Creativity under Pressure: How Compensation Schemes Interact with Task Type and Gender in Incentivizing Performance

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

PRELIMINARY WORK. PLEASE REFERENCE AS SUCH.

This draft: May 19, 2017

For updated versions of the paper, please see:

https://sites.google.com/site/jenniferannegraves/

[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*

[1] Corresponding author. Email: jennifer.graves@uam.es

ABSTRACT

Using university student participants in an experiment, we test the task-specific productivity effects of a number of incentivizing payment schemes. The incentives we use in our study are competition, high-stakes pay, time pressure and a piece rate pay, each evaluated against a flat rate pay incentive. Each of these incentives are applied in situations with participants completing three types of tasks: a routine task, a purely creative (or divergent thinking) task and a creative problem-solving (or convergent thinking) task. We also test for gender differences in these performance effects. We build on previous literature in two key ways. First, by testing these various tasks and pressures in the same experimental design, we are able to make comparisons across task types that have not been possible in previous studies. Second, we bridge a gap in the literature regarding tasks and gender. Competition has been more extensively studied in terms of gendered effects, and yet gendered effects of high-stakes pay and time pressure on productivity have been largely overlooked. What evidence does exist for other pressures motives further exploring the intersection between pressure, task types and gender. Our results show that all incentivizing payment schemes improve productivity relative to a neutral flat rate payment scheme for routine tasks. However, when performing cognitively challenging tasks - both purely creative and problem solving tasks - incentive schemes have task-specific results. For both of these tasks, we find that time pressure produces results no different than having no incentive in place, and that participant performance is decreased by high stakes pay. Competition, on the other hand, has the largest boost in productivity for both types of cognitively challenging tasks. Testing for gender differences in these findings is still in progress.

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 view predominantly held in economics and the business world regarding workforce productivity is that greater incentives yield greater work performance. For example, that higher pay will induce higher effort, or that competition will lead workers to strive harder to win, and that this increased effort results in greater productivity. We see many examples of this assumption in practice in both educational settings and the labor market. However, these incentives used to motivate effort could alternatively be seen as sources of pressure. The idea that pressure can alter performance on tasks in counter-productive ways is not a new concept in the academic literature. Many researchers have studied phenomena such as "choking under pressure" (Dandy, Brewer and Tottman, 2011; Baumeister and Showers, 1986; Dohmen, 2008). This idea is supported by many studies that find pressure to have productivity reducing effects (e.g. Glucksburg, 1962; Dandy et.al, 2011; Webb et al., 2013).[2] In psychology research, pressures are thought to diminish working memory, potentially reducing productivity. Combining the economic and psychology perspectives, the predicted effect of incentive schemes is ambiguous as the positive incentivizing effect on productivity could be offset to an unkown degree by the pressure from such incentives altering the connection between incentives and effort and/or between effort and productivity. Additionally, incentives - which provide extrinsic motivation - may impact performance through altering intrinsic motivation (Eckartsz, Kirchkamp, and Schunk, 2013).

Using university student participants in an experiment, we test the task-specific productivity effects of a number of incentivizing payment schemes. The sources of pressure-inducing incentives we use in our study are competition, high-stakes pay, time pressure and a piece rate pay, each evaluated against a flat rate pay incentive. Each of these incentives are applied in situations with participants completing three types of tasks: a routine task, a purely creative (or divergent thinking) task and a creative problem-solving (or convergent thinking task). In addition

[2] Webb et. al. (2013) finds that pressure from increased difficulty relative to available time leads to more effort, but a decreased ability in process innovation, while a lower pressure setting does not hinder this ability.

to our main effects we test for gender differences in these performance effects.

Within economics, few studies have drawn a distinction between different types of cognitively demanding tasks, such as problem-solving and pure creativity (convergent and divergent thinking) that are treated as different ways of thinking within the psychology literature. A meta-study by Byron, Khazanchi, and 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. 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. 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 task; tasks which may be used to different degrees in different businesses. Our experimental design contributes to the literature on incentivizing pressures both in our use of a common framework and also by drawing the distinction between different types of cognitively demanding tasks.

Additionally, recent evidence indicates that pressure effects may be task-specific, especially along the dimension of creativity. For example, Ariely, Gneezy, Leowenstein and Mazar (2009) find that the pressure from high-stakes payment schemes increase performance in very mechanical tasks, but reduce performance in creative, high-concentration or physical tasks. A number of studies have found differential results by creative versus non-creative tasks under pressure, many finding that high pressure reduces performance in creative tasks and increases performance on routine tasks (Byron et al., 2010).[3] A recent study by Charness and Grieco (2014), using the same two types of creative tasks as we do in this study, find that competition has task-specific effects. They find that competition relative to a flat rate pay increases

[3] Much of this literature, however looks at a single task type. For example, Erat and Gneezy (2015) find that providing a piece-rate incentive did not improve productivity, and that competition reduced productivity on creative tasks. As they note in their conclusions, however “these results do not tell us directly when incentives might interact with the task type, and if these results might be equally applicable to non-creative and possibly routine tasks.” (p.18).

performance on creative problem solving (convergent thinking), but not on pure creative tasks (divergent thinking), except for those participants with an aversion to ambiguity, for which incentives also helped in the latter case. Our research builds on these studies that suggest that productivity effects are both pressure and task-specific.

