inferencing processes: The test requires participants to draw from their vocabulary to select the most appropriate words, interpret the context of the sentence with and without the chosen words, and align and integrate the chosen words into the sentence. Thus, these questions are not WM independent, yet their reliance on prior knowledge means that performance is less reliant on WM capacity than the OSPAN, allowing us to contrast changes in WM with changes in test performance.

We made two sets of verbal reasoning questions for Blocks 1 and 2 and counterbalanced the order across participants. Order was unrelated to performance at Block 1, $t(143) = -1.58$, $p = .11$, and Block 2, $t(143) = -1.21$, $p = .23$. For each set, we chose four questions labeled as easy, four medium, five hard, and five very hard for a total of 18 questions per set. Participants received 1 point for every question, for a total possible score of 18. Partial credit (0.5 points) was provided. Verbal reasoning performance at Blocks 1 and 2 was measured as the proportion correct out of 18. Performance feedback was not provided. We did not assess response time for the verbal reasoning test.

Affect

Eleven times across Blocks 1 and 2, participants were prompted to report on the extent to which they were experiencing each of three affective states: stressed, anxious, and motivated. For each state, they were asked to use a 9-point Likert scale (1 = *Not at all anxious*; 5 = *Moderately anxious*; 9 = *Highly anxious*; Sheppes & Meiran, 2008) to "please circle the value corresponding to the extent to which you are experiencing the word right now, in this moment." Baseline anxiety, stress, and motivation are measured as the average value the participant reported for the six affect items in Block 1 before the condition manipulations, and postmanipulation scores are the average value for the five affect items occurring after the pressure manipulation. In analyses, we report change in affect as the difference between Block 2 and Block 1 averages.

Confidence

Upon conclusion of the study, we assessed participants' confidence using the same 9-point Likert scale as above (1 = *Not at all confident*; 5 = *Moderately confident*; 9 = *Highly confident*). Participants were asked to use the scale to report the extent to which they agreed with the statement: "I think I did well on the tasks."

Results

Analytic Plan

Our research questions center on the impacts of uncertainty (here, manipulated via pressure) on performance, affective, and metacognitive outcomes. First, we test main effects of the pressure manipulation on demonstrated WM capacity, and we test whether increased effort mediates the impact of pressure on WM performance. Second, we explore the effect of pressure on verbal reasoning performance. Third, we test the impact of pressure on participants' changes in self-reported affect, followed by their confidence judgments. We use regressions to explore our primary analyses of the impact of our pressure manipulation (binary coded: 0 if control, 1 if pressure) on all Block 2 performance and affective outcomes controlling for Block 1 values. School is included as a covariate in all regressions, as differences in baseline performance and affect emerged across sites.[2] Standardized β coefficients and standard errors are reported for all regressions. Counterbalanced ordering of the tasks did not predict performance.[3]

WM Performance

WM Capacity. Descriptive statistics for all measures are provided in Table 2. Overall, participants performed fairly well on the WM task: They recalled 84% ($SD = 0.14$; range = 0.28, 1.00) of the items in the correct order at baseline, which did not differ by conditions ($\beta = 0.29$, $SE = 0.18$, $p = .10$). The WM score increased to 87% ($SD = 0.12$; range = 0.24, 1.00) at Block 2. A regression revealed a main effect of condition on participants' overall WM capacity at Block 2 ($\beta = 0.49$, $SE = 0.15$, $p = .001$). Those in the pressure condition ($M = 0.89$, $SE = 0.01$) had significantly higher WM capacity scores at Block 2 than the control ($M = 0.83$, $SE = 0.01$).

We next explored whether the effects of pressure in Block 2 varied across the level of trial difficulty. WM task trials varied in length from three to seven items to remember. We first ran a repeated measure analysis of variance (ANOVA) to test the interaction of condition (between-subjects) and trial set size (within-subjects) on participants' Block 2 WM performance. The statistically significant interaction confirmed that the effect of pressure varied by trial difficulty, $F(4, 527) = 7.63$, $p < .001$, $\eta_p^2 = 0.05$. We next ran post hoc regressions for each of the set sizes, conditioning on their Block 1 performance at each set size. Regressions were not corrected for multiple comparisons. The impacts of pressure were focused on the higher difficulty items (see Figure 1). There were no effects of pressure on trials of Set size 3 ($\beta = 0.09$, $SE = 0.18$, $p = .62$) or Set size 4 ($\beta = 0.10$, $SE = 0.19$, $p = .58$). Conversely, the pressure group had significantly better performance than the control group on trials of Set size 5 ($\beta = 0.50$, $SE = 0.18$, $p = .006$), Set size 6 ($\beta = 0.56$, $SE = 0.18$, $p = .002$), and Set size 7 ($\beta = 0.59$, $SE = 0.18$, $p = .001$).

Effort. We use participants' average response time during the maintenance periods of the WM task to assess exerted effort. Participants performed well on this component of the WM task, correctly evaluating 93% ($SD = 6\%$) of the arithmetic solutions at Block 1. Pressure and control conditions did not differ in their arithmetic accuracy ($\beta = 0.23$, $SE = 0.18$, $p = .19$) nor did they differ in their average response time on these items at Block 1 ($\beta = 0.001$, $SE = 0.17$, $p = .99$; see Table 2 for means). Moreover, response time did not predict arithmetic accuracy ($\beta = 0.05$, $SE = 0.09$, $p = .58$), though it did predict WM capacity ($\beta = -0.24$, $SE = 0.08$, $p = .004$)—those who were faster at evaluating the arithmetic solution had lower recall performance at Block 1.

In Block 2, effects of pressure emerged. The pressure group had higher accuracy ($\beta = 0.33$, $SE = 0.16$, $p = .04$) and had a slower

[2] We examined differences between schools in their performance and affective responses, as one school is an elite private institution and the other a public state institution. At baseline, Study 1 participants at Chicago had lower stress ($\beta = -0.20$, $SE = 0.08$, $p = .01$), higher WM ($\beta = 0.18$, $SE = 0.08$, $p = .03$), and higher verbal reasoning ($\beta = 0.50$, $SE = 0.07$, $p < .001$) compared to Irvine participants. In Study 2, Chicago also had higher motivation ($\beta = 0.18$, $SE = 0.06$, $p = .007$) and higher verbal reasoning ($\beta = 0.97$, $SE = 0.12$, $p < .001$) compared to Irvine at baseline. Thus, we control for school in all analyses. When split by school, the relationship between pressure and performance on both tasks remained consistent with the overall analysis, though we lose power due to the smaller sample sizes.

[3] We counterbalanced the order of the WM task and the verbal reasoning test between students. In Study 1, we found no effect of task order on participants' Block 1 WM performance ($\beta = 0.19$, $SE = 0.17$, $p = .25$) or Block 1 verbal reasoning performance ($\beta = -0.02$, $SE = 0.17$, $p = .89$). Conditional on Block 1 performance, we also found no effect of task order on Block 2 WM performance ($\beta = -0.19$, $SE = 0.15$, $p = .25$) or Block 2 verbal reasoning performance ($\beta = -0.10$, $SE = 0.11$, $p = .40$). The same was true for Study 2: WM performance (Block 1: $\beta = 0.01$, $SE = 0.14$, $p = .95$; Block 2: $\beta = -0.12$, $SE = 0.10$, $p = .25$), verbal reasoning performance (Block 1: $\beta = -0.92$, $SE = 0.14$, $p = .53$; Block 2: $\beta = -0.07$, $SE = 0.09$, $p = .38$). Thus, we collapse across counterbalanced groups in all analyses.

