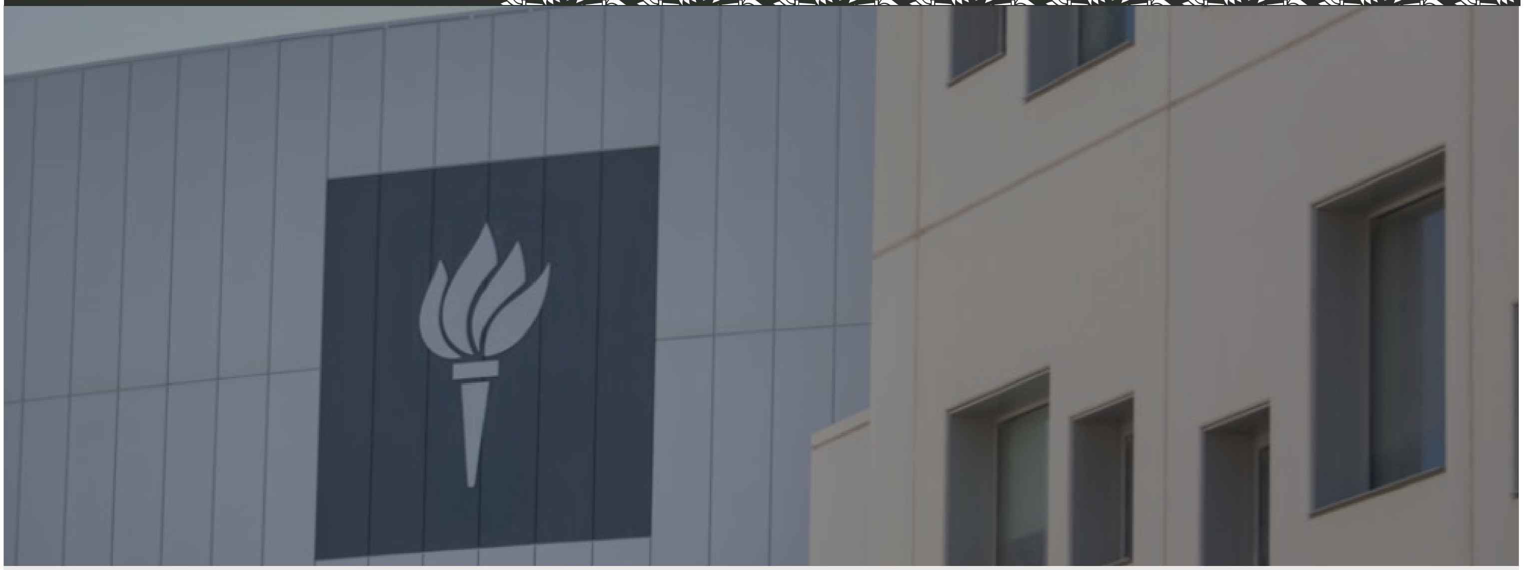
جامعة نيويورك أبوظبي
NYU | ABU DHABI



Creativity under Pressure: Performance Payments, Task Type and Productivity

Joaquín Artes, Jennifer Graves, Meryl Motika

Working Paper # 0028
October 2019

Division of Social Science Working Paper Series

New York University Abu Dhabi, Saadiyat Island P.O Box 129188, Abu Dhabi, UAE

http://nyuad.nyu.edu/en/academics/academic-divisions/social-science.html

Creativity under Pressure:
Performance Payments, Task Type and Productivity*

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

This draft: October, 2019

[a] *Universidad Complutense de Madrid-ICEI,*
Cuidad Universitaria, 28040 Madrid, Spain
and
Social Science Division, New York University, Abu Dhabi, UAE

[b] *Universidad Autónoma de Madrid, Departamento de Economía y Hacienda Pública,*
Cuidad Universitaria de Cantoblanco, 28049 Madrid, Spain
and
Social Science Division, New York University, Abu Dhabi, UAE

[c] *University of California Davis, Center for Educational Effectiveness, 1342 The Grove (Surge III), One Shields Avenue, Davis, California 95616, United States*

---

* This paper is part of a research project generously funded by the Fundación BBVA through the 2015 call for Funds for Individual Researchers and Cultural Creators (Ayudas Individual Fundación BBVA a Investigadores y Creadores Culturales), which is now part of the Leonardo Grants program (Becas Leonardo). We would like to thank both Pedro Rey-Biel and Chris Van Klaveren for feedback. We would also like to thank participants and discussants at various conferences for useful comments and suggestions, including participants at the European Society for Population Economics, European Association of Labour Economists, Association for Public Policy and Management, Association for Education Finance and Policy, and the "Education Skills and Labor Market Outcomes" Workshop hosted by the Department of Economics and Center for Economic Research at Norwegian University of Science and Technology.
† Corresponding author. Email: jennifer.graves@uam.es

ABSTRACT

When incentivizing a worker with performance pay, does the effectiveness of the pay type used vary by the type of task being completed? To answer this question, we run an experiment to test the task-specific productivity effects of various types of performance-based payments, each intended to incentivize productivity. The incentives we use are competition, high-stakes pay, time pressure and piece rate pay, each evaluated against a non-performance-based flat rate payment. Each of these incentives are applied in situations with participants completing three types of tasks: a routine task, a purely creative task and a creative problem-solving task. 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. Our results show that productivity indeed does differ across task type and incentive combinations. We find that, for routine tasks, all incentivizing payment schemes improve productivity relative to flat rate payment. In contrast, for both the purely creative and the creative problem-solving tasks, none of the payment types of piece rate, timed goals nor high stakes pay impact productivity relative to a flat rate payment, with the high pay incentive even decreasing performance on the problem-solving task. We find competition to be the one incentive-based pay scheme that boosts productivity. Participants performed as well or better under competition across all task types, with a notable increase in their performance on pure creative tasks.

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

JEL Codes:
D03, behavioral microeconomics: underlying principles
J24, human capital, skills, occupational choice, labor productivity

1. *Introduction*

In recent decades, labor markets have experienced fundamental shifts in the types of tasks required in the workforce. Fueled by computerization and automation of tasks in the workplace, skill biased technological change has moved work further from mechanical and routine tasks, toward more creative and cognitively challenging tasks (e.g. Autor, Levy and Murnane, 2003, Autor, 2015). Autor (2015) argues that "the interplay between machine and human comparative advantage allows computers to substitute for workers in performing routine, codifiable tasks while amplifying the comparative advantage of workers in supplying problem-solving skills, adaptability, and creativity" (p.5). Additionally, there is ample evidence of important connections between creativity, innovation and growth.[3] As the task content of labor market work evolves, it is worth considering how this might alter currently held assumptions on incentives. Specifically, how do we best incentivize productivity for different task types?

The view still 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. 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, 2001; 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, 2001; Beilock and Carr, 2005; Webb, Williamson and Zhang, 2013).[4] In psychology research, pressures are thought to diminish working memory, potentially reducing productivity.

---

[3] A few examples, among others, are: Lee, Florida and Acs (2004) find that, conditional on a variety of local market area controls, new firm formation is positively related to cultural creativity, as measured by the Bohemian Index or the proportion of artistically creative people in a region. Acemoglu, Akcigit and Celik (2014) present evidence that openness to unconventional and disruptive ideas is associated with more creative innovations. Azoulay, Manso and Zivin (2011) use differences in funding sources in a quasi-experimental setting to show that funding that tolerates early failure and rewards long-term success compared to funding with shorter review cycles and more predefined deliverables resulted in higher impact articles and more novel research.

[4] 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.

Combining the economic and psychology perspectives, the predicted effect of incentive schemes, such as performance-based pay, is ambiguous, as the positive incentivizing effect on productivity could be offset to an unknown degree by the pressure from such incentives altering the connection between incentives and effort and/or between effort and productivity.[5] Additionally, incentives that provide extrinsic motivation may impact performance through altering intrinsic motivation (Eckartsz, Kirchkamp, and Schunk, 2012).

In this paper, we test the task-specific productivity effects of a number of incentivizing payment schemes, each with pay contingent on performance. To do this, we use a large sample of 559 university student participants in an experiment. The sources of potentially pressure-inducing incentives we use in our study are competition, high-stakes pay, time pressure and a piece rate pay, each evaluated against a neutral (i.e. not contingent on performance) 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 task and a creative problem-solving task.

Each type of task type and performance pay method employed in our study resemble those commonly found in today's schools and workplace. Regarding tasks, for our pure creative task we use the established "unusual uses" task, which is quite similar to the act of brainstorming. This task measures divergent thinking. The participant has the objective of generating unique ideas without necessarily solving a pre-determined specific problem or goal. On the other hand, for the creative problem solving task we use a "matchsticks" problem. This task measures convergent thinking, which differs in that there is a goal or specific 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 closely resembles a convergent thinking task, with worker productivity in this area likely being particularly important for success.

The incentives used in our study also reflect real-world performance pay incentives. The high pay incentive, for example, resembles the pressure imposed by performance-contingent

[5] Bradler et al. (2019) find that wage gifts do not increase productivity, which they find is due to uncertainty regarding the connection between effort and production on the part of other individuals.

scholarships in higher education or bonus pay in the business world. Competition is regularly used as an incentive for 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 contract work. Some of these performance-based incentives when applied in real-life work and educational settings may be innate to the task at hand, while others are in place through a decision to do so. To the extent that these incentives are not innate to the job, our study therefore questions whether the incentives used for various tasks are the most effective choice of a payment scheme for incentivizing productivity.

While a number of different types of performance pay incentives and their effects on task performance have been studied in isolation or small pairings, the literature that considers multiple tasks and/or multiple payment incentives in the same context is still very limited. The literature lacks a comprehensive study of types of pressure on different types of tasks within a common context. From a practical standpoint, it would be informative for understanding the design of 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. To our knowledge, this is the first paper to study the interactions between multiple tasks and multiple performance pay incentives in a common framework. In addition, we add to the literature by distinguish between convergent and divergent thinking when studying creative tasks.[6] This is an important distinction in the psychology literature that is missing in most of the economics literature studying incentive schemes in creative tasks.

Our results confirm the presence of task-by-incentive specific performance effects. We show that for routine tasks, all incentivizing payment schemes improve productivity relative to a neutral flat rate payment scheme. We find that this is not generally the case for other tasks. For both the pure creative and the creative problem-solving tasks, neither piece rate, timed goals nor high

[6] Charness and Grieco (2019) use competition relative to flat rate pay to evaluate "open" and "closed" creative tasks. While conceptually similar, these are not synonymous with the convergent and divergent thinking tasks explored in this study. Their distinction regarding the openness of a task is based on whether restrictions are imposed, and the degree of ambiguity in how the task would be evaluated.

stakes pay impact performance, with the high pay incentive even decreasing performance on problem-solving tasks. Of the performance-based incentives, competition is the one incentive-based pay scheme that boosts productivity. Participants performed as well or better under competition across all task types, with a notable increase in their performance on pure creative tasks.

The remainder of our paper is organized as follows: We briefly discuss the most relevant literature in Section 2, then describe our experimental design in Section 3, followed by descriptive statistics of our sample in Section 4, our estimation strategy in Section 5, then we discuss our results in Section 6, followed by concluding remarks in Section 7.

2. *Literature*

Evidence from the field

Some evidence exists on the real-world impact that performance-based pay policies – such as payments determined by competition, high stakes pay, piece rate and time pressure - have in practice. Regarding competition, there is both evidence of its effects on the individual and company level. For example, Gross (2018) studies the effects of competition on individual creative production using existing logo design competitions. He finds that competition boosts production of original work, except under intense competition where individuals actually stop producing. In a review of the literature, Lazear (2018) discusses many examples of competition in the context of company performance, as opposed to individual performance. He notes a number of studies finding that both larger wage gains from promotion and competition between multiple candidates result in increased effort and productivity, measured by increased profits, higher average wages and reduced worker absenteeism (Drago and Garvey 1998, Erikson 1999, Kale, Reis and Venkateswaran 2009, and Heyman 2005, Mobbs and Reheja, 2011). The potential downside of more competition in the workplace in practice is increased volatility, risk taking and uncooperative behavior (Lazear, 2018).

