Performance under pressure:
Implications for how society organizes education and the workplace

Joaquín Artes Caselles[a], Jennifer Graves[b,1] , Maryl Motika[c]

[a] *Universidad Complutense de Madrid, Departamento de Economía Aplicada IV, Cuidad Universitaria, 28040 Madrid, Spain*

[b] *Universidad Autónoma de Madrid, Departamento de Economía y Hacienda Pública, Cuidad Universitaria de Cantoblanco, 28049 Madrid, Spain*

[c] *Saint Lawrence University, Department of Economics, 23 Romoda Dr., Canton, New York 13617, United States*

October 30, 2016

PRELIMINARY WORK. DO NOT CITE

[1] Corresponding author. US Tel.:206-618-0748, Spain Tel: 626 20 28 39. Email: jennifer.graves@uam.es

ABSTRACT

Using student participants at two universities in an experiment, we estimates the effects of multiple types of pressure-inducing incentives that are commonly used in the workplace and educational settings to measure the effects on productivity of individuals completing different tasks that are relevant to the real world. The sources of pressure we impose are competition, high-stakes pay and time pressure, each evaluated against a low-pressure piece-rate incentive. Each of these incentives are applied in situations with subjects completing three types of tasks: a routine or straightforward task, a purely creative or divergent thinking task and a creative problem-solving or convergent thinking task. Each type of task and pressure-inducing incentive mimic those commonly found in in today's schools and workplace. Previous literature has tested the effects of high-stakes pay and time pressure on creative versus non-creative tasks. However, our work is the first to test task-specific productivity effects using competition as a pressure. While there is a large literature in economics on the productivity effects of competition, it has focused largely on differences in performance across subjects and implications for wage gaps in the labor market, but has not distinguished between creative and mechanical tasks performed. As a secondary contribution to the literature, we also improve upon previous studies evaluating high-stakes pay and time pressure on creative versus non-creative tasks by drawing the distinction between types of creativity (divergent or convergent) that are treated differently in psychology. Also, by estimating the various tasks and pressures in a consistent way, within the same experiment, we are able to draw comparisons between performance effects on different tasks under different pressures that have not been possible in the current literature due to differing methodology.

Keywords: performance-based incentives, divergent and convergent thinking, creative versus mechanical tasks

JEL Codes:
D00, microeconomics, general
D03, behavioral microeconomics: underlying principles

1. ***Introduction***

The study of economics is innately tied to the analysis of how individual agents respond to incentives. In this project, we aim to discover under what conditions people respond differently than expected to the incentives commonly used within society in both educational settings and the workplace. The view predominantly held in economics and the business world regarding workforce productivity is that greater incentives yield greater work performance. For example, the classic economic assumption is that, all else the same, as an employer offers higher pay, it can recruit the highest quality workers and demand more from those workers. The reasoning is that higher pay compensates workers for their increased efforts and the company's spending on higher wages is compensated through increased productivity. While incentives are clearly strong medicine within the labor market, recent research has called into question whether businesses are applying this medicine in the most efficient way. For example, Ariely, Gneezy, Leowenstein and Mazar (2009) find that the pressure from high-stakes payment schemes can actually lead to worse performance, especially when applied to tasks that are creative in nature, rather than simple mechanical tasks. In education, we see many examples of the same types of assumptions being made about connections between evaluation incentives and student performance. High-stakes pressure innate to performance-contingent scholarship funding in higher education is one such example, driven by the same underlying assumptions about performance as many bonus-pay schemes in the business world. Additionally, time pressure is built into nearly all standardized testing settings, and yet, unless the aim is actually to evaluate speed of performance, such a setting may not yield the highest productivity from students.

Previous literature has tested the effects of high-stakes pay and time pressure on creative versus non-creative tasks. However, our work is the first to test task-specific productivity effects using competition as a pressure. While there is a large literature in economics on the productivity effects of competition, it has focused largely on differences in performance across subjects and implications for wage gaps in the labor market, but has not distinguished between creative and mechanical tasks performed. As a secondary contribution to the literature, we also improve upon previous studies evaluating high-stakes pay and time pressure on creative versus non-creative tasks by drawing the distinction between types of creativity (divergent or convergent) that are treated differently in psychology.