Table 2

Study 1 Participants' Average Affective Response, WM Performance and Response Time (RT), Verbal Reasoning Performance, and Confidence at Blocks 1 and 2, M (SD)

Measure	Control (n = 49)		Pressure (n = 96)		Overall (N = 145)	
	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2
Affect self-report (out of 9)						
Anxiety	3.43 (1.83)	3.25 (2.00)	3.53 (1.67)	3.82 (1.87)	3.49 (1.72)	3.63 (1.93)
Stress	3.63 (1.80)	3.40 (1.93)	4.07 (1.86)	4.28 (1.91)	3.92 (1.84)	3.98 (1.96)
Motivation	4.89 (2.00)	4.28 (2.09)	4.86 (1.75)	4.84 (1.93)	4.87 (1.83)	4.65 (2.00)
WM						
RT during maintenance (effort, in seconds)	2.19 (0.89)	1.70 (0.76)	2.17 (0.83)	1.88 (0.78)	2.18 (0.85)	1.82 (0.78)
Accuracy during maintenance (% correct)	0.92 (0.08)	0.91 (0.08)	0.93 (0.06)	0.94 (0.05)	0.93 (0.07)	0.93 (0.06)
WM capacity						
Set size 3	2.84 (0.56)	2.86 (0.36)	2.91 (0.42)	2.90 (0.25)	2.89 (0.47)	2.89 (0.29)
Set size 4	3.58 (0.97)	3.73 (0.53)	3.83 (0.48)	3.80 (0.46)	3.75 (0.69)	3.78 (0.48)
Set size 5	4.42 (1.25)	4.34 (0.92)	4.38 (1.31)	4.73 (0.52)	4.40 (1.29)	4.60 (0.70)
Set size 6	4.09 (2.16)	4.66 (1.34)	4.90 (1.62)	5.36 (0.71)	4.63 (1.85)	5.13 (1.02)
Set size 7	4.82 (2.12)	4.67 (1.65)	5.36 (1.88)	5.63 (1.18)	5.18 (1.98)	5.31 (1.42)
Overall % correct	0.81 (0.17)	0.82 (0.16)	0.85 (0.13)	0.90 (0.09)	0.84 (0.14)	0.87 (0.12)
Verbal reasoning performance						
Easy (out of 4)	3.15 (0.89)	3.04 (0.95)	3.34 (0.88)	3.50 (0.70)	3.28 (0.88)	3.34 (0.82)
Medium (out of 4)	2.48 (1.10)	2.35 (1.08)	2.55 (0.95)	2.82 (0.86)	2.52 (1.00)	2.66 (0.97)
Hard (out of 5)	2.42 (1.17)	2.32 (1.13)	2.74 (1.02)	2.78 (1.02)	2.63 (1.08)	2.62 (1.08)
Very hard (out of 5)	1.94 (0.93)	2.07 (0.99)	2.10 (0.93)	2.07 (0.89)	2.04 (0.93)	2.07 (0.92)
Overall % correct	0.56 (0.18)	0.54 (0.19)	0.60 (0.16)	0.62 (0.15)	0.59 (0.17)	0.59 (0.17)
Confidence self-report (out of 9)	—	4.63 (1.93)	—	4.57 (1.93)	—	4.59 (1.92)

Note. WM = working memory.

average response time (β = 0.24, SE = 0.10, p = .02) on the arithmetic items compared to the control group. In Block 2, response time predicted higher accuracy on the arithmetic items (β = 0.62, SE = 0.12, $p < .001$) and higher WM capacity (β = 0.37, SE = 0.12, p = .003): Those who spent more time evaluating the arithmetic items had higher accuracy on all parts of the WM task.

It could be the case that faster reaction times came at a cost to participants' accuracy. To address this, we reran this analysis using average response time on only the arithmetic items for which the participants were accurate. All results held,[4] suggesting the pressure condition's gains, were not due to a speed–accuracy trade-off. Rather, pressure seemed to motivate participants to take their time, which was related to better performance on the WM task.

Mediation. Given the relations between pressure, effort, and WM performance, in conjunction with theory linking performance pressure gains to effort, we next explored the mediational role of increased effort on WM capacity gains. Using Hayes (2013) PROCESS model (Model 4; 5,000 bootstrapped samples), we tested whether pressure (independent variable) increased WM capacity (dependent variable) through increased effort (response time in seconds; mediator). We conditioned our mediation analysis on participants' school and their Block 1 WM capacity and effort. We found a direct effect of pressure on participants' WM capacity (b = 0.06, SE = 0.02, p = .001) and a main effect of pressure on effort (b = 0.19, SE = 0.08, p = .02). Moreover, effort was a significant predictor of WM capacity (b = 0.05, SE = 0.02, p = .02). Critically, we found a significant indirect effect of effort, b = 0.009, SE = 0.006, 95% CI [0.0004, 0.0231]. After accounting for effort, the direct effect of pressure on WM capacity decreased slightly but remained significant (b = 0.05, SE = 0.02, p = .007), indicating a partial mediation. Pressure increased participants' WM capacity via increased effort on the task.

Verbal Reasoning Performance

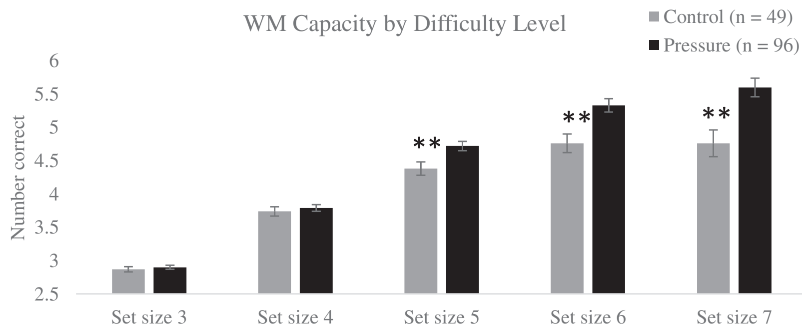
We next examine participants' performance on a relatively less WM demanding yet academically challenging verbal reasoning test. Descriptive statistics are provided in Table 2. No effects of condition emerged in Block 1 performance (β = 0.22, SE = 0.15, p = .16). However, in Block 2, the pressure group had higher accuracy than control (β = 0.29, SE = 0.12, p = .01).

Again, a repeated measures ANOVA revealed a statistically significant interaction between condition (between-subjects) and question difficulty level (within-subjects) on participants' Block 2 verbal reasoning performance, $F(3, 429) = 4.61$, p = .004, η_p^2 = 0.03. We ran separate post hoc regressions at each difficulty level. Unlike the WM patterns, verbal reasoning gains from pressure were focused on the easier items (Figure 2). The pressure group had significantly higher accuracy on the easy (β = 0.43, SE = 0.13, p = .001) and medium (β = 0.45, SE = 0.15, p = .003) items. There were comparably smaller and statistically nonsignificant effects of pressure for the hard (β = 0.28, SE = 0.14, p = .05) and very hard (β = −0.06, SE = 0.16, p = .72) items.