Some examples of high-stakes incentives in practice come from performance-contingent scholarships and accountability policies in education. Performance-contingent funding for higher education has generally been found to be effective at increasing credit accumulation, grades and persistence through college (Patel and Richburg-Hayes 2012, Dynarski and Scott-Clayton, 2013). The evidence on accountability programs, which often use high-stakes testing tied to school consequences, has found such programs to be effective at increasing testing achievement, but has also found some negative consequences such as "teaching to the test" as well as more direct manipulation of testing environments to game the system (Figlio and Loeb, 2011). In the labor market context, Booth and Frank (1999) show that performance-pay, measured as bonuses and commissions, attracts workers of higher ability and induces greater worker effort.

Real world evidence of the impacts of piece rate pay and time pressure on performance is more limited. Evidence on piece rate pay finds that this payment type generally results in increased productivity in the field, although most examples come from relatively routine tasks, such as agricultural harvesting (Moretti and Perloff 2002), car window replacement (Lazear 2000) and tree planting (Shearer 2004). Time pressure, in practice, most resembles the idea of hurried work with tight deadlines. While there is evidence that this component of work increased during the 1990s with more intense use of technology in the workplace (Green and McIntosh, 2001), little field research exists that evaluates the impacts of time pressure on productivity.[7]

Experimental Evidence

In experimental settings, 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. 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

[7] Kaur, Kremer and Mullainathan (2010) and Ariely and Wertenbroch (2002) both find that deadlines can work as desirable self-enforcement mechanisms for finishing work. However, this type of deadline is not the time pressure that we refer to, but instead is more applicable in the context on long-term projects, where procrastination could be a concern.

stressors. 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. Much of this literature separately studies a single task type. For example, Eckharts et al. (2012) finds inconclusive evidence for creative problem solving using performance incentives of a flat fee, piece-rate and competition. Bradler, Neckermann and Warnke (2019), using a divergent thinking creative task, find positive impacts of competition and no impact of wage gifts on creativity. Erat and Gneezy (2016) find that, compared to no payment, providing a piece-rate incentive did not improve productivity, and that competition reduced productivity on a creative problem-solving task.[8] 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.279).

A few experimental studies have considered multiple tasks in the same context. This 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. We model our use of high stakes pay closely off of this study. Our study differs in that the only type of creativity explored in Ariely et al (2009) is creative problem solving, while we also test another type of creativity in the high stakes context (pure creativity, measured through the divergent thinking task of "unusual uses"). We find this to be an important distinction. A study by Charness and Grieco (2019) also finds that competition has task-specific effects. Specifically, they test two variations of creativity, which they define as "open" and "closed" based on the degree of ambiguity in the method of evaluation. The open task is designed to be intentionally vague, while their closed task is more defined. They find that competition relative to a flat rate pay increases performance on the closed task, but not on the open task. While these definitions have much in common with convergent and divergent

[8] Erat and Gneezy (2016) use no payment as their neutral reference category instead of flat rate pay (as done in this study). While we find that productivity-increasing effects of competition and they find productivity-decreasing effects, these are not necessarily incompatible as we compare two types of extrinsic motivation (competition and flat rate) while they compare a case of extrinsic motivation (competition) to intrinsic motivation (no pay).

thinking, they are not the same. Our pure creative task is the commonly used divergent thinking task of "unusual uses." This type of brainstorming is not however vaguely defined, as participants are likely to be familiar with the concept and evaluation is also defined. In this sense, both of our tasks would fall under Charness and Grieco's (2019) definitions of closed tasks. Within this context where evaluation of both creative tasks is clearly defined, we find notable task-specific responses to performance pay between our creative problem solving task and our pure creative task (traditional convergent vs. divergent thinking tasks).

In our approach, we extend both the single-task literature and this smaller multiple-task literature by exploring, in a common and comparable context, the performance effects of pairings of multiple tasks and the various performance-based payments that are intended to incentivize performance on such tasks. This allows us to draw further comparisons than have been possible in previous studies. While our paper contributes to the literature through estimating the task-specific productivity effects of multiple performance pay types, which has not been done in the literature before, it is notable that on the subsets of tasks and/or incentives that we have in common, our findings are also generally consistent with the results of both Ariely et al. (2009) and Charness and Grieco (2019).[9] To our knowledge, these studies are currently the extent of literature exploring multiple tasks within a common framework. Therefore, after we present our results (Section 5), we discuss further how our findings extend upon the findings of these two studies, but also provide nuance to the larger conclusions that one would draw from this still small literature (Section 6).

3. Experimental Design

For our experiment, participants were recruited from the behavioral economics laboratory at the University of California - Irvine (UCI), where we used the Experimental Social Sciences

[9] Ariely et al. (2009) finds reduced productivity under high stakes pay for non-routine tasks, which we also find for problem-solving tasks. Charness and Grieco (2019) find positive effects of competition on creativity under certain conditions, where the type of task matters.

Laboratory (ESSL) subject pool. This subject pool came from a list of UCI students over the age of 18 who signed up to participate in research. They were invited by the ESSL staff in charge of recruitment and subject pool management, and signed up through the ESSL website. Each session took place in a computer laboratory. Upon arrival at the session, participants were asked to read and sign the informed consent form and had an opportunity to ask questions about the study. They were then seated at a computer. Initial instructions provided participants with a general description of the experiment and information on payments. At the end of the session, participants were paid the show-up fee plus their earnings in one task, which was randomly selected from those that they had completed. Paying for only one task removes the possibility of wealth or hedging effects on behavior, as it would be detrimental to measuring our intended effects if participants became less risk-averse as the session continued because they had already earned significant amounts of money.

Each participant completed three different types of tasks: routine, problem solving and pure creative. In our experiment, we assigned five different methods of payment: flat rate, piece rate, competition, time pressure, and high-stakes pay. We assigned all tasks to each subject, but randomly assigned 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.[10] Both the payment type and the order of the tasks assigned were randomized. After learning about the payment method assigned to them, participants then read instructions and completed several practice questions for their first task. After completing the first task, they read instructions and then completed practice questions for the second and third task respectively. Sample screenshots of what student participants saw on the screen in the experiment environment are provided in the appendix to this paper. At the end of the session, their payment was determined (through random selection of one of the previously completed tasks). Each subject was then paid individually and privately before leaving the room. Each

[10] We could have alternatively assigned tasks randomly across participants. However, in practice this would have required participants to perform the same task under different incentives, which presents concerns about learning across rounds. Additionally, in a pilot round of the experiment, students anecdotally expressed more difficulty in mentally separating the change in incentive type they were working under than a change in task type.

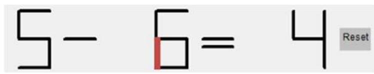
session lasted approximately 60 minutes and no follow-up visits occurred.

2.1 Tasks performed

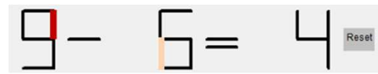
As previously mentioned, each participant completed all three tasks –a routine task, a creative problem-solving task, and a pure creative task - under a single incentive scheme. This distinction between creative problem solving and pure creativity (convergent vs. divergent thinking) is important in the psychological literature on creativity (Hocevar 1981; Byron, et al., 2010). Both of these creative tasks are quite different from routine tasks, which do not require such "outside" thinking.

The routine task in our study was completing as many problems as possible in five minutes, where the problems consisted 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. This task is a standard "letter search task" used in psychology studies. It can be considered a simplified form of the "encryption" task originating from Erkal et al. (2011), also commonly used by experimental economists as a measure of real-effort. While there are many small variants to this type of task in the literature, we like this one for its simplicity – it is clearly routine - as well as that it mixes both letters and counting, making it neither clearly a word problem nor clearly a math problem.

The cognitive problem solving task consisted of accurately completing as many matchstick questions as possible in five minutes. These questions are drawn from the psychology literature on convergent thinking, and have been used and tested in studies such as Knoblich et al. (1999). Chu and MacGregor (2011) in a review of research on insight problem-solving performance noted matchsticks problems as one of the promising new methods. An example of this task is also presented in Figure 1. The matchstick question involved an incorrect math equation, where numbers were written in "block letter" form, where each line was conceived as a matchstick that can be moved around. The participant was 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. Visually, this would be moving the "stick" highlighted in red below:



changes to



We modify the standard matchstick problem from using roman numerals to use Arabic numerals in the straight-lines form used on digital clocks. As "chunk decomposition" style problems such as these are generally considered quite difficult,[11] we viewed this formatting as easier for our student participant pool.

The pure creative 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. This method, developed by Torrance (1974), is interesting enough to engage subjects in the task (Shalley and Oldman, 1997), and has been widely used by researchers as a divergent thinking task (e.g. Bradler et al., 2019; Dutcher, 2012; Dutcher and Rodet, 2018; Eisenberger, Armeli and Pretz, 1998). In this task, participants are asked to find as many non-standard uses as possible for a household item, such as an umbrella. This task is similar to the idea of brainstorming new ideas. 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.

In scoring performance on these task, we calculated three separate measures: the total number of problems completed (to measure speed), the total number correctly answered (to measure productivity) and the proportion answered correctly (to measure accuracy). In the case of both the routine and problem-solving task, this is straightforward. In the case of the purely creative task, unusual uses, the number written down is used as the number completed, while the number answered correctly does not consider filled in responses that are either too similar to the standard

[11] We do observe that the matchsticks problem is the most difficult for students. Histograms of the number completed by task along with discussion are provided in the Appendix to this paper (Figure A1).

use of the object, is repetitive to another answer, or is implausible. Considering the example in Figure 1 of a car tire, examples of incorrect responses might be listing "to drive" (original use), listing both "as a flotation device" and "to float in the water" (would receive 1 point, but not two as it is repetitive), and listing "use as a balloon" (not possible).[12]

3.2 Payment Schemes (Incentives or Pressures)

Under the piece rate payment scheme, participants earn money for each correct response. Under competition, participants are placed 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 relatively 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. 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, et al. (2009).

Table 1 reports the break-down of payments. 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) were designed to approximate an average hourly wage of $10. For example, in the United States, where university students commonly hold jobs while studying, a

[12] While we had to grade the "unusual uses" task on site if they were randomly selected for payment in order to pay such participants at the end of the session, we later graded all unusual uses outcomes (whether selected for payment or not) using a file that did not list any other experiment details, so that grading would be blind to treatment. No notable differences were observed between those scored during the experiment and their scores from later blind scoring of responses.

typical pay for an on-campus job is likely to be at or slightly above the minimum wage, which at the time of this study was set at $7.25 for the country, and $10.50 in California, from where our sample of student participants are chosen. We use Ariely et al. (2009) as a reference for setting the high payment amounts used in this study.[13] For other payment methods, we set the expected average payouts to $10 for all task types. This can be directly imposed for the case of flat rate and competition pay, where payouts are determined before the experiment and do not depend on unknown performance levels during the experiment. For the other methods, levels were set based on how we expected students to perform. While actual average payments will naturally differ somewhat from the exact $10 average (as they depend on the outcome of the experiment), what is important for comparability across incentives is that students would expect to do similarly. Our expectations regarding how many problems a student would complete for each task type (e.g. for piece rate), as well as for setting medium and high goals for the number completed under timed goals and high pay were drawn from a pilot study on a separate university student sample. Payment levels were then set accordingly to achieve the same expected average payouts for all task types.

From Table 1, one can also see that the goals set vary by task type in a way that aligns with their difficulty level. For example, the problem solving task is a more difficult task and therefore has lower goals regarding the number completed, with only 2 correct to achieve the medium goal and 4 correct to achieve the high goal, compared to 14 and 18 correct to meet the same goals for the routine task (also see the appendix, Figure A1). The higher difficulty level of this task, as previously noted, is expected. Likewise, in real-world workplace settings, the most relevant problem solving skills would likely be for notably challenging problems. In Table 2, we report the means and standard deviations of the number done and number correctly completed by each task and pressure. From Table 2, we can see that the goals for the pure creative task (7 correct for the medium goal, and 11 correct for the high goal) were a bit easier for student participants to achieve in the experiment than for the other task types, with more students able to achieve the

[13] Anecdotally, students expressed both visually and verbally their excitement at receiving the high payment amounts, confirming our belief that the high payment amounts would be considered a large payment for this student population.

high goal (detailed percentages, by goals, tasks and incentives are reported in the appendix, Table A1).