While a number of different types of high-pressure incentives and their effects on task performance have been studied in isolation or small pairings, the literature lacks a study that combines and makes comparable different incentive systems for educational and workplace relevant tasks. By estimating the various tasks and pressures in a consistent way, within the same experiment, we are also able to draw comparisons between performance effects on different tasks under different pressures that have not been possible in the current literature due to differing methodology. In this way, we also aim be the first study to thoroughly study how multiple types of pressure-inducing incentives that are commonly used in the workplace and educational settings impact productivity of individuals completing different tasks that are relevant to the real world.

We use an experimental design to test how pressure-inducing incentives increase or decrease subjects' performance on various tasks. The sources of pressure we impose are competition, high-stakes pay and time pressure, each evaluated against a low-pressure piece-rate incentive. Each of these incentives are applied in situations with subjects completing three types of tasks: a routine or straightforward task, a purely creative or divergent thinking task and a creative problem-solving or convergent thinking task. Each pressure-inducing incentive is commonly applied in today's labor market through bonuses, commissions, as well as competition for funding and contracts, and in schools in the form of testing and grading incentives as well as in the higher education application and funding process. Additionally, the three types of tasks mimic common tasks required in today's schools and workplace. Divergent thinking tasks mimic brainstorming, with the aim of generating unique ideas without necessarily solving a pre-determined specific problem or goal. Convergent thinking tasks differ in that there is a goal or answer that needs to be achieved, but arriving at that solution requires complex thought process and creativity. Research and development on a new product most closely resembles a convergent thinking task, with worker productivity in this area likely being particularly important for results.

This difference between pure creativity and creative problem solving is important in the psychological literature on creativity (Hocevar 1981; Byron, Khazanchi, and Nazarian 2010). These creative tasks are quite different from routine tasks such as adding numbers, which do not require such “outside” thinking. The distinction between these three tasks is important for interpreting our findings in the context of preparation for and productivity in the labor market, as high-skilled work has moved further from routine tasks (many of which can now be automated and completed by computers) and more toward tasks that are creative in nature. The results of the project can potentially provide a more complete understanding of how students and workers respond to incentives and how to better use these incentives to improve learning and student evaluation in educational settings, as well as increase worker’s performance in different types of jobs.

2. *Literature*

The idea that pressure can alter performance on tasks in counter-productive ways is not a new concept in the academic literature. Many researchers, mostly within psychology, have studied phenomena such as “choking under pressure” (Dandy, Brewer and Tottman, 2011; Baumeister and Showers, 1986; Dohmen, 2008). Additionally, incentives may impact performance through altering intrinsic versus extrinsic motivation (Eckartsz, Kirchkamp, and Schunk, 2013). Existing research has shown differential results by creative versus non-creative tasks under pressure, typically finding that high pressure reduces performance in creative tasks and increases performance on routine tasks.[2] However, current studies generally focus either on only one form of pressure or one type of task at a time and use tasks less grounded in the psychology literature.

While both high-stakes pay and time pressure have been paired with creative versus non-creative tasks in previous studies, much of this literature has not drawn a distinction between types of creativity (divergent or convergent) that are treated as different ways of thinking within the psychology literature, and others have used a variety of creative tasks that are not directly motivated by evidence from psychology literature. A meta-study by Kristin Byron, Shalini

[2] See for example: Beilock and Carr (2005), Webb et al. (2013) and Byron et al. (2010).

Khazanchi, and Deborahon Nazarian (2010) on the relationship between stressors and creativity emphasizes the need for additional research on this topic to clarify differences between different types of tasks and stressors. There is also a very large literature in economics on the productivity effects of competition, but this literature has focused largely on differences in performance across subjects and implications for wage gaps in the labor market, but has not generally made the distinction between creative and mechanical tasks performed[3]. This literature has not been sufficiently evaluated in terms of potentially differing performance effects by type of task.

This study contributes to the larger literature in a number of ways. Firstly, our work is the first to test task-specific productivity effects using competition as a pressure. Previous literature has tested the effects of high-stakes pay and time pressure on creative versus non-creative tasks. However, competition, which has been extensively studied within economics, has not yet been studied as a productivity-inducing incentive for its effects on task-specific productivity. Additionally, we aim to be the first to clearly distinguish how multiple types of pressure imposing incentives that are often used in school settings and in the workplace impact productivity of individuals completing different types of tasks typically used in student evaluation and those relevant to the workplace. For example, it would be informative in designing worker incentive schemes to know whether the productivity effects from time pressure are as large as those from high-stakes pay, and whether this differs by type of creative task, which may be used to different degrees in different businesses. Additionally, we improve upon previous literature by directly incorporating recent developments in the psychology literature to distinguish between different types of pressure-inducing incentives and different types of tasks.