[4] We reran the analyses of effort after reducing the response time measure to include only the trials on which the participant was accurate (Study 1: $M_{\text{Block 1}} = 2.37$, $SD_{\text{Block 1}} = 0.97$; $M_{\text{Block 2}} = 1.96$, $SD_{\text{Block 2}} = 0.82$ and Study 2: $M_{\text{Block 1}} = 2.69$, $SD_{\text{Block 1}} = 1.22$; $M_{\text{Block 2}} = 2.31$, $SD_{\text{Block 2}} = 1.22$). In both studies, the pressure group had longer response time on Block 2, conditional on their Block 1 response time (Study 1: β = 0.25, SE = 0.10, p = .02; Study 2: β = 0.19, SE = 0.19, p = .02). Block 2 response time predicted higher accuracy on the arithmetic items (Study 1: β = 0.39, SE = 0.14, p = .005; Study 2: β = 0.34, SE = 0.11, p = .002) and on the WM task (Study 1: β = 0.30, SE = 0.14, p = .03; Study 2: β = 0.26, SE = 0.09, p = .003) conditional on their Block 1 performance and response time. All results were consistent with the original measure of effort across both studies.

Figure 1

Study 1 Participants' Average Block 2 WM Capacity at Each Set Size of the WM Task, Conditional on School and Block 1 Performance



Note. Error bars are ±1 *SE*. WM = working memory.
$^{**}p < .01$.

Affect

Regression analyses revealed no effect of condition on participants' average anxiety (β = 0.06, *SE* = 0.18, *p* = .71), stress (β = 0.25, *SE* = 0.17, *p* = .14), or motivation (β = −0.03, *SE* = 0.18, *p* = .87) at Block 1. However, at Block 2, those in the pressure condition reported higher average anxiety (β = 0.26, *SE* = 0.07, *p* = .001), stress (β = 0.26, *SE* = 0.09, *p* = .004), and motivation (β = 0.29, *SE* = 0.08, *p* < .001) compared to controls and conditional on their school and Block 1 averages. See Table 2 for descriptive statistics.

For illustrative purposes, Figure 3 depicts participants' change in average self-reported affect from Block 1 to Block 2. On average, participants in the pressure condition saw increases in self-reported anxiety and stress, whereas those in the control condition saw decreases in these experiences. Conversely, the control group saw a drastic, significant decrease in their motivation over the course of the study. Pressure seemed to buffer from this loss in motivation, as the pressure group had changes in motivation that were not different from zero.

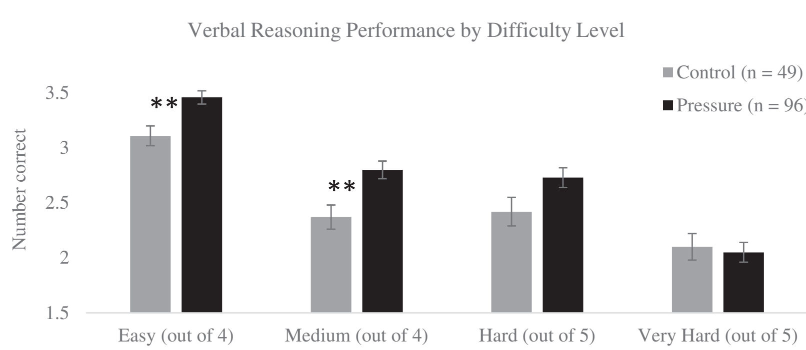
Relations Between Affect and Performance. We use partial correlations to test the relations between affect and performance at Block 2, conditional on Block 1 affect. Overall, self-reports of affect were positively correlated with each other, with the exception of changes in anxiety and motivation (see Table 3). Motivation was positively correlated with participants' overall performance on both the WM task, *r*(142) = 0.18, *p* = .03, and the verbal reasoning test, *r*(142) = 0.33, *p* < .001. Self-reported anxiety and stress had comparably smaller and statistically nonsignificant relations to task performance, suggesting motivation was particularly important to performance.

Confidence

Finally, we ask whether participants' confidence in their performance, like their actual performance, was also influenced by pressure. We compared participants' response to the item "I think I did well on the tasks," while controlling for their school and overall performance on the tasks. Overall performance here was calculated as the average of their WM capacity and verbal reasoning performance across both blocks (*M* = 0.79, *SD* = 0.10; range = 0.29, 0.95). We found no difference between pressure and control in their confidence (β = −0.09, *SE* = 0.19, *p* = .64), despite the fact that the

Figure 2

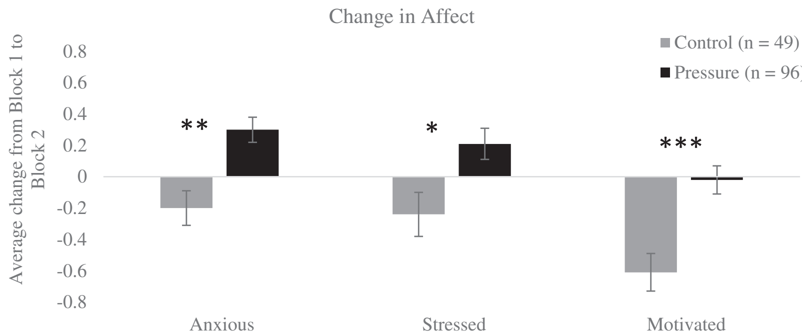
Study 1 Participants' Average Accuracy for the Easy, Medium, Hard, and Very Hard Verbal Reasoning Items, Conditional on School and Block 1 Performance



Note. Error bars are ±1 *SE*.
$^{**}p < .01$.

Figure 3

Study 1 Participants' Average Change in Affect From Block 1 to Block 2, Conditional on School



Note. Error bars are ±1 *SE*.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

pressure group actually outperformed the control on both tasks. In fact, results from a regression model simultaneously considering participants' performance, condition, and their interaction revealed that participants' confidence was unrelated to their overall performance ($\beta = -0.01$, $SE = 0.12$, $p = .92$). Additionally, condition did not interact with overall performance to predict confidence ($\beta = 0.25$, $SE = 0.18$, $p = .17$).

Discussion

Across two different higher level cognitive tasks generally used as stable measures of individual differences (e.g., Beilock, 2008; ETS, n.d.), we found that pressure predicted increased performance. WM gains under pressure were partially mediated by effort, consistent with theories of pressure as an incentive to deploy cognitive resources (Botvinick & Braver, 2015). This suggests that without pressure, participants would not have demonstrated this level of WM capacity. Their self-reported motivation supported this interpretation: Those under pressure reported higher motivation on average and bypassed the controls' significant loss in motivation, which predicted performance on both tasks. Self-reported anxiety and stress were also higher among participants with higher pressure and uncertainty (although pressure did not influence participants' confidence in their performance). Though these changes did not predict task performance and thus do not suggest a verbal WM pathway as some have suggested (Beilock, 2008), there may be affective consequences for pressure and uncertainty that are not evidenced in performance but rather through ultimate task or field persistence.