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 include risk preference as an additional potentially important control (Cadsby, Song, and Tapon, 2016), 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.[14]

4. *Sample Descriptive Statistics*

Our final sample consisted of 559 participants, comprised of students over the age of 18 at the University of California Irvine. Table 3 summarizes characteristics of our sample by treatment as well as overall. In the last column of the table, the overall sample characteristics are reported. One can see that our sample is relatively more female than male, which likely reflects both the higher proportion of females in higher education (Goldin et al. 2006) as well as more females

[14] We protected 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 were only recorded in our recruitment database. This database was password-protected and was used only to track participation and prevent the same participants from participating in multiple sessions of the same experiment. It was not linked to the results in any way.

participating in laboratory experiments at the ESSL laboratory. Roughly 40% of the sample have taken numerous math courses at the university level. In terms of majors, the two highest categories are evenly split between "social sciences" and "math, computer science and engineering". A fairly high proportion (78%) of the participants speak a second language at home. We suspect that this is largely driven by the relatively large Hispanic and Asian populations at UCI.[15]

To assess balance of characteristics across the randomly assigned incentives, we also report summary statistics by incentive in the separate columns of Table 3. T-tests are run for statistically significant differences in characteristics between the control (flat rate, reported in the 5th column) and each performance-based incentive treatment (piece rate, time pressure, competition, and high incentive, reported in columns 1-4). While some statistically significant differences are reported (by asterisks placed next to means in columns 1-4), most are not meaningful and they do not affect results. In general, the sample appears to be balanced across the randomly assigned treatment categories.

5. *Estimation*

To estimate the different effects of types of incentives 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 (speed), total number correctly answered

[15] For example, see: "https://www.forbes.com/colleges/university-of-california-irvine/" (last viewed: 3/26/2019). Also, "https://nces.ed.gov/programs/digest/d17/tables/dt17_306.30.asp?current=yes" for Table 306.30, "Fall enrollment of U.S. residents in degree-granting postsecondary institutions, by race/ethnicity: Selected years, 1976 through 2027" National Center for Education Statistics.

(productivity) and the proportion answered correctly (accuracy). Dummy variables represent the effects of the different performance-pay incentive and/or pressure treatments. T is equal to 1 for timed goals, C is competition, H is high-stakes pay and P is piece rate pay. In this specification, a non-performance-based flat rate pay is the omitted category. D1 through D3 are dummies representing the number of times the participant had 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.[16]

We control for a set of individual characteristics, X, collected in the questionnaire portion of our study, such as gender, year in school, math courses taken, major dummies, speaking a second language, spending and savings behavior and a measure of elicited risk preference. The primary coefficients of interest for the single task analysis are β_1, β_2, β_3 and β_4, which represent the effect of each type of performance-pay.

As this model is task-specific, separate regressions are 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 (559) are included in each regression, as each subject performed every task. Identification of the different incentives within the task-specific regression comes from across-participant comparisons, as each person was randomly assigned to a different payment method. All results are reported in standard deviation units to permit meaningful comparisons across results for different tasks.

6. Results

In this section, we discuss our main results, all of which come from estimation that includes a set of individual controls and major dummies (presented in Table 4 and Figure 2).[17] However, it is

[16] See Appendix Table A2 for evidence that task order does not have any impact on performance.
[17] Controls include: female, upperclassman, many math courses taken, risk preference elicitation index, non-native English speaker, whether they speak a second language, whether they have low levels of savings at the end of the month, whether they have relatively low levels of monthly spending and a dummy for pursuing a double-major. Major dummies include: Math or Engineering, Sciences, Humanities and Arts, Other major, with the omitted

first worth noting that none of these results change with the inclusion or exclusion of any set of controls (see appendix tables A3-A5 and figures A2-A4).[18] Similarly, we test for the possibility that a sub-group of students might disengage from the study and could therefore impact our findings. To this end, we define incidences when students answer nothing or answer a large number of questions, but get very few correct as “flailing”.[19] These students either may find the task too difficult or are not engaging in the experiment for some unknown reason. We also run our same specifications excluding these “flailers”. Results for this test are presented in Table 5, with additional results excluding flailers provided in the appendix. Our conclusions remain unchanged. The set of results presented and discussed in this section are therefore highly robust to alternative specifications.

5.1 Main findings

Results from estimation of equation (1) are reported in Table 4 and presented visually in Figure 2. In Table 4, results for the outcomes of number done, number correct and accuracy are reported for each task type. Each column therefore presents results from a separate regression using the full sample of participants. In column 1, for example, the outcome used in estimation is the number of routine task problems done, while in column 2, the outcome used in estimation is the number of routine task problems answered correctly and in column 3 the outcome used in estimation is a measure of accuracy on the routine task. In this same way, results for the routine task are reported in columns 1-3, for the problem solving task in columns 4-6 and in the pure creative task in columns 7-9. The estimated coefficients for incentive-pay dummies are reported,

---

category of Social Sciences.

[18] In the appendix to this paper, we report results repeated without any controls (Table A3), with demographic controls but no major dummies (Table A4) and a fully saturated model including demographics, major dummies and 36 separate session dummies (Table A5). Only in the fully saturated model do any estimates remain similar but loose significance, and even so, only for high stakes pay. Therefore, the inclusion or exclusion of controls do not change our overall findings in this study.

[19] A student can flail on one task, but not on others. We define the student as a “flailer” if they flailed on any one task and exclude these students from estimation using any task to maintain the same sample across specifications. Appendix Table A6 reports the percent of flails by pressure and task type pairs. Flailing occurs predominantly on the more difficult problem solving task at 11%, with lower incidence on the other two tasks (1%-2%).

where the omitted category (control group) is the non-performance-based flat rate pay. Figure 2 is constructed using the same estimates as presented in Table 4, and are included in the paper simply due to the ease of interpretation. From Figure 2, the main conclusions can easily be seen at a glance.

One can see from the first panel of Figure 2 (columns 1-3 of Table 4), that for the routine task, all incentivizing payment schemes improve performance, or productivity (measured as the number correct), relative to the neutral flat rate payment. In this case, the results align with the standard economic model of incentives. We can see that both the total number done and number correct are higher with performance-based pay, with no corresponding statistically significant difference in accuracy. In the presence of performance-based pay, participants simply work faster, successfully completing more.

The second panel of Figure 2 (columns 4-6 of Table 4) presents results for the creative problem-solving task. We now instead see that incentives have no measurable difference in performance, except for a choking effect under high payment levels. In this case, we hypothesize that the lack of statistically significant effects found for other performance-based payments is less likely to be driven by such payments having no incentivizing impact on students, and more likely driven by any positive incentivizing impact being outweighed by whatever pressure the incentive also imposes. We cannot however directly test this hypothesis with our experimental design.[20] Nonetheless, there are reasons to believe this to be the more likely explanation of our findings. As previously noted, the hourly payment is comparable to hourly wages for our participant sample and our results are unchanged when "flailing" students, or those who we suspect might disengage from the study are excluded. Additionally, behavioral impacts on piece rate and timed goals do occur; we see that participants work slower (statistically significant fewer done), just not so much so as to have an impact on their overall productivity.

[20] Future work could extend our study to directly ask participants about how incentivizing, as well as how stressful they found the payment type they were assigned. Stress responses could be further measured using biological measures (e.g. eye movements, heart rate or sweat) that can be measured in a laboratory setting.

Lastly, results for the purely creative task are presented in the third panel of Figure 2 (columns 7-9 of Table 4). In this case, we see that competition increases performance on this purely creative task. Competition results in an increased number of correct responses that appears to be driven by increased accuracy rather than an increase in the number done. Recall that this task resembles brainstorming. Therefore, this could suggest that participants focused on higher quality brainstorming (innovative responses) rather than producing quantity.[21]

Taken together, we find robust evidence of task- and incentive-specific productivity effects. We see that incentives work for routine tasks. However, for both the pure creative and the creative problem solving tasks, neither piece rate, timed goals nor high-stakes pay impact productivity relative to a flat rate pay scheme. What is also clear is that competition is the weakly dominant incentive-pay scheme across task types. Competition performs similar to other performance-based schemes relative to flat rate pay on routine tasks, performs no worse than any other payment scheme on problem-solving tasks (or slightly better, although insignificant) and significantly outperforms all other incentives on the pure creative task.

6. Discussion

A few things are particularly notable of our findings in the context of the existing literature. Firstly, we do see that 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. However, a primary reason for studying task-specific productivity effects is that, with changing labor market 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

[21] The main effect of competition increasing productivity is robust across specifications. However, this interpretation of the mechanism should be taken cautiously as it holds up in most, but not all specifications.

and involve creativity (for which we find that it does not).

Secondly, many of our findings for specific tasks or incentives are consistent with the most related existing multi-task studies in the literature: Ariely et al (2009) and Charness and Grieco (2019). 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. Although this specific finding in our paper is not as pronounced or robust as found in Ariely et al. (2009), it can be seen in comparing productivity on the routine task to productivity on the problem-solving task. We do not see this choking effect of high stakes pay for the pure creative task. This type of task was not studied in Ariely et al (2009) where creativity was tested only using a problem-solving task. The main conclusion in Ariely et al. (2009) in that choking under high stakes pay does not occur under very mechanical tasks, but occurs across a variety of tasks requiring concentration. Our finding adds nuance to this broader conclusion, as we find that for our pure creative task, which measures divergent thinking, does not suffer from this problem.

Ariely et al (2009) posit various mechanisms through which choking under pressure on high concentration tasks might occur, including both the possibility that (i) the high incentive increases arousal to a sub-optimal level for performance, and (ii) that the incentive causes a narrowing of focus which might limit insight. Considering these same mechanisms, in the first case, it may be that generating a list of unusual uses, as well as similar divergent thinking tasks that involve a flow of new ideas, simply requires a higher level of arousal than creative problem solving tasks. If high-stakes pay raises arousal more than other performance-pay incentives, it would not necessarily produce the choking effect for the pure creative task as was observed for the creative problem solving task. Considering the second proposed mechanism, where the narrowing of focus under pressure is to blame, our differing effects by type of creative task could be explained by Dual Process Theory (e.g. Stanovich and West, 2000, Kahneman, 2011). The creative problem solving task, with its single solution as a goal, triggers one to think in a controlled, slower and concentrated manner, while the open generation of ideas under our pure creative task uses a faster more automatic and free-flowing process. This distinction could explain why the creative problem solving, which involves a cognitive process characterized by focus, would be susceptible to faltering when overly narrowed focus limits insight, while the pure creative task is unaffected, as it uses a distinct cognitive process.[22]

The other research paper that serves as a main reference on evaluating financial incentives using multiple tasks is Charness and Grieco (2019). They also draw the distinction between different types of creativity under the performance-pay incentive of competition. However, the contribution of our paper in this regard is different than theirs because we focus on different types of creativity. Charness and Grieco (2019) define their two types of creative tasks as "open" versus "closed." While conceptually similar, these are not synonymous with the convergent and divergent thinking tasks explored in this study. Their distinction regarding the openness of a task is based on whether restrictions are imposed, and the degree of ambiguity in how the task would be evaluated. Given their definition, both our creative problem solving task and our pure creative task would fall under their definition of a closed task, since the objective and metrics of evaluation are defined. Charness and Grieco (2019) find that competition works to increase creativity under "closed" creativity, but does not do so for "open" creativity. However, our results show that this is not the end of the story. Within a context where evaluation of both creative tasks is clearly defined, we show that there are differences in how competition affects performance on creative problem solving tasks (convergent thinking) versus pure creative tasks (divergent thinking). Specifically, competition does not make a notable difference (positive, but statistically insignificant estimate) for creative problem solving tasks, but does increase productivity for pure creative tasks (positive and statistically significant estimate).