An additional advantage of testing performance effects of various tasks and pressures in the same experiment is that it allows us to bridge a gap in the literature regarding tasks and gender. In contrast with the other pressure types studied here, competition has been more extensively studies in terms of gendered effects,[4] and yet gendered effects of high-stakes pay and time pressure on productivity have been largely overlooked.[5] There are some exceptions, such as Azmat, Calsamiglia and Iriberri (2016) who use a natural experiment to find that female students perform relatively worse on high stakes tests than male students. In another notable study, Shurchkov (2012) finds that gender differences in competition can be explained in part by preferences for gendered topics (math vs. reading) and partly by time constraints. In this study, she notes that the seminal papers on competition and gender predominantly use both math problems and time pressure, which when relaxed in her study (verbal and low time pressure) result in a reversal of gendered preferences. This further motivates the need to explore task-specific performance effects, as well as the possibility of differences by gender in these results.

While our study is conducted in an experimental laboratory setting, each type of task and pressure-inducing incentive used in our study mimic those commonly found in in today's schools and workplace. High stakes pay, for example, resembles the pressure imposed by performance-contingent scholarships in higher education or bonus pay in the business world. One can regularly see competition used as an incentive for contracts, promotions and prizes. Time pressure is imposed in work deadlines and nearly all standardized testing settings, and piece rate pay is similar to payment schemes used for simple contract work. Regarding tasks, creative (divergent thinking) tasks mimic brainstorming, with the aim of generating unique ideas without

[4] (See for example: Niederle and Vesterlund, 2007; Delfgaauw et al., 2009; Lavy, 2008)

[5] This is especially relevant since a number of studies relating performance pay contracts to the gender wage gap, have found mixed results, largely due to the different performance pay systems studied (Kangasneimi and Kauhanen, 2013; de la Rica et. al. 2010; Jijahn and Stephan, 2004; Booth and Frank, 1999).

necessarily solving a pre-determined specific problem or goal. Cognitive problem solving (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. The research and development process on a new product likely closely resembles a convergent thinking task, with worker productivity in this area likely being particularly important for success. 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 cognitively demanding and creative in nature.

Our results show that all incentivizing payment schemes improve productivity relative to a neutral flat rate payment scheme for routine tasks. However, when performing cognitively challenging tasks - both purely creative and problem solving tasks - incentive schemes have task-specific results. For both of these tasks, we find that time pressure produces results no different than having no incentive in place, and that participant performance is decreased by high stakes pay. Competition, on the other hand, has the largest boost in productivity for both types of cognitively challenging tasks. Testing for gender differences in these findings is still in progress.

The remainder of our paper is organized as follows: We describe our experimental design in Section 2, followed by our estimation strategy in Section 3, then we discuss our results in Section 4 and robustness of findings and extensions in Section 5, followed by concluding remarks in Section 6.

2. *Experimental Design*

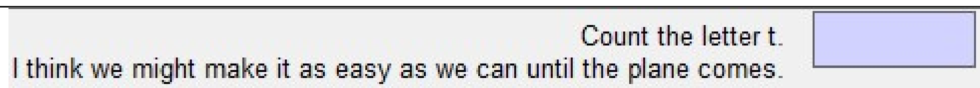
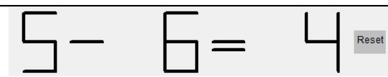
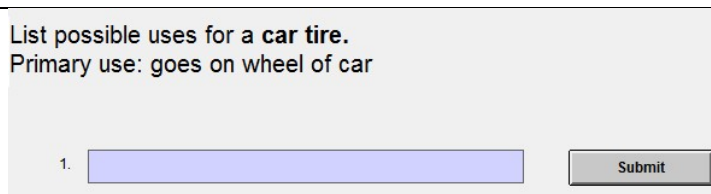
For our experiment, participants have been recruited from the behavioral economics laboratory at the University of California - Irvine (UCI), where we use the Experimental Social Sciences Laboratory (ESSL) subject pool. This subject pool comes from a list of UCI students over the age of 18 who sign up to participate in research. They are invited by the ESSL staff in charge of recruitment and subject pool management, and sign up through the ESSL website. Each session takes place in a computer laboratory. Upon arrival at the session, participants are asked to read and sign the informed consent form and have an opportunity to ask questions about the study. They are then be seated at a computer. Initial instructions provide participants with a general description of the experiment and information on payments. Participants are paid the show-up fee plus their earnings in one task, which is randomly selected from those that they have completed, at the end of the session. Paying for only one task removes the possibility of wealth or hedging effects on behavior, as it would be detrimental to our research if participants became less risk-averse as the session continued because they had already earned significant amounts of money.