Consistent with prior work (Heitz et al., 2008; Pochon et al., 2002), pressure-related gains in performance were focused on the most difficult trials of the WM task, but the easier items of the verbal reasoning task. These differences further support the role of effort in performance changes under pressure. Bonner and Sprinkle's (2002) framework posits that, to the extent that task performance can change with the deployment of additional cognitive resources, one could see increases in task performance under pressure. Thus, performance at all levels of the WM task could be improved with additional effort. However, for the verbal reasoning test, where item difficulty is partially a function of prior knowledge, additional effort cannot produce changes in performance for more difficult items requiring advanced vocabulary. Accordingly, we found the easier verbal reasoning items were more responsive to effort (see also Kiplinger & Linn, 1995; O'Neil et al., 1995). Verbal reasoning scores were relatively normally distributed at every level. However, because most participants received perfect WM scores on Set sizes 3 and 4, it is possible that effort-induced WM performance gains on the easiest trials were constrained by ceiling effects (see Heitz et al., 2008).

Table 3

Partial Correlations Between Study 1 Participants' Self-Reported Affect and Performance at Block 2, Conditional on Block 1 Self-Reported Affect

Measure	1	2	3	4	5
1. Anxiety	1				
2. Stress	0.64***	1			
3. Motivation	0.06	0.24**	1		
4. WM—overall	−0.02	−0.03	0.18*	1	
5. Verbal reasoning—overall	−0.15	0.07	0.33***	0.27***	1

Note. WM = working memory.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Study 2

In Study 1, we find evidence to suggest pressure can incentivize participants to increase their effort and demonstrate greater cognitive performance than they otherwise would. In Study 2, we explored whether further increasing uncertainty by removing another contextual factor—feedback—may also influence performance and affect.

Feedback signals the discrepancy between one's goal and one's current performance (see Hattie & Timperley, 2007). If one receives negative feedback or no feedback, signaling a potential gap between current and desired performance, then it may induce uncertainty about one's ability to perform. Receiving positive feedback may assuage feelings of uncertainty, indicating a more achievable goal.

The extent to which feedback catalyzes effort expenditure may depend on whether the feedback leads one to believe success is achievable (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Informative feedback can alter WM performance (Acklin, 2012; Adam & Vogel, 2016, 2018) and metacognitive awareness of one's performance (Adam & Vogel, 2018).

In Study 1, we assumed that the pressure to perform above 90% on the tasks would heighten uncertainty. Given that 39% of participants in Study 1 received a score above 90% on the Block 1 WM task, perhaps many felt certain in their ability to achieve the 90% goal on Block 2. Put differently, pressure may have improved performance because participants knew there was little discrepancy between their current and goal performance. Anecdotal insight from debriefing conversations suggested that Study 1 participants were more certain of their WM performance, because it provided trial-by-trial feedback, than they were for verbal reasoning, which provided no feedback.

To elucidate the conditions under which participants best perform, we ran an experiment identical to Study 1 with one key modification: We further heightened uncertainty in the performance context by removing the trial-by-trial feedback participants received during the WM task. Study 2 participants would not receive any performance feedback.

Participants

In accordance with the power analysis conducted in Study 1, 230 participants were recruited from Chicago and Irvine sites.[5] We recruited from the same online research study databases and flyers, and under the same cover story, as Study 1. Participants were dropped from analyses for misunderstanding the task instructions ($n = 2$), experimenter error ($n = 2$), computer malfunction ($n = 4$), or suspecting the other participant was a confederate ($n = 5$; pressure condition only), for a total of 217 participants ($n_{\text{Chicago}} = 101$; $n_{\text{Irvine}} = 116$; $M_{\text{age}} = 22.29$ years, $SD_{\text{age}} = 5.29$ years; 168 women). Again, we had originally assigned participants to one of three conditions (pressure, pressure + reappraisal, control). However, we again found no differences between the pressure condition and the pressure plus reappraisal condition on any of our measures (see Footnote 1), so we collapsed across these two conditions to create one pressure condition ($n = 137$) and a control condition ($n = 80$). All participants provided informed consent.

Procedure

Procedures and measures were identical to Study 1, with one exception. For both the abbreviated OSPAN (Block 1) and full-length OSPAN (Block 2) WM tasks, we removed the feedback provided during the test trials. Typically, after each test trial, the task provides accuracy feedback for the recall items and the arithmetic solution items of the WM test trial (e.g., "You got 5 out of 7 letters correct. You made 3 math error(s) for this set of trials."). We removed these feedback statements. Instead, after each test trial in Blocks 1 and 2, participants would see a blank screen displayed for the same duration. Feedback provided during the practice trials were not removed. No other changes were made to the tasks. As in Study 1, feedback was not provided on the verbal reasoning test.

Results

WM Performance

WM Capacity. The analytical plan is identical to that used in Study 1. Again, counterbalanced ordering of the tasks did not predict performance (see Footnote 3). Descriptive statistics for all outcomes are provided in Table 4. At Block 1, there were no differences in WM performance by condition ($\beta = 0.22$, $SE = 0.14$, $p = .11$). We found that the pressure condition had somewhat higher Block 2 WM capacity compared to control; however, this effect was less than half the magnitude of Study 1 and was no longer statistically significant ($\beta = 0.22$, $SE = 0.11$, $p = .05$). Moreover, the effect of pressure did not interact with trial set size to predict WM performance, $F(4, 848) = 0.46$, $p = .77$.

Effort. We again used average response time on the arithmetic items to examine exerted effort. In Block 1, there were no differences in average response time ($\beta = -0.09$, $SE = 0.14$, $p = .56$) between pressure and control, though the pressure group had higher accuracy on the arithmetic items at Block 1 ($\beta = 0.32$, $SE = 0.14$, $p = .03$). Response time did not predict arithmetic accuracy ($\beta = 0.11$, $SE = 0.07$, $p = .12$), but it did predict overall WM performance ($\beta = -0.23$, $SE = 0.07$, $p = .001$). Again, those with slower arithmetic speed performed better on the WM task.

At Block 2, the pressure group had somewhat slower response times on average than control ($\beta = 0.18$, $SE = 0.09$, $p = .05$), though this effect was again smaller in magnitude compared to Study 1 and was not statistically significant. Moreover, differences by condition in response time did not correspond to differences in arithmetic accuracy at Block 2 ($\beta = 0.23$, $SE = 0.14$, $p = .11$). Response time at Block 2 predicted arithmetic accuracy ($\beta = 0.69$, $SE = 0.09$, $p < .001$) and overall WM capacity ($\beta = 0.26$, $SE = 0.08$, $p = .001$). There was no evidence for a speed–accuracy trade-off (see Footnote 4).

Verbal Reasoning Performance

Pressure and control participants performed equally well on the test at Block 1 ($\beta = -0.08$, $SE = 0.14$, $p = .56$) and Block 2 ($\beta = 0.05$, $SE = 0.09$, $p = .60$). There was no interaction between condition and item difficulty on verbal reasoning performance, $F(3, 645) = 0.61$, $p = .61$, again suggesting that the pressure group performed equally to the control at all difficulty levels (see Table 4 for means).