Our findings have the potential to inform educational policies, as well as workplace compensation policies. They may also motivate a wide range of future work that can help to explain commonly discussed, yet still not fully understood, labor market concerns, such as wage gaps and inequality. We hypothesize that it is likely to find heterogeneity in productivity under pressure across study subjects. In this case, the findings of this study may help explain why some firms thrive under high-pressure compensation schemes, while others may not. It is quite

[3] See for example: Niederle and Vesterlund (2007), Delfgaauw, Dur, Sol and Verbeke (2009), Lavy (2008) and De la Rica, Dolado and Vegas (2010).

possible that by implementing a particular incentive-scheme, a company may distort the productivity of its workers, making some workers less efficient than they would be otherwise. Additionally, we see this project as a necessary first step in a broader series of studies that can explore many widely-observed phenomena in the workplace. Do results vary by gender or race? Is there a threshold at which different people "crack" under pressure? The answer to these questions could have implications for current societal concerns such as the gender wage gap, the "glass ceiling" or other observed inequalities.

3. Experimental Methodology

Subjects will be recruited from two separate behavioral economics laboratories on university campuses in the United States. The first laboratory is at the University of California, Irvine where we use the Experimental Social Sciences Laboratory subject pool. This subject pool is a list of UCI students over the age of 18 who sign up to participate in research. They will be invited by the ESSL staff in charge of recruitment and subject pool management, and will sign up through the ESSL website. The second laboratory is at Saint Lawrence University, where student subjects are recruited in a similar manner.

Each session will take place in a computer laboratory. Upon arrival at the session, subjects will be asked to read and sign the informed consent form and will have an opportunity to ask questions about the study. They will then be seated at a computer. The session will begin with more detailed instructions about the kinds of choices subjects will make and how they will be paid. They will then participate in a series of tasks. At the end of the session their payment will be determined. Each subject will then be paid individually and privately before leaving the room. Each session will last approximately 30-60 minutes. No follow-up visits will occur.

Immediately after subjects are seated at the computers, initial instructions will give subjects a general description of the experiment and information on payments. Subjects will be paid the show-up fee plus their earnings in one task which will be randomly selected from those they have completed at the end of the session. Paying for only one task removes the possibility of wealth or hedging effects on behavior; it would be detrimental to our research if subjects became less risk-averse as the session continued because they had already earned significant amounts of money. After learning about the payments, subjects will then read instructions and complete several practice questions for their first task. After completing the first task, they will read instructions and complete practice questions for the second task, and so on.

We will assign three different types of task: routine, convergent thinking, and divergent thinking. Furthermore, we will use four different methods of payment: piece rate, tournament, time pressure, and high incentive. Pairing each type of task with each type of payment makes 12 different tasks a subject might be asked to complete. Each subject will do no more than 8 tasks total – we plan to have each participant do 6 tasks. Different subjects may receive different sets of tasks.

The routine task is completing as many problems as possible in five minutes, where the problems are counting the number of times a particular letter appears in a sentence. The convergent thinking task consists of accurately completing as many matchstick questions (Knoblich et al. 1999) as possible in five minutes. These questions are drawn from the psychology literature on convergent thinking. A matchstick question takes the form II – III = IV where each line is conceived as a matchstick that can be moved around. The subject is asked to move one stick to make the equation accurate. The divergent thinking task measures a more frequently studied type of creativity: the ability to find original responses. We will use the unusual uses task to measure divergent thinking. In this task, participants are asked to find as many non-standard uses as possible for a household item such as an umbrella. Participants receive a point for each different use they list.

Under the piece rate, subjects earn money for each correct response. Under tournament pay, subjects are placed anonymously in groups of 4 via the computer and compete with other members of their group. They receive nothing if they lose and a high payment if they win. For time pressure, subjects receive a moderate payment if they complete a moderate number of problems correctly in a given time, a high payment if they complete a large number, and no payment if they are unable to complete the moderate number of problems. High incentives is the same as time pressure with the difference that the amounts of money offered are $100 or $200 as opposed to $10 or $20 for simple time pressure. The effect of the payments themselves is therefore measured as the difference between the high payment effect and the time pressure effect. This use of high payments in conjunction with time pressure mirrors the methods of other papers examining effects of high payments, in particular Ariely, Gneezy, Loewenstein and Mazar (2009). Setting aside the high incentive treatment, each of the payment systems be scaled so that, on average, subjects earn an average of $10 from the task. All of the payment levels used in this experiment are motivated by empirical literature both in behavioral economics and in psychology. While per-question payments may appear small, they also take very little time to complete. To provide context, payment amounts (except for high-stakes payments) amount to an average hourly wage of 16,67€.[4]