Naturally, the real world implications of our results are therefore also different from Charness and Grieco (2019). They provide examples of their closed creativity as "finding a way to decrease the size of a computer or developing a new drug for a specific purpose" and their open creativity as work on "an abstract painting" (pp.455-456). Our types of creativity relate less to the distinction between the art world versus research and development in the business world, but rather would be more directly applicable to understanding how incentives impact the different

[22] This possibility of Dual Process Theory as a potential explanation for incentives to differentially impact different forms of creativity was also discussed by Charness and Greico (2019).

types of tasks likely to be employed at different stages of the research and development process within the business and research context.[23]

Lastly, our findings are also notable in the context of the current literature due to our (possibly) surprising lack of impacts found for *other* performance-based pay types besides competition. To our knowledge, this is the first study to experimentally test the task-specific performance effects for these various performance pay incentives within the same experiment.[24] Despite results similar to competition on routine tasks, we do not find evidence that participants improve under these alternative incentives when completing cognitively challenging tasks. Participant performance on problem solving and creative tasks under other incentives are not distinguishable from having no performance-based incentive in place at all. We hypothesize that such pay types likely produce an incentivizing effect, but also 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. Nonetheless, this finding could also be interpreted as reflecting positively on the reference payment group of flat rate pay, which is stable pay that is not tied to performance outcomes, and yet produces similar outcomes to incentives that are more directly tied to performance.

7. Concluding Remarks

Our findings provide a more complete look at how students and workers respond to incentives. Each of the three types of tasks reflect common tasks required in today's schools and workplace. Additionally, each potentially pressure-inducing performance-pay incentive we tested are

[23] It is worth noting that our context for studying divergent differs from Charness and Grieco (2019) in two additonal key ways: first, our more "open" task measures brainstorming, as opposed to a final product, and second, our scoring of the task does not involve the social component of work being judged by a panel for creativity. These differences in context could also contribute to some of the differences in our findings.

[24] In addition to their experimental evidence of the effects of competition on creativity, Charness and Grieco (2019) also present a model regarding potential effects of alternative performance pay methods, but do not experimentally test such performance pay methods directly. As it was developed simultaneously to our study, we did not use their theoretical framework for developing our experimental design. Nonetheless, their theory serves as motivation for the need for our experimental evidence in the larger literature.

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. We find competition to be particularly effective as an incentive across task types. It should be noted that the form of competition used in this study was anonymous; participants could not identify their competitors. There is evidence that factors such as group composition and information provided on performance can matter for performance under competition (e.g. Booth and Nolan, 2012, Iriberri and Rey-Biel, 2017, Niederle and Vesterlund, 2007, Siddique and Vlassopoulos, 2017). Therefore, further experiments testing variations to the competition setting could prove fruitful for understanding the contexts in which our findings can be best applied.[25] Our understanding of how to apply these results in real-world settings could also further benefit from exploring potential heterogeneity across demographic groups in society (e.g. race or gender) in task-specific responses to pressure.[26] This is especially relevant considering the abundance of relatively recent studies showing that many of the individual performance-pay incentives used in this study exhibit gender differences on certain task types.[27]

Our results provide evidence to consider when designing incentive schemes in education and the labor market, as these incentives are used under the assumption that it provides motivation for performance. It is quite possible that by implementing a particular performance pay incentive, a company may influence the productivity of its workers, or similarly the form of educational funding could alter student performance. The findings of the current study are supportive of the use of competition to incentivize productivity on multiple types of tasks, both creative and routine. While other performance-based pay incentives did no better (or in some cases worse)

---

[25] As in many similar experimental settings, we also limit students' ability to distract themselves with internet or their phone. Goerg et al (2019) find that performance effects differ when implicit effort costs, such as ability to use the internet or leave early, are introduced. To the extent that the work or school environments freely allow such distractions, our results may not generalize as well to those contexts.

[26] Current work by the authors is underway to evaluate gender differences in task-specific performance effects of performance-pay incentives.

[27] See, for example, Azmat, Calsamiglia and Iriberri (2016) who study gender differences under high stakes pay and Shurchkov (2012) who studies gender differences in performance (both quality and quantity) under competition and time constraints. Other examples of recent research on gender differences by incentives and their role in the workplace include: Delfgaauw, Dur, Sol, and Verbeke, 2013; De la Rica, Dolado, and Vegas, 2010; Grosse and Gerhard, 2010; Jirjahn and Stephan, 2004; Kangasniemi and Kauhanen, 2013 and Lavy, 2013.

than a flat rate payment, we find competition to improve productivity across all settings.

References:

1. Acemoglu, Daron, Ufuk Akcigit and Murat Alp Celit (2014). “Young, Restless and Creative: Openess to Disruption and Creative Innovations.” NBER Working Paper # 19894.
2. Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar (2009). “Large Stakes and Big Mistakes.” *Review of Economic Studies*, 76 (2), 451–469.
3. Ariely, Dan and Klaus Wertenbroc (2002). “Procrastination, Deadlines, and Performance: Self-control by Precommitment.” *Psychological Science*, 13(3), 219-224.
4. Azmat, Ghazala, Caterina Calsamiglia and Nagore Iriberri (2016). “Gender Differences in Response to Big Stakes.” *Journal of the European Economic Association*, 14(6), 1372-1400.
5. Azoulay, Pierre, Joshua S. Graff Zivin, and Gustavo Manso (2011). “Incentives and Creativity: Evidence from the Academic Life Sciences.” *RAND Journal of Economics*, 42(3), 527-544.
6. Autor, David H. (2015). “Why Are There Still So Many Jobs? The History and Future of Workplace Automation.” *Journal of Economic Perspectives*, 29(3), 3–30
7. Autor, David H., Frank Levy and Richard J. Murnane (2003). “The Skill Content of Recent Technological Change: An Empirical Exploration.” *The Quarterly Journal of Economics*, 118(4), 1279–1333.
8. Baumeister, Roy F., and Carolin J. Showers (1986). “A review of paradoxical performance effects: Choking under pressure in sports and mental tests,” *European Journal of Social Psychology*, 16, 361-383.
9. Beilock, Sian L. and Thomas H. Carr (2005). “When High-Powered People Fail,” *Psychological Science*, 16 (2), 101–105.
10. Booth, Alison L. and Jeff Frank (1999). “Earnings, productivity, and performance-related pay,” *Journal of Labor Economics*, 17, 447–63.
11. Booth, Alison and Patrick Nolan (2012). “Choosing to compete: How different are girls and boys?” *Journal of Economic Behavior & Organization*, 81, 542–555.
12. Bradler, Christiane, Susanne Neckermann, and Arne Jonas Warnke (2019).

"Incentivizing Creativity: a Large-Scale Experiment with Tournaments and Gifts," *Journal of Labor Economics*, 37 (3), 793-851.

13. Byron, Kristin, Shalini Khazanchi, and Deborah Nazarian (2010). "The relationship between stressors and creativity: A meta-analysis examining competing theoretical models," *Journal of Applied Psychology*, 95 (1), 201–212.
14. Cadsby, C. Bram, Fei Song, and Francis Tapon (2016). "The Impact of Risk Aversion and Stress on the Incentive Effect of Performance Pay." *Experiments in Organizational Economics* (*Research in Experimental Economics, Vol. 19*), Emerald Group Publishing Limited, pp. 189-227.
15. Charness, Gary and Daniela Grieco (2019). "Creativity and Incentives," *Journal of the European Economic Association*, 17(2), 454-496.
16. Chu, Yun and James N. MacGregor (2011). "Human Performance on Insight Problem Solving: A Review," *Journal of Problem Solving*, 3(2), 119-150.
17. Dandy, Justine, Neil Brewer, and Robin Tottman (2001). "Self-Consciousness and Performance Decrements Within a Sporting Context," *The Journal of Social Psychology*, 141 (1), 150-152.
18. Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke (2013). "Tournament Incentives in the Field: Gender Differences in the Workplace," *Journal of Labor Economics,* 31 (2), 305-326.
19. De la Rica, Sara, Juan J. Dolado, and Raquel Vegas (2010). "Gender gaps in performance pay. New evidence from Spanish workers." *Annals of Economics and Statistics,* No. 117/118, SPECIAL ISSUE ON THE ECONOMICS OF GENDER, pp. 41-59.
20. Dohmen, Thomas J. (2008). "Do professionals choke under pressure?" *Journal of Economic Behavior & Organization*, 65 (3-4), 636–653.
21. Drago, Robert, and Gerald T. Garvey (1998)."Incentives for Helping on the Job: Theory and Evidence." *Journal of Labor Economics,* 16(1): 1–25.
22. Dutcher, E. Glenn (2012). "The effects of telecommuting on productivity: An experimental examination. The role of dull and creative tasks." *Journal of Economic Behavior & Organization*, 84(1), 355-363.

23. Dutcher, E. Glenn and Cortney S. Rodet. (2018). "Which Two Heads are Better than One? Uncovering the Positive Effects of Diversity in Creative Teams." Unpublished manuscript.
24. Dynarski, Susan and Judith Scott-Clayton (2013). "Financial aid policy: Lesson from research." *The Future of Children*, 23(1): 67-92.
25. Eisenberger, Robert, Stephen Armeli, and Jean Pretz (1998). "Can the promise of reward increase creativity?" *Journal of Personality and Social Psychology, 74*(3), 704-714.
26. Eckartz, Katharina and Kirchkamp, Oliver and Schunk, Daniel (2012). "How Do Incentives Affect Creativity?" CESifo Working Paper Series No. 4049, CESifo Group Munich.
27. Erat, S., & Gneezy, U. (2016). Incentives for creativity. *Experimental Economics*, *19*(2), 269-280.
28. Eriksson, Tor (1999). "Executive Compensation and Tournament Theory: Empirical Tests on Danish Data." *Journal of Labor Economics* 17(2): 262–80.
29. Erkal, Nisvan, Lata Gangadharan, and Nikos Nikiforakis (2011). "Relative Earnings and Giving in a Real-Effort Experiment." *American Economic Review*, 101 (7), 3330-48.
30. Figlio, David and Susanna Loeb (2011). *Chapter 8 - School Accountability* in Handbook of the Economics of Education, Vol. 3, Pages 383-421.
31. Goerg, Sebastian, Sebastian Kube and Jonas Radbruch (2019). "The Effectiveness of Incentive Schemes in the Presence of Implicit Effort Costs" *Management Science*, 65(9), 4063–4078
32. Glucksburg, Sam (1962). "Influence of Strength of Drive on Functional Fixedness and Perceptual Recognition," Journal of Experimental Psychology, 63(1), 36-41.
33. Goldin, Claudia, Lawrence F. Katz and Ilyana Kuziemko (2006). "The Homecoming of American College Women: The Reversal of the College Gender Gap." *The Journal of Economic Perspectives,* 20(4): 133-156.
34. Green, Francis and Steven McIntosh (2001). "The intensification of work in Europe." *Labour Economics*, 8: 291–308.
35. Gross, Daniel (2018). "Creativity under fire: The effects of competition on creative production." NBER Working Paper number 25057.

36. Grosse, Niels D. and Riener, Gerhard (2010). "Explaining Gender Differences in Competitiveness: Gender-Task Stereotypes" *Jena Economic Research Papers* 2010-017, Friedrich-Schiller-University Jena.
37. Hocevar, Dennis (1981). "Measurement of creativity: review and critique." *Journal of Personality Assessment*, 45(5), 450-64.
38. Harrison, Glenn W. (2002). "Estimating Individual Discount Rates in Denmark: A Field Experiment," *American Economic Review*, 92 (5), 1606–1617.
39. Heyman, Fredrik (2005). "Pay Inequality and Firm Performance: Evidence from Matched Employer-Employee Data." *Applied Economics,* 37(11): 1313–27.
40. Holt, Charles A. and Susan K. Laury (2002). "Risk Aversion and Incentive Effects," *The American Economic Review*, 92 (5), pp. 1644–1655.
41. Iriberri, Nagore and Pedro Rey-Biel (2017). "Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision." *Journal of Economic Behavior & Organization*, 135, 99-111.
42. Jirjahn, U. and Stephan, G. (2004). "Gender, piece rates and wages: evidence from matched employer-employee data", *Cambridge Journal of Economics*, 28, 683–704.
43. Kale, Jayant R., Ebru Reis, and Anand Venkateswaran (2009). "Rank-Order Tournaments and Incentive Alignment: The Effect on Firm Performance." *Journal of Finance* 64(3): 1479–1512.
44. Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan (2010). "Self-Control and the Development of Work Arrangements" *American Economic Review: Papers & Proceedings*, 100: 624–628.
45. Kangasniemi, Mari and Antti Kauhanen (2013). "Performance-related pay and gender wage differences." *Applied Economics*, 45(36), 5131-5143.
46. Knoblich, Gunther, Stellan Ohlsson, Hilde Haider, and Detlef Rhenius (1999). "Constraint relaxation and chunk decomposition in insight problem solving." *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 1999, 25, 1534-1555.
47. Lavy, Victor (2013). "Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-based Pay Tournaments among Teachers," *The Economic Journal*, 123(569), 540-573.