Affect

At Block 1, regression analyses revealed no effect of condition on participants' self-reported levels of anxiety ($\beta = -0.02$, $SE = 0.14$, $p = .87$), stress ($\beta = -0.10$, $SE = 0.14$, $p = .51$), and motivation ($\beta = -0.07$, $SE = 0.14$, $p = .63$). At Block 2, however, those in the pressure condition had significantly higher self-reported anxiety ($\beta = 0.19$, $SE = 0.08$, $p = .01$), stress ($\beta = 0.23$, $SE = 0.07$, $p = .001$), and motivation ($\beta = 0.29$, $SE = 0.07$, $p < .001$) controlling for Block 1

[5] During Spring 2020, we were in the process of recruiting more participants for Study 1 to match our Study 2 sample size. However, due to the onset of the coronavirus disease (COVID-19) pandemic, we had to halt our data collection prematurely. Thus, the sample sizes for Studies 1 and 2 both met our power threshold, however, our end-of-quarter stopping rule led to different sizes.

Table 4
Study 2 Participants' Average Affective Response, WM Performance and Response Time (RT), Verbal Reasoning Performance, and Confidence at Blocks 1 and 2, M (SD)

Measure	Control (*n* = 80)		Pressure (*n* = 137)		Overall (*N* = 217)	
	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2
Affect self-report (out of 9)						
Anxiety	3.63 (1.90)	3.61 (2.13)	3.54 (1.73)	3.96 (1.98)	3.57 (1.79)	3.82 (2.04)
Stress	3.94 (2.11)	3.93 (2.18)	3.71 (2.02)	4.30 (2.07)	3.80 (2.05)	4.16 (2.11)
Motivation	5.28 (1.83)	4.63 (2.05)	5.01 (1.77)	4.90 (1.98)	5.11 (1.80)	4.80 (2.01)
WM						
RT during maintenance (effort, in seconds)	2.52 (1.11)	2.00 (1.01)	2.48 (1.18)	2.20 (1.21)	2.50 (1.15)	2.13 (1.14)
Accuracy during maintenance (% correct)	0.91 (0.09)	0.90 (0.09)	0.93 (0.06)	0.92 (0.09)	0.92 (0.07)	0.91 (0.09)
WM capacity						
Set size 3	2.85 (0.51)	2.83 (0.48)	2.88 (0.44)	2.88 (0.34)	2.87 (0.46)	2.86 (0.40)
Set size 4	3.65 (0.89)	3.68 (0.63)	3.73 (0.76)	3.82 (0.45)	3.70 (0.81)	3.77 (0.53)
Set size 5	4.33 (1.39)	4.26 (1.11)	4.46 (1.31)	4.49 (0.82)	4.41 (1.34)	4.40 (0.94)
Set size 6	4.39 (1.95)	4.78 (1.40)	4.83 (1.65)	5.07 (1.00)	4.67 (1.77)	4.97 (1.17)
Set size 7	4.74 (2.02)	4.62 (1.66)	4.93 (2.01)	4.97 (1.55)	4.86 (2.01)	4.84 (1.60)
Overall % correct	0.80 (0.18)	0.81 (0.17)	0.83 (0.16)	0.85 (0.11)	0.82 (0.16)	0.84 (0.14)
Verbal reasoning performance						
Easy (out of 4)	3.27 (0.92)	3.22 (0.94)	3.32 (0.80)	3.27 (0.92)	3.30 (0.84)	3.25 (0.93)
Medium (out of 4)	2.65 (0.99)	2.63 (0.96)	2.42 (0.90)	2.55 (0.83)	2.51 (0.94)	2.58 (0.88)
Hard (out of 5)	2.43 (1.10)	2.62 (1.02)	2.61 (1.40)	2.52 (1.07)	2.54 (1.30)	2.56 (1.05)
Very hard (out of 5)	2.12 (1.08)	2.22 (0.96)	1.93 (0.88)	2.08 (1.08)	2.00 (0.96)	2.13 (1.04)
Overall % correct	0.59 (0.18)	0.59 (0.17)	0.57 (0.15)	0.58 (0.16)	0.58 (0.16)	0.58 (0.17)
Confidence self-report (out of 9)	—	4.98 (1.84)	—	4.26 (1.72)	—	4.52 (1.80)

Note. WM = working memory.

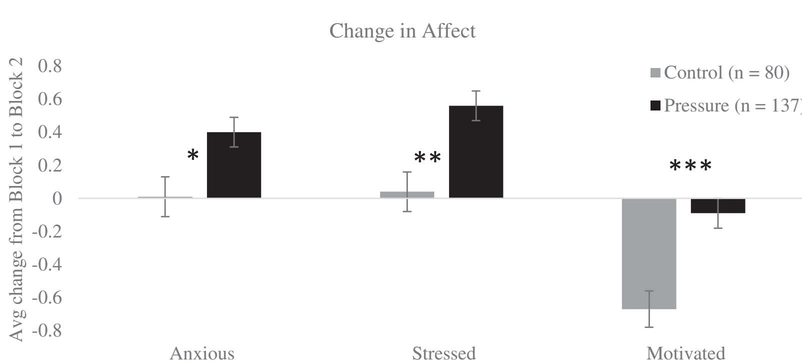
levels. Figure 4 depicts the change in average affect from Block 1 to Block 2 for illustrative purposes. Again, pressure increased participants' stress and anxiety on average and buffered participants from a substantial loss in motivation experienced by those in control.

Relations Between Affect and Performance. Table 5 displays the partial correlation coefficients of Block 2 affective responses and WM and verbal reasoning performance, controlling for Block 1 affective responses. Only self-reported anxiety and stress were positively related to each other, $r(214) = 0.78$, $p < .001$. Self-reported motivation was not related to anxiety or stress, but it did significantly positively correlate with participants' performance on the verbal reasoning test, $r(214) = 0.20$, $p = .003$, and, to a lesser degree, their performance on the WM task, $r(214) = 0.12$, $p = .08$. There were no significant relations to anxiety and stress on WM performance.

Confidence

Conditional on their school and overall performance on the WM and verbal reasoning tasks across both blocks ($M = 0.77$, $SD = 0.12$; range = 0.20, 0.96), the pressure group had lower self-reported confidence in their performance than control (β = −0.39, $SE = 0.14$, $p = .005$). This again was despite the fact that the pressure group performed equally to the control group on both tasks and at both blocks. Again, participants' confidence was unrelated to their overall performance (β = 0.18, $SE = 0.09$, $p = .06$) and condition

Figure 4
Study 2 Participants' Average Change in Affect From Block 1 to Block 2, Conditional on School



Note. Error bars are ±1 *SE*.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5
Partial Correlations Between Study 2 Participants' Self-Reported Affect and Performance at Block 2, Conditional on Block 1 Self-reported Affect

Measure	1	2	3	4	5
1. Anxiety	1				
2. Stress	0.78***	1			
3. Motivation	0.06	0.09	1		
4. WM—overall	−0.001	−0.01	0.12	1	
5. Verbal reasoning—overall	−0.15*	−0.12	0.20**	0.33***	1

Note. WM = working memory.
* $p < .05$. ** $p < .01$. *** $p < .001$.

did not interact with overall performance to predict confidence ($\beta = 0.09$, $SE = 0.13$, $p = .48$).

Discussion

In Study 2, we heightened uncertainty during performance by removing all feedback, replicating most Study 1 results and yielding larger magnitudes in affective changes. Like Study 1, those in the pressure condition reported greater anxiety and stress, and greater levels of motivation throughout the experiment relative to control, the latter of which again predicted performance on both tasks. Unlike Study 1, those in the pressure condition reported lower confidence in their performance than controls, though confidence did not relate to actual performance. Pressure still was related to increased participant effort on the WM task in Study 2. However, without feedback on the WM task, there were no longer statistically significant effects of pressure on either the WM task nor the verbal reasoning task, which was surprising given that feedback was never provided on the latter. In addition, and in contrast to Study 1, pressure neither boosted nor harmed WM or verbal reasoning performance in the absence of feedback. This suggests that feedback may potentially play a role in determining whether and how pressure may incentivize performance.