In addition to these tasks, subjects will complete a survey about other characteristics that might plausibly affect outcomes such as gender, age, year in school, major, fluency in English, and typical weekly spending to assess the impact of the high incentive. To include risk preference as an additional potentially important control (Cadsby, Song, and Tapon 2009), we will also use a risk preference elicitation method established in the literature (Holt and Laury,2002; Harrison, 2002). This involves giving subjects $3 with which they can choose to gamble. They then make a series of choices between more-risk and less-risky gambles. For example, one choice might be between Lottery A offering a 50% chance of $2 and a 50% chance of $4 and Lottery B offering a 50% chance of $1 and a 50% chance of $7. After they make their choices, a random number generator determines which choice determines their earnings and whether or not they win.

At the end of the session, a task to be compensated will be selected from those the subject completed in the first part of the study and subjects will be paid the total of their earnings for that task, the show-up payment, and the risk preference elicitation. We protect individual participants, by not including names or identifying information in our data. These data will be shared between the three researchers and kept on password-protected computers.[5] However, participants are offered the opportunity to sign up to receive a summary of the experiment results at a later date.

---

[4] How much is this worth for a student? In the United States, for example, where students commonly hold jobs while studying, a typical pay for a minimum wage on-campus job is roughly $8,50.

[5] We intend to keep the data indefinitely, as the current standard in experimental economics is to keep non-identifying data so that it can be shared with other researchers engaged in replication or meta-studies. Sharing non-identifying data has been crucial for methodological research and has contributed to our field's understanding of issues such as publication bias and the need for power calculations. The names of participating subjects will be recorded in our recruitment database. This database will be password-protected, only accessible by Meryl Motika

4. *Estimation*

After experiments are run, estimation will be done using standard estimation techniques, namely regression analysis with a series of dummy variables for interactions of each type of task and pressure. We will also include dummy variables to account for experimental design features that could otherwise impact results. For example, dummy variables representing the number of times the participant has completed that task in order to account for learning effects, as well as an individual fixed effect to account for the subject's ability level as well as other unobserved time-invariant characteristics.

To estimate the differential effect of types of pressures on creative versus mechanical tasks, a difference in differences approach can be used. With this approach one would be comparing each individual's performance under pressure to their piece-rate performance, and then comparing this difference in response to pressure calculated for a creative task to the same difference in response calculated for the routine task.

The outcome we are interested in is the interaction between type of task and pressure. We will observe outcomes at the individual-subtask level. To see this more formally, a model for the routine task is provided below.

$$Y = \alpha + \beta_1 T + \beta_2 C + \beta_3 I + \beta_4 D1 + \beta_5 D2 + \beta_6 D3 + F_i + \epsilon$$

Y is the outcome (number of problems completed for speed, proportion correct for accuracy). Dummy variables represent the effects of the different pressure treatments. T is time pressure and is equal to one for high-stakes pay as well as time pressure treatments, C is competitive pressure, and I is high-stakes pay. D1 through D3 are dummies representing the number of times the participant has completed that task in order to account for learning effects. We use separate

---

and a student research assistant, and will be used only to track participation and prevent the same subjects from participating in multiple sessions of the same experiment. It will not be linked to the results in any way.

dummies rather than a scalar because learning effects are typically nonlinear. We assume that learning effects are independent of pressure. Between-task order effects are controlled through randomization. An individual fixed effect is also included to account for the subject's ability level as well as other unobserved time-invariant characteristics. Including the individual fixed effect for the task makes the analysis of pressure within- rather than between-subjects; we are comparing each individual's performance under pressure to their piece-rate performance. The primary coefficients of interest for the routine task analysis are β_1, β_2, and β_3, which represent the effect of each type of pressure.