48. Lazear, Edward P. (2018). “Compensation and Incentives in the Workplace.” *Journal of Economic Perspectives*, 32(3): 195–214.

49. Lazear, Edward P. (2000). “Performance Pay and Productivity.” *American Economic Review* 90(5): 1346–61.

50. Lee, Sam Youl, Richard Florida and Zoltan J. Acs (2004). “Creativity and Entrepreneurship: A Regional Analysis of New Firm Formation” *Regional Studies*, 38(8), 879–891.

51. Mobbs, Shawn, and Charu G. Raheja (2011).“Internal Managerial Promotions: Insider Incentives and CEO Succession.” *Journal of Corporate Finance* 18(5), 1337–53.

52. Moretti, Enrico, and Jeffrey M. Perloff (2002). “Efficiency Wages, Deferred Payments, and Direct Incentives in Agriculture.” *American Journal of Agricultural Economics* 84(4), 1144–55.

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

54. Patel, Reshma and Lashawn Richburg-Hayes (2012). “Performance-Based Scholarships: Emerging Findings from a National Demonstration” Policy Report, (New York: MDRC, May).

55. Shalley, Christina E. & Greg R. Oldman (1997). “Competition and creative performance: Effect of competitor presence and visibility” *Creative Research Journal*, 10, 337-345.

56. Shearer, Bruce (2004). “Piece Rates, Fixed Wages, and Incentives: Evidence from a Field Experiment.” *Review of Economic Studies* 71(2): 513–34.

57. Shurchkov, Olga (2012). “Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints,” *Journal of the European Economic Association*, 10(5), 1189-1213.

58. Siddique, Abu and Michael Vlassopoulos (2017). “Competitive Preferences and Ethnicity: Experimental Evidence from Bangladesh.” *Economic Journal*, forthcoming.

59. Torrance, Paul E. (1974). *The Torrance Tests of Creative Thinking. Norms-Technical Manual Research Edition. Verbal Tests, Forms A and B-Figural Tests, Forms A and B.* Personnel Press, Princeton, NJ.

60. Webb, R. Alan, Michael G. Williamson, and Yue (May) Zhang (2013). "Productivity-Target Difficulty, Target-Based Pay, and Outside-the-Box Thinking," *The Accounting Review*, 88 (4), 1433-1457.

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

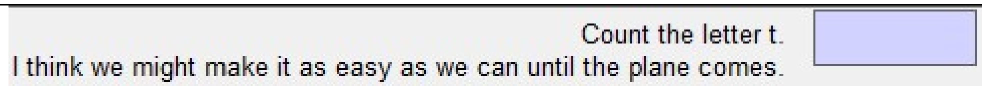
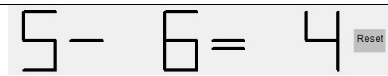
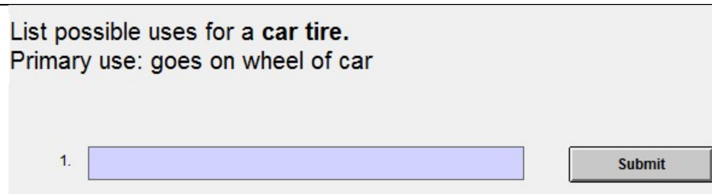
Routine	
Problem solving	
Pure creative	

Figure 2: visual representation of results from Table 4, with demographics controls and major dummies

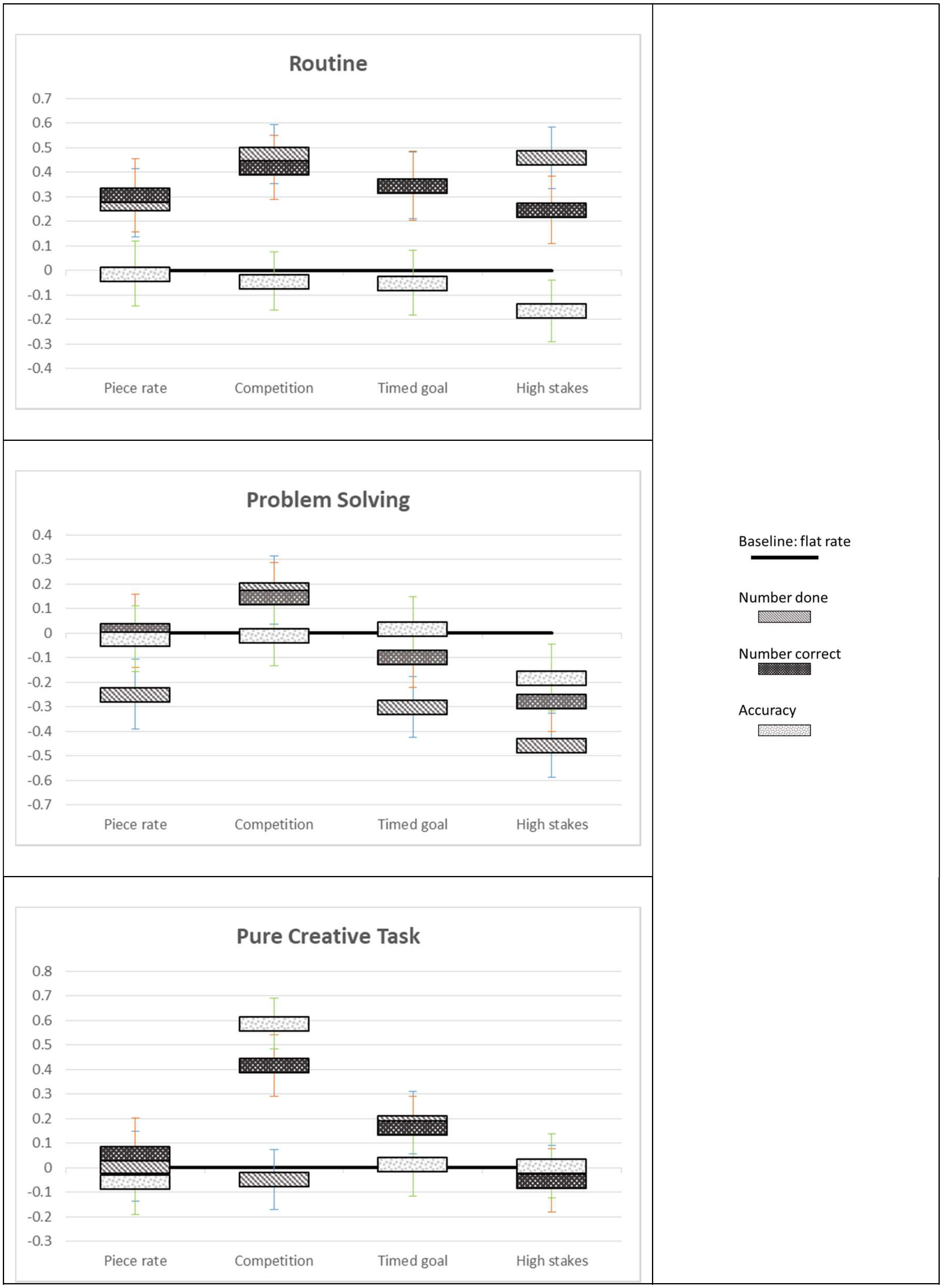


Table 1: Summary of payments under different types of incentives (or pressures)

Type of Incentive Pressure	*Payment Amount*
Flat rate	$10
Piece rate	$0.70, routine task $1.75, problem-solving $1.50, creative
Competition *(groups of 4 with no interaction or information on each other)*	$40 to winner $0 to others
Timed goal	$0 if medium goal not achieved $10 if medium goal is achieved $20 if high goal is achieved
High-pay incentive *(timed goal with higher payouts)*	$0 if medium goal not achieved $100 if medium goal is achieved $200 if high goal is achieved

Note: The goals for routine task are 14 correct for the medium and 18 correct for the high goal. The goals for problem-solving task are 2 correct for the medium and 4 correct for the high goal. The goals for routine task are 7 correct for the medium and 11 correct for the high goal.

Table 2: Mean and standard deviations of number completed and correct by incentive and task type

		Number done			Number correct		
Incentives	Observations	Routine	Problem solving	Pure Creative	Routine	Problem solving	Pure Creative
Piece Rate	84	18.93	3.89	13.13	12.6	1.69	10.37
		(5.01)	(1.73)	(3.96)	(4.24)	(1.64)	(3.88)
Time Pressure	97	19.07	3.8	13.82	12.6	1.53	10.84
		(4.87)	(1.4)	(3.76)	(4.1)	(1.35)	(3.41)
Competition	108	19.56	4.59	12.81	12.94	1.91	11.41
		(4.38)	(1.8)	(3.62)	(4.05)	(1.71)	(3.57)
High Incentive	92	19.47	3.59	12.85	12.29	1.32	9.92
		(4.22)	(1.55)	(3.86)	(3.81)	(1.27)	(3.28)
Flat Rate	178	17.51	4.33	13.08	11.35	1.7	10.28
		(4.45)	(1.56)	(3.69)	(3.55)	(1.41)	(3.35)
Total	559	18.71	4.1	13.13	12.21	1.64	10.55
		(4.63)	(1.64)	(3.76)	(3.93)	(1.48)	(3.5)

Table 3: Sample means by randomized incentive type

Incentive	Piece rate	Time pressure	Competition	High incentive	Flat rate	Overall
Observations	84	97	108	92	178	559
Percent of sample	15%	17.40%	19.30%	16.50%	31.80%	100%
Female	57.14	51.55	66.67	61.96	59.55	59.57
	(49.78)	(50.24)	(47.36)	(48.82)	(49.22)	(49.12)
Many math courses	42.86	39.18	36.11	38.04	39.33	39
	(49.78)	(49.07)	(48.26)	(48.82)	(48.99)	(48.82)
Non-native English speaker	23.81	24.74	26.85	29.35	23.6	25.4
	(42.85)	(43.38)	(44.53)	(45.79)	(42.58)	(43.57)
Second language	80.95	77.32	75	76.09	78.65	77.64
	(39.5)	(42.09)	(43.5)	(42.89)	(41.09)	(41.7)
Upperclassman	61.9	70.1	55.56	67.39	61.24	62.79
	(48.85)	(46.02)	(49.92)	(47.13)	(48.86)	(48.38)
Major: Social Science	29.76	31.96	30.56	31.52	32.58	31.48
	(46)	(46.87)	(46.28)	(46.71)	(47)	(46.49)
Major: Math or Engineering	29.76	39.18	25	29.35	33.15	31.48
	(46)	(49.07)	(43.5)	(45.79)	(47.21)	(46.49)
Major: Science	14.29	7.22**	17.59	13.04	18.54	14.85
	(35.2)	(26.01)	(38.25)	(33.86)	(38.97)	(35.59)
Major: Hummanities and Art	5.95	3.09	4.63	10.87**	3.37	5.19
	(23.8)	(17.4)	(21.11)	(31.3)	(18.1)	(22.2)
Major: Other	20.24*	18.56	22.22**	15.22	11.8	16.82
	(40.42)	(39.08)	(41.77)	(36.12)	(32.35)	(37.43)
Double major	30.95	30.93	34.26	40.22	34.27	34.17
	(46.51)	(46.46)	(47.68)	(49.3)	(47.59)	(47.47)
Few savings at the end of the month	29.76	21.65	17.59	23.91	21.35	22.36
	(46)	(41.4)	(38.25)	(42.89)	(41.09)	(41.7)
Low monthly spending categories	72.62	78.35	76.85	79.35	70.22	74.78
	(44.86)	(41.4)	(42.37)	(40.7)	(45.86)	(43.47)
Number of participants in session	9.79	9.13***	10.07	9.07***	10.66	9.89
	(4.21)	(4.18)	(3.86)	(3.36)	(3.97)	(3.97)
Risk elicitation	45.43*	40.34	40.2	41.73	31.9	38.62
	(65.22)	(46.58)	(42.62)	(53.15)	(55.23)	(52.98)