General Discussion

Across two studies, we explored how uncertainty can change performance by increasing effort. We manipulated uncertainty by adding performance pressure with some (Study 1) or no (Study 2) feedback while participants completed a WM task and a verbal reasoning test. Performance was optimized when there was some uncertainty (pressure with performance feedback; Study 1), but not under conditions of too much uncertainty (pressure without performance feedback; Study 2). Imposing pressure sustained participants' motivation, which positively predicted performance on both tasks. Critically, pressure predicted increased effort on the WM task, which predicted higher demonstrated WM capacity when feedback was provided in Study 1. Pressure-induced uncertainty may have optimized the arousal–performance relationship (Howard-Jones & Jay, 2016; Miendlarzewska et al., 2016), moving participants to the peak zone of the inverted-U, but only with feedback. This begs the question: When we test participants' cognitive abilities using performance measures, are we measuring their true capacity?

Pressure and Effort

Our data highlight the important mediating role of effort in WM capacity under pressure (Botvinick & Braver, 2015). Using a pre–post design, we extend this by showing that pressure can predict increased effort while simultaneously predicting increased stress and anxiety and, under some conditions, lowered confidence. Pressure improves performance by determining whether, how long, and to what extent individuals will deploy limited cognitive resources (Bonner & Sprinkle, 2002). Accordingly, when differences in performance were evident, participants under pressure saw higher demonstrated WM capacity on the higher WM-demanding items in particular (see also Heitz et al., 2008; Pochon et al., 2002), and gains on the verbal reasoning test were concentrated on the easier items, where additional effort can immediately facilitate performance without acquiring new vocabulary (Kiplinger & Linn, 1995; O'Neil et al., 1995). However, we are cautious without a direct measure of effort for verbal reasoning.

Feedback

Our findings suggest that removing feedback in addition to an imposed pressure may induce too much uncertainty, pushing individuals beyond the optimal amount of arousal in the inverted-U (Miendlarzewska et al., 2016; Yerkes & Dodson, 1908). Indeed, feedback is theorized to be an important moderator of the relationship between incentives, effort, and performance across many performance domains (see Bonner & Sprinkle, 2002; Byron & Khazanchi, 2012). Performance feedback indicates how far away one is from a desired goal, consequently signaling the need to put forth greater effort (Hattie & Timperley, 2007). Feedback can be quite motivating if the distance is small (Kluger & DeNisi, 1996), which was true for most Study 1 participants' performance on the WM task. When we removed feedback on the WM task in Study 2, the incentivizing effects of pressure on effort and task performance were much smaller in magnitude and no longer statistically significant as compared to Study 1. The relation between the manipulation of feedback on the WM task and performance on the verbal reasoning test was unexpected. Results from our counterbalancing tests showed no effect of task order on verbal reasoning or WM performance, thus this unexpected relation could not be attributed to a motivating spillover effect of receiving feedback on the WM task to verbal reasoning performance in Study 1.

Despite differences in pressure-related performance changes across studies, the effects of pressure on participants' affect were of similar magnitudes: With and without feedback, those under pressure reported greater stress and anxiety and lower motivation. Conversely, we note that the effect of pressure on participants' confidence did seem to change with feedback, with participants under pressure reporting less confidence than controls only when they completed the tasks without any feedback. Feedback is key for calibrating confidence with performance (Hattie, 2012), in part because feedback indicates the discrepancy between one's actual performance and goal performance (Hattie & Timperley, 2007). Yet, we provided outcome-level feedback, which can counterintuitively lower metacognitive accuracy during WM performance, possibly by reducing the need for self-monitoring (Adam & Vogel, 2018). Conversely, process-level feedback—indicating the discrepancy and providing guidance on how to reduce said discrepancy—can promote more adaptive mastery approaches (Hattie & Timperley, 2007; Pekrun et al., 2014). Beyond merely ameliorating uncertainty, such growth-focused feedback may encourage self-monitoring for those under pressure, which could further promote performance and improve metacognitive awareness.

We address two limitations of the feedback manipulation. First, feedback was only manipulated across studies, making interpretations of the effect of feedback and its relation to performance pressure challenging. Second, our manipulation of feedback was partial, as participants never received feedback on the verbal reasoning task in either study. This was because participants completed the task using pen and paper, disenabling us to score their tests and provide feedback in a timely manner. We are addressing both limitations in a follow-up experiment.

Choking Under Pressure?

We pursued WM as a mechanism for disentangling the role of uncertainty during performance, as prior work has shown WM performance can be facilitated (e.g., Heitz et al., 2008) and hindered (see Schmader et al., 2008) by pressure and feedback. Importantly, we did replicate that WM measures were not stable within person but rather varied according to the level of pressure and uncertainty, and thereby argue that these must be considered in any measurement endeavor.

Regarding the direction of these effects, we used similar social-evaluative pressure manipulations that others (e.g., Beilock et al., 2004) have reliably used to demonstrate the choking under pressure phenomenon, where increased anxiety generates worry that reduces available WM resources (Ashcraft & Kirk, 2001; Beilock, 2008). Our pressure manipulation differed from Beilock and colleagues in three ways: First, our participants under pressure saw the confederate with whom they would collaborate. Pilot participants told us that they did not believe there was another participant; seeing a confederate allayed this suspicion. Relatedly, because we had participant and confederate start tasks concurrently, we did not tell pressured participants that the confederate already achieved the performance target. Finally, participants were not videotaped while they performed. Nonetheless, our pressure manipulation maintained the key social (peer pressure) and evaluative (live performance monitoring by authority and peer) threat components that reliably increase indices of stress and anxiety (see Dickerson & Kemeny, 2004). Surprisingly, however, we found that this pressure improved performance on the highest WM-demanding trials (Heitz et al., 2008; Jimura et al., 2010; Pochon et al., 2002). We propose two explanations.

First, in line with inverted-U models (Sapolsky, 2015; Yerkes & Dodson, 1908), the social-evaluative manipulation could have induced an optimal amount of pressure—one that sufficiently motivates effort and increases performance (so long as feedback is provided) without inducing excessive concern. Though participants under pressure reported greater stress and anxiety than controls, their scores were typically below the midpoint of each scale—numerically consistent with research using a nearly identical manipulation to examine choking under pressure (Beilock et al., 2004; Sattizahn et al., 2016) and suggesting that these pressure manipulations might induce a moderate, optimal pressure. As we and others have argued, individual-level (effort; Eysenck et al., 2007) and task-level (feedback; Bonner & Sprinkle, 2002) factors may also explain when these purportedly anxiety-inducing manipulations threaten or facilitate performance. Still, a more debilitating pressure—or as we found, pressure without feedback—might not increase effort and should be explored in future work.

Alternatively, participants may not have choked by preemptively regulating their emotions (Johns et al., 2008; Pochon et al., 2002), perhaps biasing our understanding of their internal states. Moreover, increased effort expenditure itself may indicate that one is attempting to compensate for heightened anxiety (Eysenck et al., 2007). Yet, rarely do studies examining choking under pressure measure effort; those that do emphasize the importance of simultaneously considering anxiety–effort relations (Hardy et al., 2007; Putwain & Symes, 2018).