Each participant completes three different types of tasks: routine, cognitive problem solving (convergent thinking), and creative (divergent thinking). In our experiment, we assign five different methods of payment: flat rate, piece rate, competition, time pressure, and high-stakes pay. We assign all tasks to each subject, but randomly assign only one payment type to each participant. Assigning incentives across participants limits the amount of new information each subject must understand to complete the rounds assigned to them. Both the payment type and the order of the tasks assigned are randomized. After learning about the payment method assigned to them, participants then read instructions and complete several practice questions for their first task. After completing the first task, they will first read instructions and then complete practice questions for the second and third task respectively. At the end of the session their payment is determined (through random selection of one of the previously completed tasks). Each subject is then be paid individually and privately before leaving the room. Each session lasts between approximately 30 and 60 minutes and no follow-up visits occur.

2.1 Tasks performed

The routine task is completing as many problems as possible in five minutes, where the problems consist of counting the number of times a particular letter appears in a sentence. An example of how this appears to the student is provided in Figure 1 (below). The cognitive problem solving (convergent thinking) task consists of accurately completing as many matchstick questions as possible in five minutes (Knoblich et al. 1999). These questions are drawn from the psychology literature on convergent thinking. An example of this is also presented in Figure 1. The matchstick question involves an incorrect math equation, where numbers are written in “block letter” form, where each line is conceived as a matchstick that can be moved around. The participant is asked to move one stick to make the equation accurate. In the case presented in Figure 1, the student can move the bottom left stick on the 6 to the top right position on the 5, turning the equation into 9-5=4.

Figure 1: Examples of how tasks appear to students by task types

Routine	
Problem solving	
Pure creative	

The creative (divergent thinking) task measures a more frequently studied type of creativity: the ability to find original responses. We 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. An example of a prompt for a student for this task is also presented in Figure 1. This difference between pure creativity and cognitive 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.

2.2 Sources of pressure (payment schemes)

Under the piece rate payment scheme, participants earn money for each correct response. Under competition, participants are placed anonymously in groups of four via the computer and compete with other members of their group, anonymously, without interaction or information on each other. They receive nothing if they lose and a high payment if they win. For the time pressure, we used timed goals. Participants 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 in the time provided. High-stakes pay is the same as timed goals pressure, with the difference that the amounts of money offered are $100 or $200 as opposed to $10 or $20 for simple time pressure.

Table 1: Summary of payments under different types of incentives (or pressures)

Type of Incentive Pressure	**Payment Amount**
Piece rate	$0.80 - $2.00 per point earned *(depending on task)*
Timed goal	$10 is achieve medium goal $20 if achieve high goal $0 if do not achieve medium goal
Competition *(groups of 4 with no interaction or information on each other)*	$40 to winner $0 to others
High-stakes pay *(timed goal with higher payouts)*	$100 is achieve medium goal $200 if achieve high goal $0 if do not achieve medium goal

The effect of the payments themselves can therefore be measured as the difference between the high payment effect and the timed goals 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-stakes pay treatment, each of the payment systems are scaled so that, participants earn an average of $10 from the task paid out. The break-down of payments are summarized in Table 1.

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. For example, in the United States, where university students commonly hold jobs while studying, a typical pay for an on-campus job is likely to be at or slightly above the minimum wage, which is set at $7.25 for the country, and $10.50 in California, from where our sample of student participants are chosen.

In addition to these tasks, participants 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-stakes pay. To include risk preference as an additional potentially important control (Cadsby, Song, and Tapon 2009), we also use a risk preference elicitation method established in the literature (Holt and Laury,2002; Harrison, 2002). This involves giving participants $3 with which they can choose to gamble. They then make a series of choices between more-risky 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 complete a series of choices of this nature, a random number generator determines which of their previously-made choices will be used to determine their earnings and then again for whether or not they win.

At the end of the session, a task to be compensated is selected from those the subject completed in the first part of the study and participants are paid the total of their earnings for that task, the show-up payment, and the risk preference elicitation.[6]

3. *Estimation*

[6] We protect individual participants by not including names or identifying information in our data. However, participants are offered the opportunity to sign up to receive a summary of the experiment results at a later date. The names of participants are only recorded in our recruitment database. This database is password-protected, only accessible by Meryl Motika and a student research assistant, and was only used only to track participation and prevent the same participants from participating in multiple sessions of the same experiment. It will not be linked to the results in any way. The non-identifying data are shared between the three researchers and kept on password-protected computers.

After experiments are run, estimation is done using a series of dummy variables for interactions of each type of task and pressure. We also include dummy variables to account for experimental design features that could otherwise impact results. For example, we include dummy variables representing the number of times the participant has completed a different task type (rounds in the experiment) in order to account for learning effects or fatigue, as well as an individual fixed effect to account for the subject's ability level as well as other unobserved time-invariant characteristics. Because each subject performs various task under only one pressure, pressure effects are identified across participants (with pressure randomized), while task effects are identified within-subject.