Table 4: Estimated effects of incentive type on number comleted, correct and accuracy by task type, demographic and major controls

	Routine task			Problem solving task			Pure creative task		
	1	2	3	4	5	6	7	8	9
VARIABLES	done	correct	accuracy	done	correct	accuracy	done	correct	accuracy
Piece rate	0.275**	0.306**	-0.0135	-0.249*	0.00901	-0.0235	0.00549	0.0572	-0.0587
	(0.138)	(0.150)	(0.132)	(0.143)	(0.148)	(0.135)	(0.143)	(0.145)	(0.134)
Timed goal	0.345**	0.345**	-0.0511	-0.301**	-0.0977	0.0165	0.184	0.162	0.0126
	(0.136)	(0.140)	(0.132)	(0.124)	(0.125)	(0.133)	(0.127)	(0.128)	(0.129)
Competition	0.473***	0.418***	-0.0438	0.176	0.147	-0.00818	-0.0488	0.416***	0.587***
	(0.121)	(0.131)	(0.118)	(0.139)	(0.139)	(0.125)	(0.122)	(0.126)	(0.104)
High pay	0.459***	0.246*	-0.165	-0.457***	-0.278**	-0.181	-0.0450	-0.0534	0.00720
	(0.125)	(0.137)	(0.126)	(0.130)	(0.122)	(0.136)	(0.135)	(0.129)	(0.131)
demographic controls	X	X	X	X	X	X	X	X	X
major dummies	X	X	X	X	X	X	X	X	X
session controls/dummies									
Observations	559	559	559	559	559	559	559	559	559
R-squared	0.070	0.064	0.046	0.056	0.047	0.037	0.043	0.092	0.106

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: repeated excluding flailers , demographic and major controls included

	Routine task			Problem solving task			Pure creative task		
	1	2	3	4	5	6	7	8	9
VARIABLES	done	correct	accuracy	done	correct	accuracy	done	correct	accuracy
Piece rate	0.316**	0.289**	-0.00143	-0.213	-0.0355	-0.0615	0.0468	-0.00673	-0.133
	(0.123)	(0.142)	(0.125)	(0.149)	(0.147)	(0.140)	(0.140)	(0.140)	(0.129)
Timed goal	0.423***	0.426***	0.0517	-0.283**	-0.124	0.0116	0.0638	0.0424	-0.0192
	(0.120)	(0.138)	(0.121)	(0.144)	(0.142)	(0.135)	(0.135)	(0.136)	(0.125)
Competition	0.479***	0.387***	-0.0606	0.215	0.139	-0.0425	-0.00358	0.362***	0.457***
	(0.116)	(0.134)	(0.118)	(0.139)	(0.137)	(0.130)	(0.131)	(0.131)	(0.121)
High pay	0.429***	0.298**	-0.0732	-0.344**	-0.301**	-0.189	-0.0538	-0.0783	-0.0434
	(0.123)	(0.142)	(0.124)	(0.146)	(0.144)	(0.137)	(0.139)	(0.140)	(0.128)
demographic controls	X	X	X	X	X	X	X	X	X
major dummies	X	X	X	X	X	X	X	X	X
session controls/dummies									
Observations	487	487	487	487	487	487	487	487	487
R-squared	0.099	0.073	0.054	0.051	0.045	0.042	0.038	0.082	0.090

Standard errors in parentheses

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

Appendix: Creativity under Pressure

Appendix A: Tables and Figures

Figure A1: distributions of the number of problems answered correctly and number done, by task types

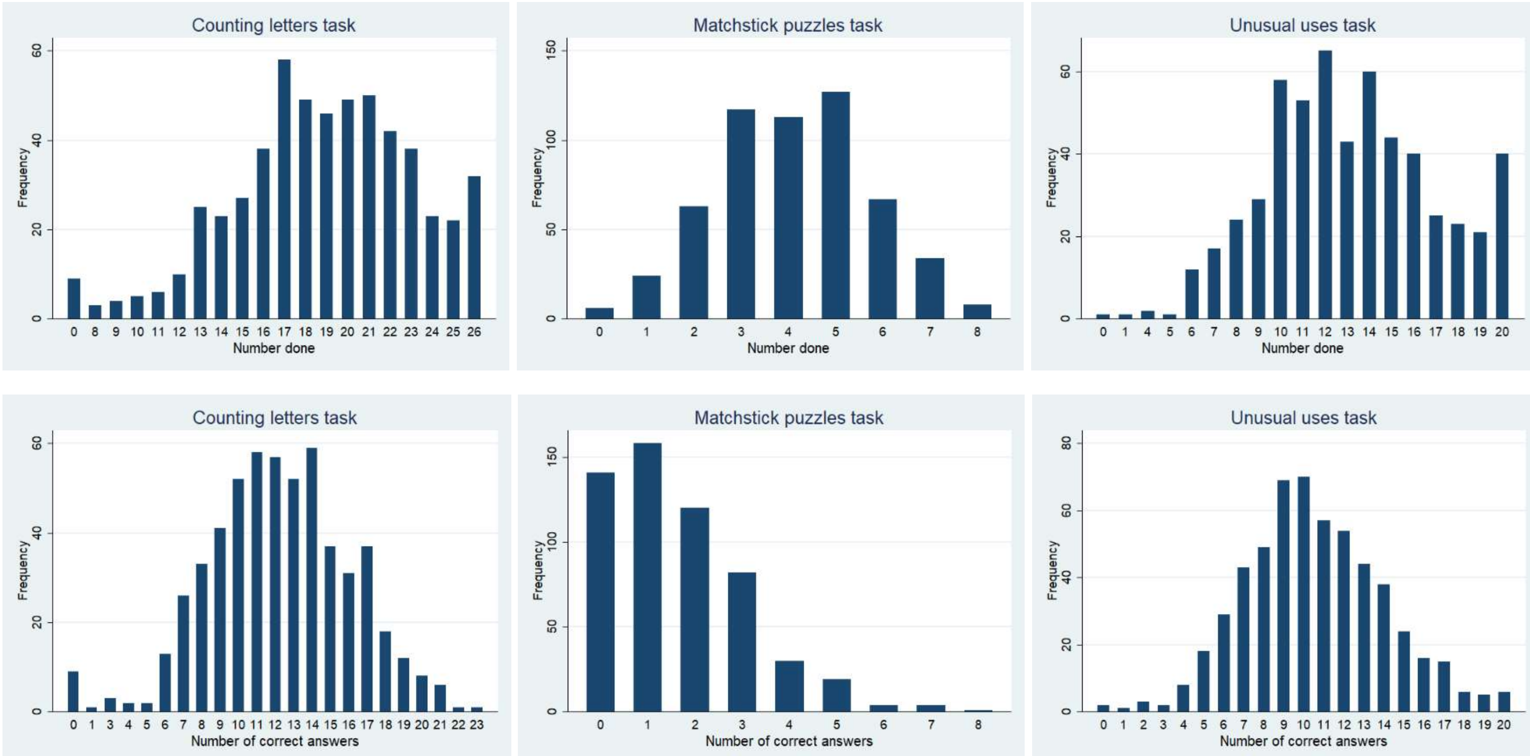


Figure A2: visual representation of results from Table A3, without controls

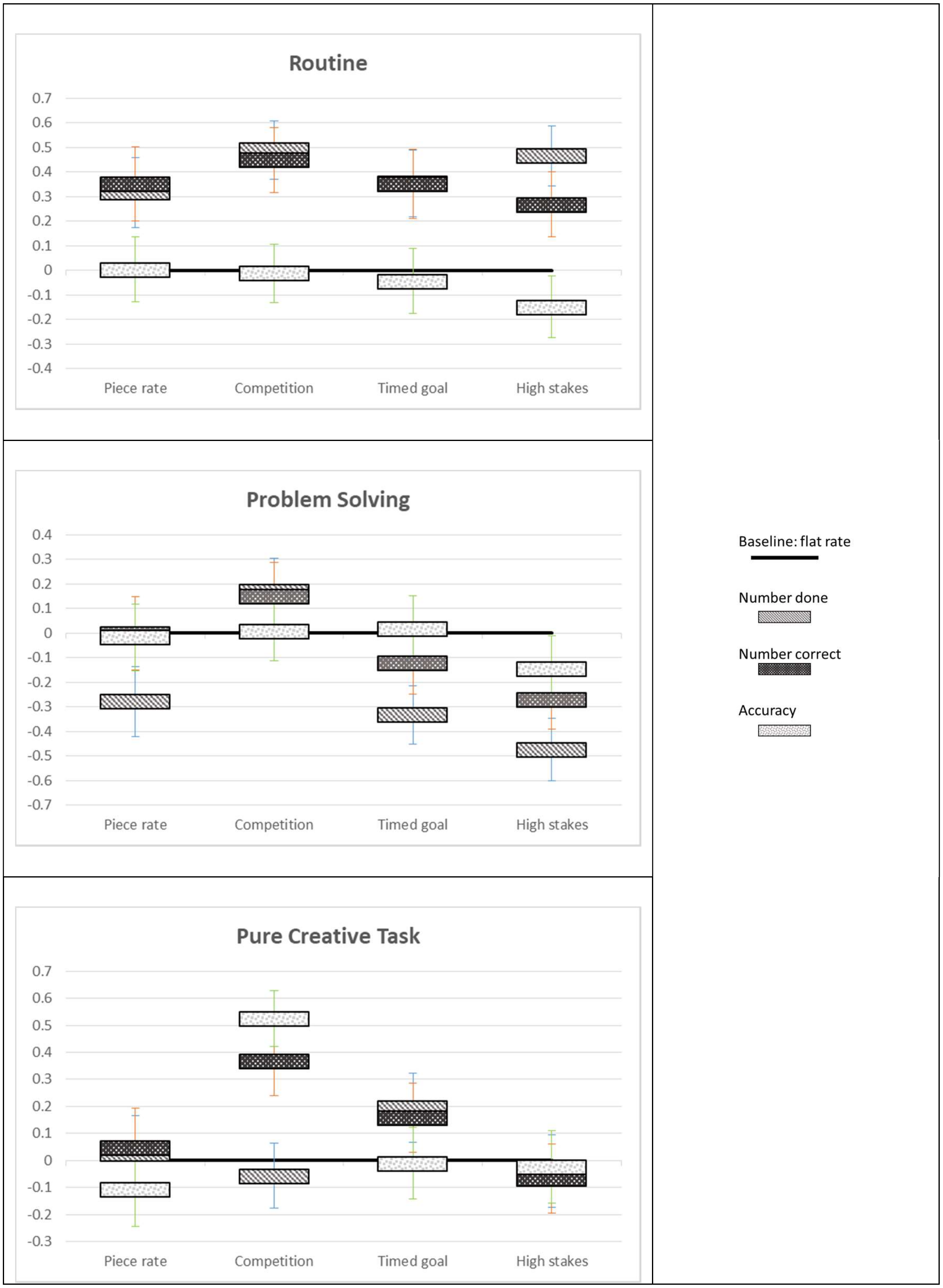


Figure A3: visual representation of results from Table A4, with minimal student demographic controls