Practical Implications

These findings present practical implications for researchers and practitioners interested in assessing cognitive performance. Though cognitive measures are typically described as characterizing individuals' cognitive capacity in a static way, these results bear on the crucial nature of context on recorded measurements. That there were systematic differences in performance based on the pressure and feedback context elucidates the importance of recognizing the malleability of these measures in their interpretation. For example, researchers and practitioners oftentimes identify individuals as high or low WM (e.g., Beilock, 2008) or as more or less prepared for academic programs (ETS, n.d.) on the basis of a single assessment. Scores on standardized tests like the GRE are highly relied upon in graduate school admissions yet are unreliable indices of student success (e.g., see Moneta-Koehler et al., 2017), and as revealed in this paper, may be systematically unreliable based on the relative uncertainty in the test design.

Moreover, the predictive relationship between measurements of WM and other task performance indices may be complicated by pressure and uncertainty in the performance context. Cognitive psychological research examining choking under pressure often uses participants' WM capacity (as measured by a singular assessment) as a proxy for baseline cognitive capacity, with some (e.g., Beilock & Carr, 2005; Gimmig et al., 2006) arguing that individuals' baseline WM capacity may moderate the relation between pressure and cognitive performance. Yet, the influence of the performance context on the initial categorization of individuals into "high" and "low" WM remains underconsidered. Moreover, rather than directly predict students' achievement, more recent work argues that WM capacity may indirectly influence students' achievement by helping make more accessible and integrated their intrinsic and explicit motivations to succeed (Gareau et al., 2019). These relations between WM, uncertainty and motivation in the performance context, and demonstrated performance, and their implications in both the classroom and the laboratory, warrants further attention.

As a research design implication, experimenters are keenly aware that volunteer participants may not be highly motivated to perform (Sharp et al., 2006). Many have endeavored to mitigate the effects low motivation on the quality of experimental data via exclusion criteria or analytic corrections (e.g., Curran, 2016; Oppenheimer et al., 2009). We instead argue that a key goal for researchers is to optimize performance contexts so that all participants are incentivized to try their best. We align with learning and memory researchers (Howard-Jones & Jay, 2016; Luria et al., 2020; Miendlarzewska et al., 2016) who argue that gamifying performance contexts can maximize the validity of cognitive measures. We extend this to WM and verbal reasoning performance, showing that gamifying task instructions—

including establishing performance goals coupled with feedback and performance-contingent rewards (Luria et al., 2020)—can increase uncertainty about reward attainment, which increases arousal, motivation, and goal-directed effort (Miendlarzewska et al., 2016).

Researchers interested in optimizing participants' performance via gamified design should consider the minimum effective dose of uncertainty. We modeled our manipulation of uncertainty after prior social-evaluative pressure inductions (Beilock et al., 2004), which increased participants' motivation and effort but also increased negative affect. The benefits of gamification may manifest with less drastic changes to the performance context (Cheng et al., 2020), particularly regarding the evaluative nature of the pressure. Most neuroscientific and behavioral economic research revealing gains in cognitive performance following the promise of reward (see Botvinick & Braver, 2015) use incentive schemes that reward participants for goal attainment without punishing them for failing to meet said goals as we did. Inducing uncertainty with positive, not negative, reinforcement may most optimally improve performance with reduced concern for choking under pressure (Luria et al., 2020) and warrants future research.

Conclusion

Pressure while performing a cognitive task can increase performance by increasing effort. Our work shows that cognitive measures including WM and verbal reasoning, often used to characterize stable individual differences, are quite malleable to context and thus performance might not reflect true capacity or potential. Thus, we posit that individual difference measures should be administered in optimal performance contexts—where participants are maximally motivated yet not overwhelmed. We propose one mode for optimizing these contexts: adding uncertainty through pressure.

References

Acklin, W. T. (2012). *The effects of feedback on working memory capacity* (Publication No. 9949332834502959) [Master's thesis, University of Georgia]. Ex Libris.

Adam, K. C. S., & Vogel, E. K. (2016). Reducing failures of working memory with performance feedback. *Psychonomic Bulletin & Review*, *23*(5), 1520–1527. https://doi.org/10.3758/s13423-016-1019-4

Adam, K. C. S., & Vogel, E. K. (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, Perception & Psychophysics*, *79*(5), 1506–1523. https://doi.org/10.3758/s13414-017-1331-8

Adam, K. C. S., & Vogel, E. K. (2018). Improvements to visual working memory performance with practice and feedback. *PLOS ONE*, *13*(8), Article e0203279. https://doi.org/10.1371/journal.pone.0203279

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology*, *130*(2), 224–237. https://doi.org/10.1037/0096-3445.130.2.224

Attali, Y. (2016). Effort in low-stakes assessments: What does it take to perform as well as in a high-stakes setting? *Educational and Psychological Measurement*, *76*(6), 1045–1058. https://doi.org/10.1177/0013164416634789

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science*, *17*(5), 339–343. https://doi.org/10.1111/j.1467-8721.2008.00602.x

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science*, *16*(2), 101–105. https://doi.org/10.1111/j.0956-7976.2005.00789.x

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*, *133*(4), 584–600. https://doi.org/10.1037/0096-3445.133.4.584

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology*, *72*(2), 123–132. https://doi.org/10.1016/j.biopsycho.2005.08.007

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society*, *27*(4–5), 303–345. https://doi.org/10.1016/S0361-3682(01)00052-6

Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: From behavior to neural mechanism. *Annual Review of Psychology*, *66*(1), 83–113. https://doi.org/10.1146/annurev-psych-010814-015044

Byron, K., & Khazanchi, S. (2012). Rewards and creative performance: A meta-analytic test of theoretically derived hypotheses. *Psychological Bulletin*, *138*(4), 809–830. https://doi.org/10.1037/a0027652

Castro, J. F., Yamada, G., Contreas, H., Linares, F., & Watson, H. (2018). *The effect of monetary incentives on cognitive effort, emotions and test-solving performance* (Working Paper No. 131). Peruvian Economic Association.

Cheng, S., Jiang, T., Xue, J., Wang, S., Chen, C., & Zhang, M. (2020). The influence of rewards on incidental memory: More does not mean better. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *27*(11), 462–466. https://doi.org/10.1101/lm.051722.120

Crum, A. J., Salovey, P., & Achor, S. (2013). Rethinking stress: The role of mindsets in determining the stress response. *Journal of Personality and Social Psychology*, *104*(4), 716–733. https://doi.org/10.1037/a0031201

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*(8), 627–634. https://doi.org/10.1037/0003-066X.43.8.627

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*(3), 355–391. https://doi.org/10.1037/0033-2909.130.3.355

Educational Testing Service. (n.d.). *Skills measured*. https://www.ets.org/gre/institutions/about/general/skills/

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–236. https://doi.org/10.3758/s13421-014-0461-7

Gareau, A., Gaudreau, P., & Boileau, L. (2019). Past academic achievement contributes to university students' autonomous motivation (AM) which is later moderated by implicit motivation and working memory: A Bayesian replication of the explicit-implicit model of AM. *Learning and Individual Differences*, *73*, 30–41. https://doi.org/10.1016/j.lindif.2019.05.004

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*(6), 1005–1010. https://doi.org/10.3758/BF03213916

Hardy, L., Beattie, S., & Woodman, T. (2007). Anxiety-induced performance catastrophes: Investigating effort required as an asymmetry factor. *British Journal of Psychology*, *98*(1), 15–31. https://doi.org/10.1348/000712606X103428

Hattie, J. (2012). Calibration and confidence: Where to next? *Learning and Instruction*, *24*(1), 62–66. https://doi.org/10.1016/j.learninstruc.2012.05.009

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.

Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, *45*(1), 119–129. https://doi.org/10.1111/j.1469-8986.2007.00605.x

Hodges, W. F., & Spielberger, C. D. (1969). An indicant of trait or state anxiety? *Journal of Consulting and Clinical Psychology*, *33*(4), 430–434. https://doi.org/10.1037/h0027813

Howard-Jones, P. A., & Demetriou, S. (2009). Uncertainty and engagement with learning games. *Instructional Science*, *37*(6), 519–536. https://doi.org/10.1007/s11251-008-9073-6

Howard-Jones, P. A., & Jay, T. (2016). Reward, learning, and games. *Current Opinion in Behavioral Sciences*, *10*, 65–72. https://doi.org/10.1016/j.cobeha.2016.04.015

Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal of Experimental Social Psychology*, *46*(1), 208–212. https://doi.org/10.1016/j.jesp.2009.08.015

Jimura, K., Locke, H. S., & Braver, T. S. (2010). Prefrontal cortex mediation of cognitive enhancement in rewarding motivational contexts. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8871–8876. https://doi.org/10.1073/pnas.1002007107

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, *137*(4), 691–705. https://doi.org/10.1037/a0013834

Kane, M. J., & McVay, J. C. (2012). What mind wandering reveals about executive-control abilities and failures. *Current Directions in Psychological Science*, *21*(5), 348–354. https://doi.org/10.1177/0963721412454875

Kiplinger, V. L., & Linn, R. L. (1995). Raising the stakes of test administration: The impact on student performance on the National Assessment of Educational Progress. *Educational Assessment*, *3*(2), 111–133. https://doi.org/10.1207/s15326977ea0302_1

Kluger, A. N., & DeNisi, A. S. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *199*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Konheim-Kalkstein, Y. L., & van den Broek, P. (2008). The effect of incentives on cognitive processing of text. *Discourse Processes*, *45*(2), 180–194. https://doi.org/10.1080/01638530701792883

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 730–746. https://doi.org/10.1037/0278-7393.9.4.730

Luria, E., Shalom, M., & Levy, D. A. (2020). Cognitive neuroscience perspectives on motivation and learning: Revisiting self-determination theory. *Mind, Brain and Education: The Official Journal of the International Mind, Brain, and Education Society*, *15*(1), 5–17. https://doi.org/10.1111/mbe.12275

Maloney, E. A., Sattizahn, J. R., & Beilock, S. L. (2014). Anxiety and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(4), 403–411. https://doi.org/10.1002/wcs.1299

Miendlarzewska, E. A., Bavelier, D., & Schwartz, S. (2016). Influence of reward motivation on human declarative memory. *Neuroscience and Biobehavioral Reviews*, *61*, 156–176. https://doi.org/10.1016/j.neubiorev.2015.11.015

Moneta-Koehler, L., Brown, A. M., Petrie, K. A., Evans, B. J., & Chalkley, R. (2017). The limitations of the GRE in predicting success in biomedical graduate school. *PLOS ONE*, *12*(1), Article e0166742. https://doi.org/10.1371/journal.pone.0166742

O'Neil, H. F., Jr., Sugrue, B., & Baker, E. L. (1995). Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance. *Educational Assessment*, *3*(2), 135–157. https://doi.org/10.1207/s15326977ea0302_2

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Ozcelik, E., Cagiltay, N. E., & Ozcelik, N. S. (2013). The effect of uncertainty on learning in game-like environments. *Computers & Education*, *67*, 12–20. https://doi.org/10.1016/j.compedu.2013.02.009

Pekrun, R., Cusack, A., Murayama, K., Elliot, A. J., & Thomas, K. (2014). The power of anticipated feedback: Effects on students' achievement goals and achievement emotions. *Learning and Instruction*, *29*, 115–124. https://doi.org/10.1016/j.learninstruc.2013.09.002

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, *78*(4), 1255–1264. https://doi.org/10.1111/j.1467-8624.2007.01064.x

Pochon, J. B., Levy, R., Fossati, P., Lehericy, S., Poline, J. B., Pillon, B., Le Bihan, D., & Dubois, B. (2002). The neural system that bridges reward and cognition in humans: An fMRI study. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(8), 5669–5674. https://doi.org/10.1073/pnas.082111099

Putwain, D. W., & Symes, W. (2018). Does increased effort compensate for performance debilitating test anxiety? *School Psychology Quarterly*, *33*(3), 482–491. https://doi.org/10.1037/spq0000236

Randall, J. G., Oswald, F. L., & Beier, M. E. (2014). Mind-wandering, cognition, and performance: A theory-driven meta-analysis of attention regulation. *Psychological Bulletin*, *140*(6), 1411–1431. https://doi.org/10.1037/a0037428

Sapolsky, R. M. (2015). Stress and the brain: Individual variability and the inverted-U. *Nature Neuroscience*, *18*, 1344–1346. https://doi.org/10.1038/nn.4109

Sattizahn, J. R., Moser, J. S., & Beilock, S. L. (2016). A closer look at who "chokes under pressure." *Journal of Applied Research in Memory and Cognition*, *5*(4), 470–477. https://doi.org/10.1016/j.jarmac.2016.11.004

Schlosser, A., Neeman, Z., & Attali, Y. (2019). Differential performance in high vs low stakes tests: Evidence from the GRE test. *Economic Journal (London)*, *129*(623), 2916–2948. https://doi.org/10.1093/ej/uez015

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*(2), 336–356. https://doi.org/10.1037/0033-295X.115.2.336

Sharp, E. C., Pelletier, L. G., & Levesque, C. (2006). The double-edged sword of rewards for participation in psychology experiments. *Canadian Journal of Behavioural Science*, *38*(3), 269–277. https://doi.org/10.1037/cjbs2006014

Sheppes, G., & Meiran, N. (2008). Divergent cognitive costs for online forms of reappraisal and distraction. *Emotion*, *8*(6), 870–874. https://doi.org/10.1037/a0013711

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199. https://doi.org/10.1177/1745691615569000

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

Wise, S. L., & DeMars, C. E. (2010). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, *10*(1), 1–17. https://doi.org/10.1207/s15326977ea1001_1

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, *18*(2), 163–183. https://doi.org/10.1207/s15324818ame1802_2

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *The Journal of Comparative Neurology and Psychology*, *18*(5), 458–482. https://doi.org/10.1002/cne.920180503

Zedelius, C. M., Veling, H., & Aarts, H. (2011). Boosting or choking—How conscious and unconscious reward processing modulate the active maintenance of goal-relevant information. *Consciousness and Cognition*, *20*(2), 355–362. https://doi.org/10.1016/j.concog.2010.05.001

Received March 20, 2021
Revision received February 4, 2022
Accepted March 1, 2022 ■