To estimate the different effects of types of pressures on task types, we first consider a model for a single task (e.g. the routine task only), provided below.

$$Y = \alpha + \beta_1 T + \beta_2 C + \beta_3 H + \beta_4 P + \beta_5 D1 + \beta_6 D2 + \beta_7 D3 + \beta_8 X + \varepsilon \qquad (1)$$

In this equation, Y is the outcome of interest. We use three different outcomes in separate regressions: the total number of problems completed (to measure speed), total number correctly answered (to measure productivity) and the proportion answered correctly (to measure accuracy). Dummy variables represent the effects of the different pressure treatments. T is equal to 1 for our timed goals pressure, C is competitive pressure, H is high-stakes pay and P is piece rate pay. In this specification, a pressure-free flat rate pay is the omitted category. D1 through D3 are dummies representing the number of times the participant has completed a task type (rounds in the experiment) in order to account for learning effects or fatigue. We use separate dummies rather than a scalar to account for the possibility that fatigue or learning effects are nonlinear. Because all comparisons of pressure treatments in this specification are made across-participants, there are no individual fixed effects included in this specification. We instead control for a set of individual characteristics, X, collected in the questionnaire portion of our study, including gender, age, year in school, major, fluency in English, typical weekly spending, and a measure of elicited risk preference. The primary coefficients of interest for the single task analysis are $_1$, $_2$, and $_3$, which represent the effect of each type of pressure. Because this model is task-specific, separate regressions would be run for performance on routine tasks, cognitive problem solving tasks and creative tasks (each separately measured in terms of total completed, total correct and proportion correct). All participants are included in each regression, as each subject performed every task, each randomly assigned to a different pressure.

In order to determine the differential effects of cognitively demanding tasks (cognitive problem solving task or pure creative task) relative to a routine task, changes in accuracy and speed can be measured for a cognitively demanding versus routine task using a difference in differences approach. With a difference-in-differences approach, one would therefore be comparing across individuals' performance under a pressure treatment (e.g. piece rate, competition, timed goal, or high-stakes pay) to other individuals' performance on a comparison pressure (e.g. flat rate) for a given task type, and then comparing this difference in response to pressure calculated for a single task to the same difference in response calculated for a different task type. The outcomes we are interested in are the interactions between type of task and pressure. We observe outcomes at the individual-subtask level.

For example, suppose that in a piece-rate environment the average number of routine problems solved in five minutes is 25, and the average number of cognitive problem solving problems solved is 10. If moving from piece-rate to competitive pay has no effect on routine but reduces the average number of cognitive problem solving problems solved to 8, then this type of pressure might be said to reduce cognitive problem solving speed more than routine task speed. This difference in differences approach that yields relative results for different types of tasks is shown in the model below (we use the example of cognitive problem solving relative to routine tasks).

$$
\begin{aligned}
Y = \alpha + \beta_1 T + \beta_2 C + \beta_3 H + \beta_4 P + \beta_5 PS + \beta_6 PS * T + \beta_7 PS * C + \beta_8 PS * H + \beta_9 PS * P \\
+ \beta_{10} D1 + \beta_{11} D2 + \beta_{12} D3 + F_i + \varepsilon
\end{aligned}
\tag{2}
$$

Regressions of this form would include all outcomes from routine tasks and cognitive problem solving tasks. The new dummy variable *PS* is equal to one if the observation is for the cognitive problem solving task and zero for the routine task. The coefficients of interest for the relative effects of cognitive problem solving in this model are β_6, β_7, β_8 and β_9. These coefficients represent the differential effects of pressure on the cognitive problem solving task compared to the routine task. In this specification, as each subject performs multiple tasks under a single pressure, an individual fixed effect is also included to account for the subject's ability level as well as other unobserved time-invariant characteristics (replacing time-invariant controls, X). To measure the relative effect of the pure creative task relative to the routine task, the dummy for cognitive problem solving (PS) could be replaced with one for the pure creative task, using only outcomes for the pure creative task and the routine task in estimation.

As previously mentioned, we are also interested in whether these task-specific performance effects differ by gender. This can be estimated simply by incorporating a gender interaction with our key variables of interest. For example, equation (1) would become equation (3) below:

$$\begin{aligned} Y = \alpha + \beta_1 T + \beta_2 C + \beta_3 H + \beta_4 P + \beta_5 T*F + \beta_6 C*F + \beta_7 H*F + \beta_8 P*F \\ + \beta_9 D1 + \beta_{10} D2 + \beta_{11} D3 + \beta_{12} X + \varepsilon \end{aligned} \tag{3}$$

Within each task-specific regression, the interaction terms for each pressure type with gender measures the differential performance effect of women relative to men in that pressure. For example, if the estimated coefficient ($_5$) on time pressure (T) interacted with a female dummy (F) were negative, that would indicate that women do relatively worse than men on time pressured tasks. Likewise, we can adapt equation (2) to measure gendered effects, as shown in equation (4) below:

$$\begin{aligned} Y = \alpha + \beta_1 T + \beta_2 C + \beta_3 H + \beta_4 P + \beta_5 PS + \beta_6 PS*T + \beta_7 PS*C + \beta_8 PS*H + \beta_9 PS*P \\ + \beta_{10} T*F + \beta_{11} C*F + \beta_{12} H*F + \beta_{13} P*F \\ + \beta_{14} PS*T*F + \beta_{15} PS*C*F + \beta_{16} PS*H*F + \beta_{17} PS*P*F \\ + \beta_{18} D1 + \beta_{19} D2 + \beta_{20} D3 + F_i + \varepsilon \end{aligned} \tag{4}$$

In this case, the estimated coefficient on an interaction term, such as β_{14}, on the interaction between problem solving, PS, time pressure, T, and female, F, would measure whether changing from a routine to a problem-solving task under time pressure affected women differentially from men.