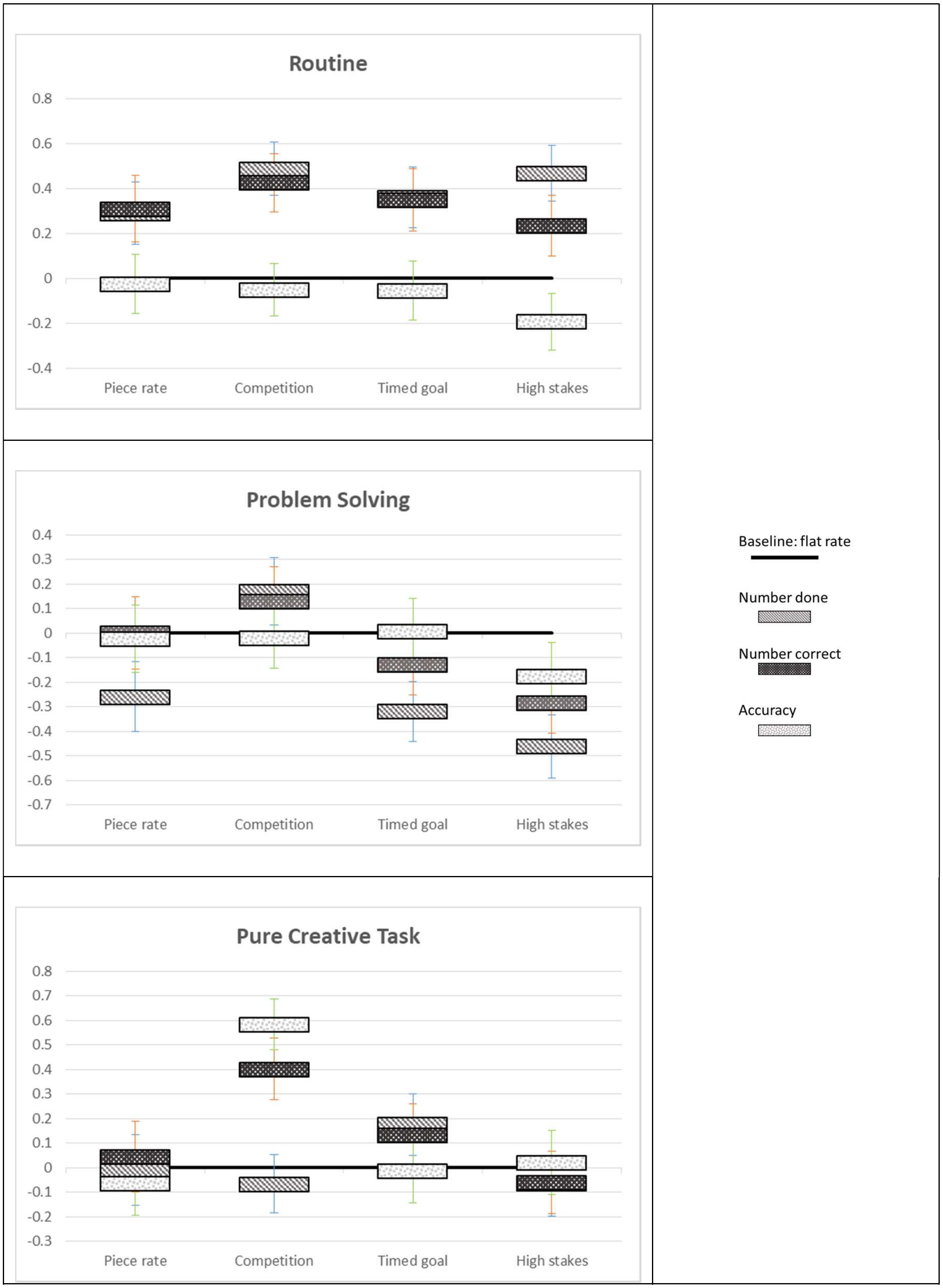


Figure A4: visual representation of results from Table A5, with full controls: demographics, major dummies, session controls and session dummies

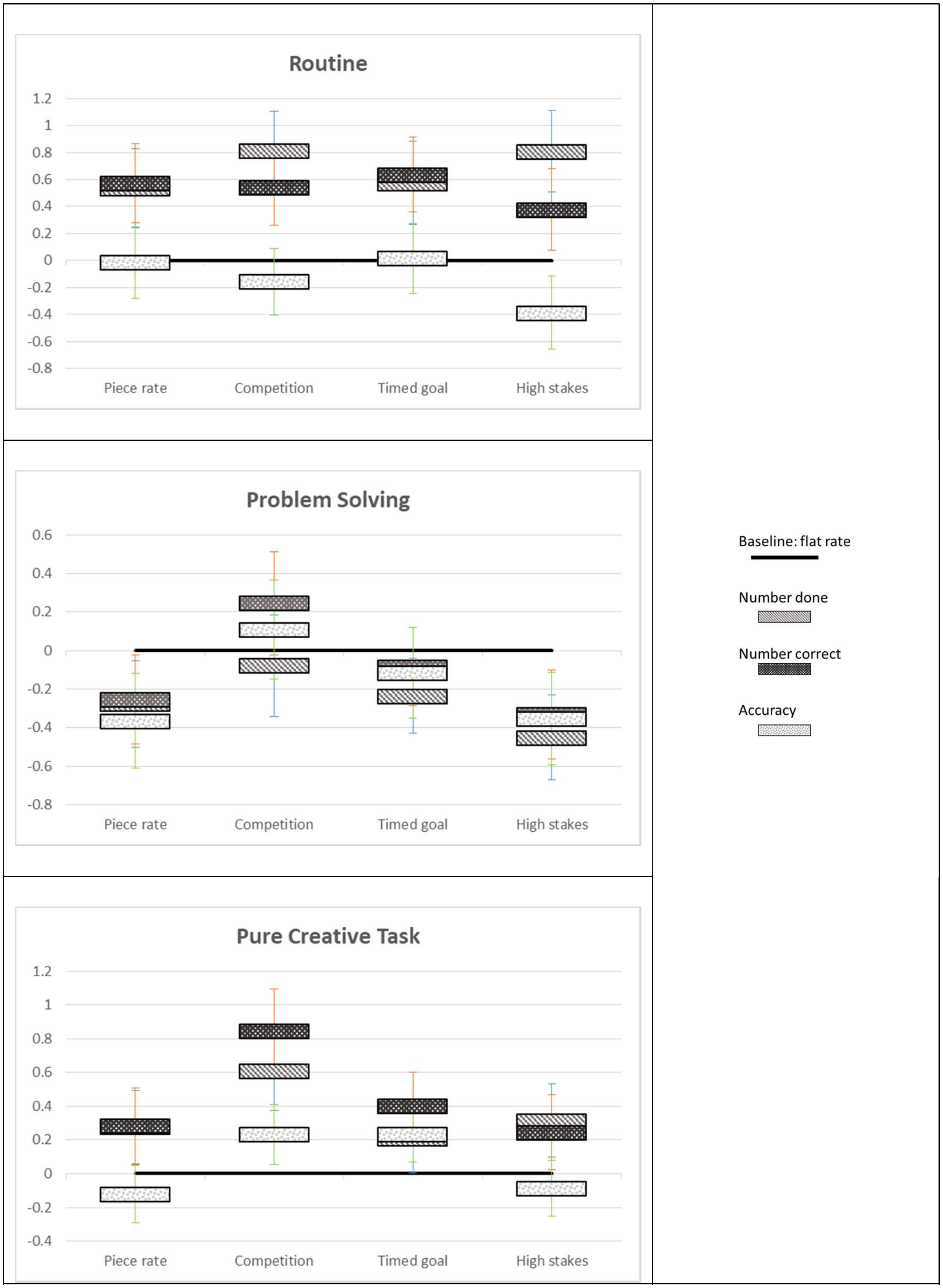


Figure A5: visual representation of results from Table 5, excluding “flailing” students

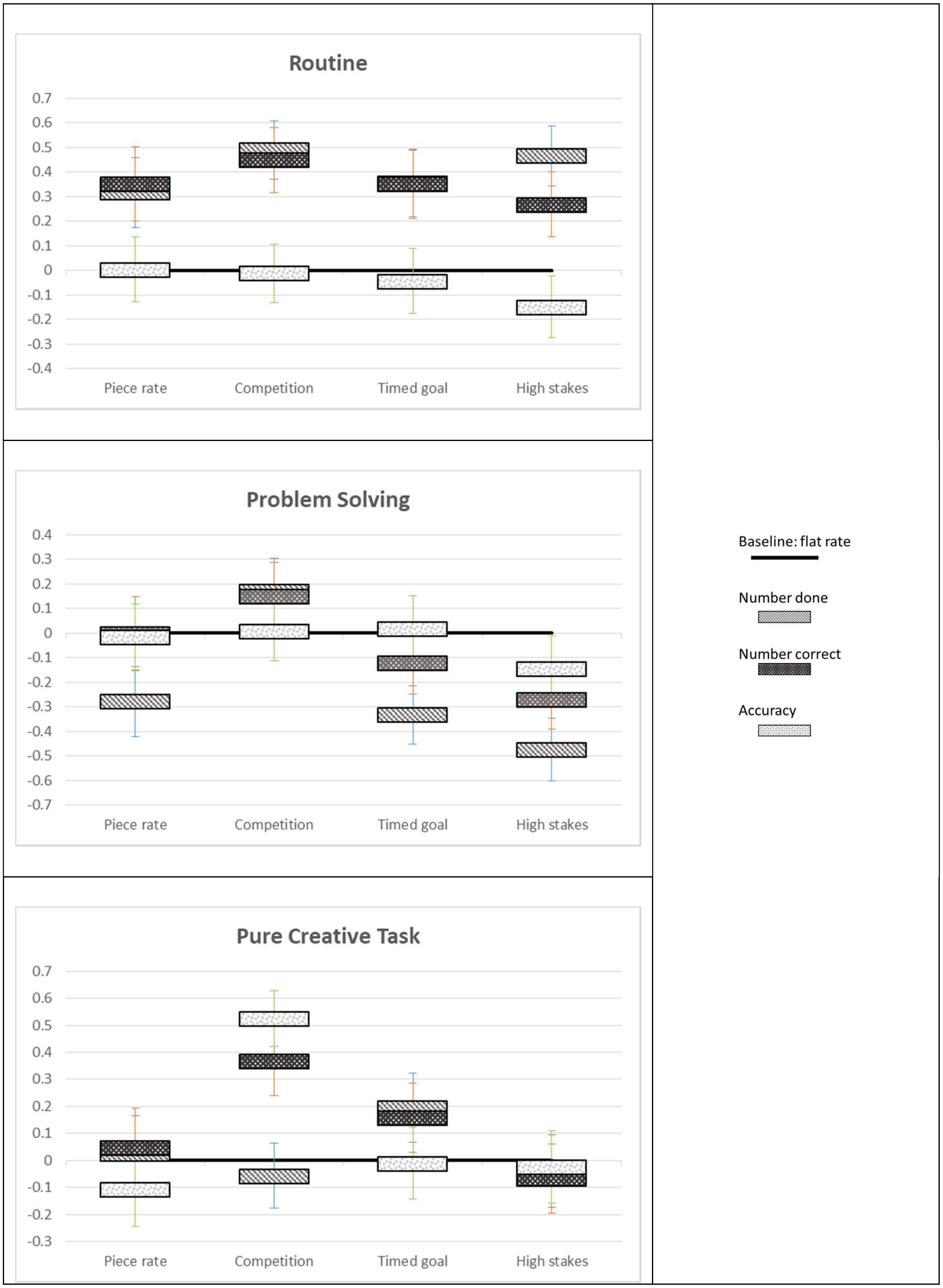


Table A1: Percent of participants scoring below the medium goal, above the medium goal and above the high goal by task and pressure

		Routine			Problem Solving			Pure Creative		
Incentives	Observations	below medium goal	above medium goal	above high goal	below medium goal	above medium goal	above high goal	below medium goal	above medium goal	above high goal
Piece Rate	84	55%	35%	11%	54%	35%	12%	15%	39%	45%
Time Pressure	97	62%	29%	9%	58%	35%	7%	9%	40%	51%
Competition	108	53%	33%	14%	48%	35%	17%	6%	35%	58%
High Incentive	92	65%	26%	9%	62%	33%	5%	16%	42%	41%
Flat Rate	178	71%	26%	3%	50%	40%	10%	11%	46%	43%
Total	559	62%	29%	8%	53%	36%	10%	11%	41%	47%