In order to determine the differential effects on creative tasks, changes in accuracy and speed can be measured for creative versus mechanical tasks using a difference in differences approach. When we do this, we are comparing each individual's performance under pressure to their piece-rate performance, and then comparing this response to pressure calculated for a creative task to this response calculated for the routine task. For example, suppose that in a piece-rate environment the average number of arithmetic problems solved in five minutes is 25, and the average number of convergent thinking problems solved is 10. If moving from piece-rate to competitive pay has no effect on arithmetic but reduces the average number of convergent thinking problems solved to 8, then this type of pressure might be said to reduce creative problem solving speed more than mechanical speed. This difference in differences approach that yields relative results for different types of tasks is shown in the model below.

$$Y = \alpha + \beta_1 T + \beta_2 C + \beta_3 I + \beta_4 CT + \beta_5 CT * T + \beta_6 CT * C + \beta_7 CT * I + \beta_8 D1 + \beta_9 D2 + \beta_{10} D3 + \epsilon$$

The coefficients of interest in this model are β_5, β_6, and β_7. These coefficients represent the differential effects of pressure on the convergent thinking task compared to the routine task. The new dummy variable CT is equal to one if the observation is for the convergent thinking and zero for the routine task. Individual fixed effects are included separately for the creative versus the routine task because subjects might be better at one than at the other.

5. Preliminary Results

TO BE ADDED SOON. EXPERIMENT IN PROGRESS.

6. Discussion

7. Conclusions

References:

1. Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar, "Large Stakes and Big Mistakes," *Review of Economic Studies*, 2009, 76 (2), 451–469.
2. Baumeister, Roy F., and Carolin J. Showers, "A review of paradoxical performance effects: Choking under pressure in sports and mental tests," *European Journal of Social Psychology*, XVI (October-December 1986), 361-383.
3. Beilock, Sian L. and Thomas H. Carr, "When High-Powered People Fail," *Psychological Science*, 2005, 16 (2), 101–105.
4. Bolte, A., Goschke, T., & Kuhl, J.. Emotion and intuition: Effects of positive and negative mood on implicit judgments of semantic coherence. *Psychological Science*, 2003, 14, 416-421.
5. Byron, Kristin, Shalini Khazanchi, and Deborah Nazarian, "The relationship between stressors and creativity: A meta-analysis examining competing theoretical models," *Journal of Applied Psychology*, 2010, 95 (1), 201–212.
6. Cadsby, C. Bram, Fei Song, and Francis Tapon, "The Impact of Risk Aversion and Stress on the Incentive Effect of Performance Pay," Technical Report 2009.
7. Dandy, Justine, Neil Brewer, and Robin Tottman, "Self-Consciousness and Performance Decrements Within a Sporting Context," *The Journal of Social Psychology*, 2001 Vol. 141 (1).
8. Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke (2009). "Tournament Incentives in the Field: Gender Differences in the Workplace," Tinbergen Institute Discussion Papers 09-069/1, Tinbergen Institute August.
9. De la Rica, Sara, Juan J. Dolado, and Raquel Vegas (2010). "Performance Pay and the Gender Wage Gap: Evidence from Spain," IZA Discussion Papers 5032, Institute for the Study of Labor (IZA) June.
10. Dohmen, Thomas J., "Do professionals choke under pressure?," *Journal of Economic Behavior & Organization*, March 2008, 65 (3-4), 636–653.
11. Grosse, Niels D. and Riener, Gerhard, Explaining Gender Differences in Competitiveness: Gender-Task Stereotypes, No 2010-017, Jena Economic Research Papers, Friedrich-Schiller-University Jena, Max-Planck-Institute of Economics, 2010.
12. Hocevar, D. "Measurement of creativity: review and critique." *Journal of Personality Assessment*, Oct 1981, 45(5), 450-64.
13. Harrison, Glenn W., "Estimating Individual Discount Rates in Denmark: A Field Experiment,"*American Economic Review*, 2002, 92 (5), 1606–1617.

14. Holt, Charles A. and Susan K. Laury, “Risk Aversion and Incentive Effects,” *The American Economic Review*, 2002, 92 (5), pp. 1644–1655.

15. Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 1999, 25, 1534-1555.

16. Lavy, Victor (2008). “Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-based Pay Tournaments among Teachers,” NBER Working Papers 14338, National Bureau of Economic Research, Inc September.

17. Niederle, Muriel and Lise Vesterlund. (2007). “Do Women Shy Away from Competition? Do Men Compete Too Much?,” The Quarterly Journal of Economics, 08, 122 (3), 1067–1101.

18. Webb, R. Alan, Michael G. Williamson, and Yue (May) Zhang “Productivity-Target Difficulty, Target-Based Pay, and Outside-the-Box Thinking,” *The Accounting Review*, July 2013, Vol. 88, No. 4, pp. 1433-1457.