4. Preliminary Results

At the time of this draft, we have run roughly half of our planned laboratory sessions. Remaining sessions are currently underway. This section therefore summarizes only preliminary results, with smaller sample sizes than will be used in the future. We currently have results from 327 participants, 57 of which were randomized into piece rate pay, 77 to the timed goal incentive, 56 to competition, 61 to high stakes pay and 76 to a neutral flat rate payment. Power tests indicate that current observations are sufficient to detect our desired main performance effects in this study. Therefore, data from the experiment are intended to permit reliable estimation of gender effects (which should be analyzed and added to this paper shortly).

Summary statistics for some of the sample characteristics are presented in Table 2 below. One can see that our sample is relatively balanced in terms of gender of participants, and consists primarily of students in the middle of their university career. In terms of majors, the most common major in our sample is in the category of "math, computer science and engineering".

Table 2: Summary statistics for subject sample

Variable	***Percent (rounded to the nearest whole number)***
Female	55%
Native English speaker	74%
1 or fewer college math classes	20%
Age:	
18-19	31%
20-21	52%

22 or older	17%
Majors:	
Economics or business	14%
Natural Science	11%
Math, computer science, engineering	36%
Psychology	10%
Other social science	9%
Humanities	4%
Arts	1%
Other	15%

Our results from estimation of equation (1), task-specific estimation are presented in Tables 3 - 5. Each table presents results using the full sample of participants and using the neutral flat rate pay as the omitted group for compensation schemes, each done for a different task type. Estimates from regressions for the routine task only are presented in Table 3, for the problem solving task only in Table 4, and for the pure creative task in Table 5. For each outcome - number correct, number completed and accuracy – estimation is done with and without a set of individual controls collected from student questionnaires at the end of the experiment. Results from the estimation of equation (1) are also presented visually in Figure 2 by task type. Figure 2 is constructed using the same estimates as presented in Tables 3-5 from specifications including controls, and are included in the paper for ease of interpretation.

Table 3: Full sample, relative to flat rate pay, routine task only

	correct		done		accuracy	
	1	2	3	4	5	6
Piece rate	0.298	0.259	0.401**	0.388**	-0.109	-0.166
	(0.199)	(0.198)	(0.192)	(0.186)	(0.178)	(0.172)
Timed goal	0.394**	0.364**	0.454***	0.426***	-0.00643	-0.0226
	(0.168)	(0.164)	(0.162)	(0.160)	(0.163)	(0.158)
Competition	0.448**	0.372*	0.588***	0.566***	-0.0868	-0.176
	(0.202)	(0.202)	(0.180)	(0.187)	(0.175)	(0.172)
High stakes	0.302*	0.246	0.497***	0.458***	-0.113	-0.154
	(0.180)	(0.180)	(0.163)	(0.161)	(0.172)	(0.167)
Controls included		X		X		X
Observations	327	327	327	327	327	327
R-squared	0.021	0.041	0.044	0.057	0.006	0.036

Robust se in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Full sample, relative to flat rate pay, Problem solving task only

	correct		done		accuracy	
	1	2	3	4	5	6
Piece rate	0.274	0.214	-0.120	-0.119	0.0643	0.0160
	(0.233)	(0.232)	(0.198)	(0.199)	(0.186)	(0.185)
Timed goal	0.0606	0.0582	-0.0945	-0.113	0.0256	0.0218
	(0.181)	(0.187)	(0.158)	(0.162)	(0.169)	(0.171)
Competition	0.640**	0.591**	0.387*	0.386*	0.196	0.126
	(0.256)	(0.265)	(0.201)	(0.204)	(0.185)	(0.192)
High stakes	-0.166	-0.177	-0.189	-0.201	-0.156	-0.176
	(0.172)	(0.170)	(0.176)	(0.176)	(0.169)	(0.170)
Controls included		X		X		X
Observations	327	327	327	327	327	327
R-squared	0.041	0.054	0.033	0.037	0.014	0.031

Robust se in parenthese Robust se in parentheses
*** p<0.01, ** p<0.05 *** p<0.01, ** p<0.05, * p<0.1

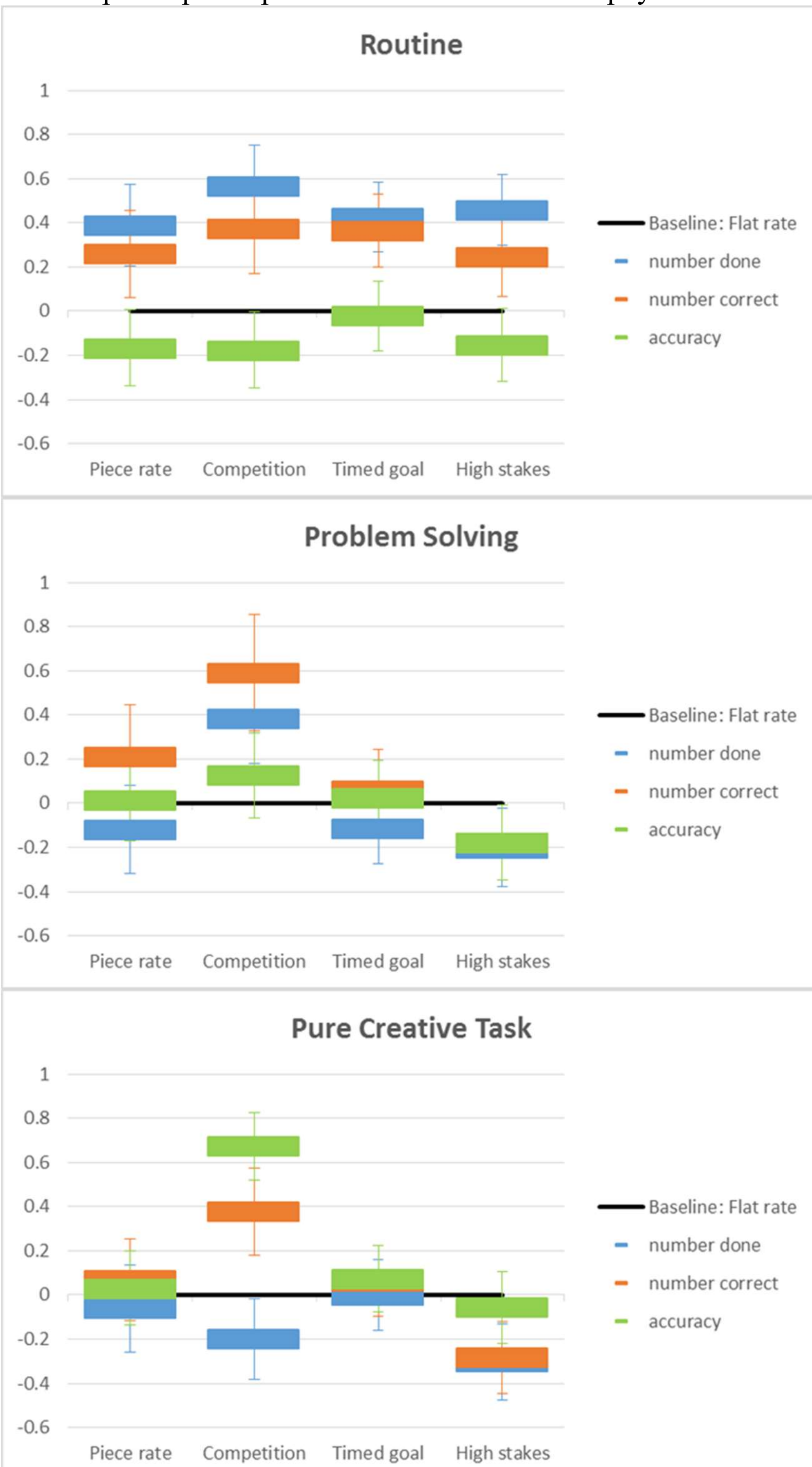
Table 5: Full sample, relative to flat rate pay, creative task only

	correct		done		accuracy	
	1	2	3	4	5	6
Piece rate	0.0522	0.0677	-0.0167	-0.0620	-0.0650	0.0314
	(0.183)	(0.185)	(0.187)	(0.196)	(0.171)	(0.169)
Timed goal	0.0769	0.0627	0.0360	-0.000246	0.0494	0.0754
	(0.159)	(0.159)	(0.164)	(0.162)	(0.153)	(0.151)
Competition	0.419**	0.377*	-0.135	-0.200	0.632***	0.673***
	(0.197)	(0.199)	(0.184)	(0.183)	(0.148)	(0.155)
High stakes	-0.229	-0.283*	-0.247	-0.302*	-0.0612	-0.0568
	(0.167)	(0.164)	(0.177)	(0.173)	(0.169)	(0.164)
Controls included		X		X		X
Observations	327	327	327	327	327	327
R-squared	0.043	0.067	0.013	0.034	0.068	0.121

Robust se in parentheseRobust se in parentheses

*** p<0.01, ** p<0.05 *** p<0.01, ** p<0.05, * p<0.1

Figure 2: Graphs of estimates and error bars corresponding to Tables 3-5, by task type using the full sample of participants and a neutral flat rate pay as the baseline payment incentive



From these estimation results, we are able to draw a number of conclusions regarding the task-specific relative performance effects of incentivizing payment schemes. Our first two results confirm previous findings in the literature, while our second two are new contributions to our understanding of task-specific performance effects.

Result 1: All incentivizing payment schemes improve productivity relative to a neutral flat rate payment scheme for routine tasks.

This is not surprising since the flat rate is designed specifically to apply *no pressure* (and also no incentive), as students receive payments regardless of what they complete. This is consistent with the standard economic theory motivating the widespread use of each of the incentivizing payment schemes that we test in our experiment – namely, that incentivizing a task induces productivity. As previously mentioned, one of the reasons for studying task-specific productivity effects is that, with forces such as automation, the labor market increasingly demands high-skilled work that resembles less the routine tasks (for which this specific result holds) and more tasks that are cognitively challenging and/or involve creativity (for which it does not).

Result 2: High stakes pay reduces performance on non-routine tasks

Contrary to our first result, when performing cognitively challenging tasks, we do see that incentive schemes have task-specific results. For both purely creative and problem solving tasks, participant performance is decreased by high stakes pay (See Figure 2, where the number correct for high stakes pay is higher than the baseline for routine tasks and lower for cognitively challenging tasks). Consistent with previous findings by Ariely, et al. (2009), we confirm that participant performance is decreased by high stakes pay when considering tasks beyond simply very routine tasks.

Result 3: Competition increases performance on both creative and problem solving tasks.

We find that competition has the largest boost in productivity for both types of cognitively challenging tasks studied here – both problem solving tasks and purely creative tasks. There are very few studies that measure the performance effects of competition. As previously mentioned, in a recent study by Erat and Gneezy (2015), using only creative tasks, the authors find that competition fares worse than a piece rate or flat rate payment in resulting productivity. Like their study, we too find no difference between piece rate and flat rate pay for productivity on creative tasks. However, contrary to their findings, we do see a performance boost in creative tasks resulting from competition.

Result 4: Timed goals are not incentivizing on cognitively challenging tasks

For timed goals, despite results similar to competition on routine tasks, participants perform notably poorly on cognitively challenging tasks. Participant performance on problem solving and creative tasks, under timed goals, is similar to having no incentive in place at all. We suspect that timed goals, while providing an incentive also result in a choking under pressure effect large enough to counteract any incentive provided. As we can only observe a final task-specific performance effect and not the degree to which this is driven by a combination of incentivizing forces versus counter-acting pressure, this explanation is merely a hypothesis.

5. *Robustness of results and extensions*

Our results are robust to the inclusion and exclusion of control variables. We also test whether individuals that either did not take the tasks seriously or found them too difficult to be meaningful could be impacting our findings. To this end, we define students who answer nothing or who answer a large number of questions, but get very few correct as "flailers". These students either may find the task too difficult or are not engaging in the experiment for some unknown reason. We run our same specifications as previously presented excluding these "flailers". Results for this test are presented in Table 6 and Figure 3. One can see that our conclusions remain unchanged.

Planned extensions:

- Relative results as defined in equation (2) still need to be run and incorporated into the paper

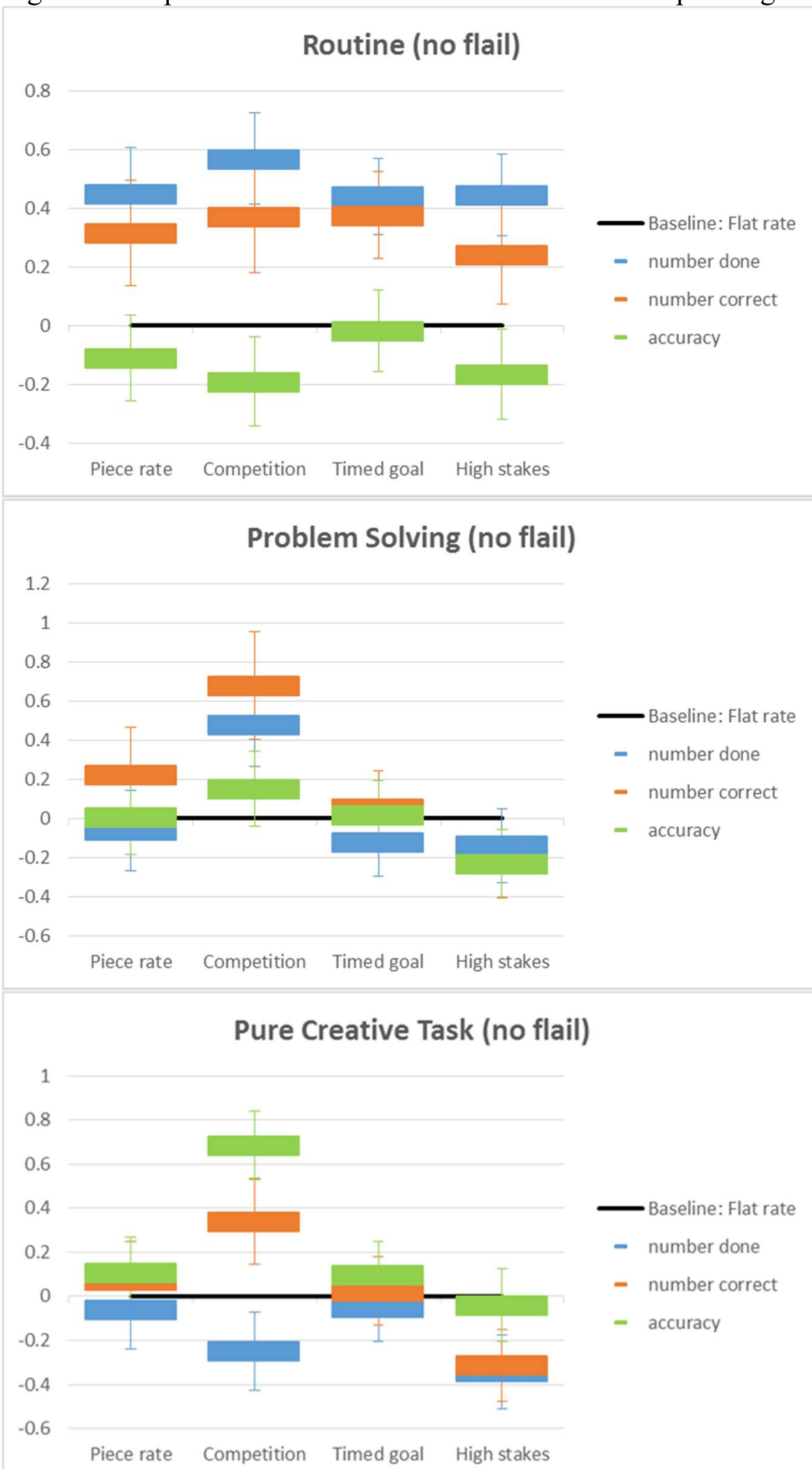
- We currently still have more rounds of our experiment to incorporate into the paper. This will increase our sample sizes and allow us to identify whether there is a gender effect to our task-specific performance responses. We view task-specific estimates especially relevant in the case of competition, which has been more extensively studied in the context of gender labor market outcomes and the gender wage gap. Likewise, we view gendered effects as particularly interesting for other pressure sources, which have not yet been fully explored along this dimension.

Table 6: Full sample excluding "flailers", relative to flat rate pay, all tasks

	Routine task			Problem solving task			Creative task		
	1	2	3	4	5	6	7	8	9
VARIABLES	done	correct	accuracy	done	correct	accuracy	done	correct	accuracy
Piece rate	0.448***	0.316*	-0.110	-0.0612	0.222	0.00636	-0.0586	0.0721	0.109
	(0.159)	(0.180)	(0.146)	(0.207)	(0.243)	(0.188)	(0.183)	(0.178)	(0.160)
Timed goal	0.440***	0.377**	-0.0168	-0.118	0.0526	0.0192	-0.0503	0.0255	0.0960
	(0.131)	(0.147)	(0.137)	(0.178)	(0.192)	(0.175)	(0.155)	(0.155)	(0.152)
Competition	0.569***	0.370**	-0.189	0.482**	0.680**	0.153	-0.248	0.341*	0.686***
	(0.156)	(0.188)	(0.153)	(0.218)	(0.276)	(0.193)	(0.178)	(0.196)	(0.155)
High stakes	0.447***	0.241	-0.166	-0.139	-0.231	-0.228	-0.343**	-0.313*	-0.0388
	(0.139)	(0.169)	(0.153)	(0.187)	(0.176)	(0.174)	(0.168)	(0.162)	(0.164)
Controls	X	X	X	X	X	X	X	X	X
Observations	319	319	319	290	290	290	325	325	325
R-squared	0.080	0.041	0.032	0.041	0.069	0.042	0.044	0.063	0.115

Robust se in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 3: Graphs of estimates and standard errors corresponding to Table 6 (excluding “flailers”)



6. Conclusions

While our results are currently still preliminary, our findings 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. 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. We find competition to be particularly effective as an incentive across task types. On the contrary, time pressure, in the form of timed goals, is remarkably non-incentivizing. On cognitively challenging problem-solving and creative tasks, participants performed no differently than when given no incentive at all.

This is an important finding for designing incentive schemes in education and the labor market, as time pressure is used in many settings under the assumption that it provides a motivating source of pressure. It is quite possible that by implementing a particular incentive-scheme, a company may influence the productivity of its workers, or similarly an educator could alter the performance of his or her students. Additionally, an avenue of research that could build on our study is to explore commonly discussed, yet still not fully understood, labor market concerns, such as wage gaps and inequality. To this end, we intend to expand our study to look at gender differences in task-specific responses to pressure.

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. 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.
5. 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.
6. 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).
7. 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.
8. 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.
9. Dohmen, Thomas J., "Do professionals choke under pressure?," *Journal of Economic Behavior & Organization*, March 2008, 65 (3-4), 636–653.
10. 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.

11. Hocevar, D. "Measurement of creativity: review and critique." *Journal of Personality Assessment*, Oct 1981, 45(5), 450-64.
12. Harrison, Glenn W., "Estimating Individual Discount Rates in Denmark: A Field Experiment,"*American Economic Review*, 2002, 92 (5), 1606–1617.
13. Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *The American Economic Review*, 2002, 92 (5), pp. 1644–1655.
14. 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.
15. 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.
16. 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.
17. 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.