Table A2: Mean and SD of number done and correct by task type and task order

	Task order 1	Task order 2	Task order 3	Totals
Panel A: number done				
Routine				
	Task order 1	Task order 2	Task order 3	Totals
mean	19.1	18.43	18.57	18.71
SD	(4.54)	(4.69)	(4.65)	(4.63)
observations	194	167	198	559
Problem Solving				
	Task order 1	Task order 2	Task order 3	Totals
mean	4.1	4.11	4.09	4.1
SD	(1.72)	(1.56)	(1.67)	(1.64)
observations	169	208	182	559
Pure Creativity				
	Task order 1	Task order 2	Task order 3	Totals
mean	12.8	13.21	13.4	13.13
SD	(3.61)	(3.97)	(3.69)	(3.76)
observations	196	184	179	559
Panel B: number correct				
Routine				
	Task order 1	Task order 2	Task order 3	Totals
mean	12.35	11.78	12.44	12.21
SD	(4.03)	(4)	(3.77)	(3.93)
observations	194	167	198	559
Problem Solving				
	Task order 1	Task order 2	Task order 3	Totals
mean	1.64	1.68	1.6	1.64
SD	(1.53)	(1.4)	(1.53)	(1.48)
observations	169	208	182	559
Pure Creativity				
	Task order 1	Task order 2	Task order 3	Totals
mean	10.11	10.82	10.75	10.55
SD	(3.35)	(3.69)	(3.44)	(3.5)
observations	196	184	179	559

Table A3: Estimated effects of incentive type on number comleted, correct and accuracy by task type, no controls

	Routine task			Problem solving task			Pure creative task		
	1	2	3	4	5	6	7	8	9
VARIABLES	done	correct	accuracy	done	correct	accuracy	done	correct	accuracy
Piece rate	0.316**	0.351**	0.00339	-0.279*	-0.00132	-0.0182	0.0256	0.0481	-0.107
	(0.143)	(0.150)	(0.133)	(0.143)	(0.148)	(0.137)	(0.140)	(0.146)	(0.138)
Timed goal	0.354***	0.352**	-0.0446	-0.333***	-0.123	0.0165	0.195	0.158	-0.0101
	(0.135)	(0.140)	(0.132)	(0.119)	(0.125)	(0.134)	(0.128)	(0.128)	(0.132)
Competition	0.489***	0.449***	-0.0129	0.171	0.149	0.00860	-0.0568	0.367***	0.524***
	(0.120)	(0.132)	(0.118)	(0.134)	(0.139)	(0.123)	(0.121)	(0.128)	(0.103)
High pay	0.465***	0.268**	-0.149	-0.474***	-0.270**	-0.145	-0.0396	-0.0666	-0.0235
	(0.122)	(0.133)	(0.124)	(0.128)	(0.120)	(0.134)	(0.134)	(0.129)	(0.135)
demographic controls									
major dummies									
session controls/dummies									
Observations	559	559	559	559	559	559	559	559	559
R-squared	0.039	0.025	0.009	0.048	0.016	0.004	0.011	0.027	0.048

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A4: Estimated effects of incentive type on number comleted, correct and accuracy by task type, demographic controls

	Routine task			Problem solving task			Pure creative task		
	1	2	3	4	5	6	7	8	9
VARIABLES	done	correct	accuracy	done	correct	accuracy	done	correct	accuracy
Piece rate	0.291**	0.310**	-0.0231	-0.260*	0.000901	-0.0221	-0.00963	0.0442	-0.0627
	(0.138)	(0.148)	(0.131)	(0.143)	(0.147)	(0.137)	(0.145)	(0.144)	(0.133)
Timed goal	0.361***	0.348**	-0.0550	-0.319***	-0.128	0.00770	0.175	0.134	-0.0148
	(0.135)	(0.139)	(0.131)	(0.122)	(0.125)	(0.132)	(0.126)	(0.127)	(0.129)
Competition	0.488***	0.426***	-0.0509	0.170	0.130	-0.0204	-0.0660	0.402***	0.584***
	(0.120)	(0.130)	(0.117)	(0.136)	(0.139)	(0.124)	(0.120)	(0.124)	(0.103)
High pay	0.466***	0.235*	-0.192	-0.461***	-0.284**	-0.175	-0.0652	-0.0592	0.0205
	(0.124)	(0.136)	(0.125)	(0.129)	(0.123)	(0.136)	(0.133)	(0.127)	(0.130)
demographic controls	X	X	X	X	X	X	X	X	X
major dummies									
session controls/dummies									
Observations	559	559	559	559	559	559	559	559	559
R-squared	0.066	0.062	0.040	0.052	0.034	0.027	0.035	0.082	0.097

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A5: Estimated effects of incentive type on number comleted, correct and accuracy by task type, full set of controls

	Routine task			Problem solving task			Pure creative task		
	1	2	3	4	5	6	7	8	9
VARIABLES	done	correct	accuracy	done	correct	accuracy	done	correct	accuracy
Piece rate	0.536*	0.571*	-0.0139	-0.276	-0.253	-0.364	0.277	0.281	-0.120
	(0.294)	(0.294)	(0.265)	(0.224)	(0.230)	(0.245)	(0.218)	(0.226)	(0.171)
Timed goal	0.575*	0.636**	0.0146	-0.235	-0.0839	-0.117	0.209	0.401**	0.234
	(0.307)	(0.277)	(0.261)	(0.195)	(0.204)	(0.235)	(0.205)	(0.203)	(0.166)
Competition	0.813***	0.543*	-0.158	-0.0778	0.245	0.108	0.608***	0.846***	0.232
	(0.293)	(0.281)	(0.247)	(0.264)	(0.269)	(0.256)	(0.234)	(0.248)	(0.176)
High pay	0.807***	0.377	-0.388	-0.452**	-0.331	-0.354	0.314	0.245	-0.0860
	(0.302)	(0.303)	(0.270)	(0.220)	(0.230)	(0.239)	(0.217)	(0.222)	(0.165)
demographic controls	X	X	X	X	X	X	X	X	X
major dummies	X	X	X	X	X	X	X	X	X
session controls/dummies	X	X	X	X	X	X	X	X	X
Observations	559	559	559	559	559	559	559	559	559
R-squared	0.143	0.203	0.172	0.164	0.171	0.134	0.186	0.227	0.389

Robust standard errors in parentheses

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

Table A6: Mean and standard deviations of number completed and correct by incentive and task typ

Incentives	Observations	Routine	Percent Flailers Problem solving	Pure Creative
Piece Rate	84	0.02	0.1	0.01
		(0.15)	(0.3)	(0.11)
Time Pressure	97	0.03	0.1	0.01
		(0.17)	(0.31)	(0.1)
Competition	108	0.01	0.12	0
		(0.1)	(0.33)	(0)
High Incentive	92	0.01	0.11	0.01
		(0.1)	(0.31)	(0.1)
Flat Rate	178	0.01	0.11	0.01
		(0.11)	(0.32)	(0.11)
Total	559	0.02	0.11	0.01
		(0.13)	(0.31)	(0.09)

Table A7: repeated excluding flailers , no controls included

	Routine task			Problem solving task			Pure creative task		
	1	2	3	4	5	6	7	8	9
VARIABLES	done	correct	accuracy	done	correct	accuracy	done	correct	accuracy
Piece rate	0.334***	0.320**	0.0152	-0.224	-0.0249	-0.0361	0.0737	0.0150	-0.141
	(0.123)	(0.142)	(0.125)	(0.146)	(0.145)	(0.139)	(0.138)	(0.141)	(0.129)
Timed goal	0.419***	0.417***	0.0481	-0.311**	-0.128	0.0381	0.0675	0.0234	-0.0547
	(0.119)	(0.137)	(0.121)	(0.141)	(0.140)	(0.134)	(0.134)	(0.136)	(0.125)
Competition	0.506***	0.436***	-0.0249	0.225*	0.166	0.0103	-0.00137	0.336**	0.409***
	(0.115)	(0.133)	(0.117)	(0.135)	(0.134)	(0.129)	(0.129)	(0.131)	(0.120)
High pay	0.440***	0.301**	-0.0805	-0.368***	-0.299**	-0.151	-0.0664	-0.0974	-0.0593
	(0.121)	(0.140)	(0.123)	(0.142)	(0.141)	(0.135)	(0.137)	(0.139)	(0.127)
demographic controls									
major dummies									
session controls/dummies									
Observations	487	487	487	487	487	487	487	487	487
R-squared	0.058	0.031	0.008	0.042	0.020	0.006	0.006	0.026	0.041

Standard errors in parentheses

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

Table A8: repeated excluding flailers , demographic controls included

	Routine task			Problem solving task			Pure creative task		
	1	2	3	4	5	6	7	8	9
VARIABLES	done	correct	accuracy	done	correct	accuracy	done	correct	accuracy
Piece rate	0.330***	0.291**	-0.0125	-0.222	-0.0296	-0.0428	0.0341	-0.00788	-0.122
	(0.123)	(0.141)	(0.124)	(0.148)	(0.146)	(0.139)	(0.139)	(0.140)	(0.129)
Timed goal	0.434***	0.422***	0.0394	-0.307**	-0.142	0.0204	0.0612	0.0194	-0.0466
	(0.119)	(0.137)	(0.120)	(0.143)	(0.141)	(0.134)	(0.134)	(0.135)	(0.124)
Competition	0.502***	0.394***	-0.0723	0.208	0.133	-0.0371	-0.0233	0.348***	0.460***
	(0.116)	(0.133)	(0.117)	(0.138)	(0.136)	(0.129)	(0.130)	(0.131)	(0.121)
High pay	0.455***	0.286**	-0.110	-0.362**	-0.306**	-0.172	-0.0802	-0.0914	-0.0289
	(0.121)	(0.139)	(0.123)	(0.144)	(0.142)	(0.135)	(0.137)	(0.138)	(0.127)
demographic controls	X	X	X	X	X	X	X	X	X
major dummies									
session controls/dummies									
Observations	487	487	487	487	487	487	487	487	487
R-squared	0.090	0.069	0.047	0.046	0.038	0.034	0.030	0.069	0.073

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A9: repeated excluding flailers , demographic and major controls included

	Routine task			Problem solving task			Pure creative task		
	1	2	3	4	5	6	7	8	9
VARIABLES	done	correct	accuracy	done	correct	accuracy	done	correct	accuracy
Piece rate	0.622***	0.733***	0.201	-0.355	-0.336	-0.415	0.219	0.312	0.0332
	(0.232)	(0.256)	(0.227)	(0.273)	(0.272)	(0.265)	(0.248)	(0.251)	(0.204)
Timed goal	0.696***	0.905***	0.329	-0.269	-0.117	-0.127	0.0511	0.359	0.385**
	(0.215)	(0.238)	(0.211)	(0.252)	(0.251)	(0.244)	(0.231)	(0.234)	(0.190)
Competition	0.717***	0.603**	-0.00760	-0.0622	0.193	0.0526	0.617**	0.811***	0.206
	(0.240)	(0.265)	(0.235)	(0.282)	(0.280)	(0.273)	(0.258)	(0.262)	(0.212)
High pay	0.823***	0.610**	-0.0933	-0.353	-0.343	-0.348	0.286	0.290	-0.0305
	(0.223)	(0.246)	(0.218)	(0.262)	(0.261)	(0.254)	(0.242)	(0.245)	(0.198)
demographic controls	X	X	X	X	X	X	X	X	X
major dummies	X	X	X	X	X	X	X	X	X
session controls/dummies	X	X	X	X	X	X	X	X	X
Observations	487	487	487	487	487	487	487	487	487
R-squared	0.199	0.247	0.218	0.194	0.175	0.131	0.221	0.245	0.416

Standard errors in parentheses

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

Appendix: Creativity under Pressure

Appendix B: Sample screen shots

Instructions

In this study, you will receive a series of different tasks. These tasks consist of solving problems within a time limit. Your earnings will depend on how many problems you solve in a task. They may also depend on how many problems other participants solve in the same task.

A shorter second section of the study will involve choices between gambles and a survey. Your earnings in this section may depend on chance as well as the choices you make.

If at any time during this session you wish to leave, please talk to the monitor to receive your show-up payment of $7.

If you leave during the session for any reason, you will receive the show-up payment. You will not receive other earnings unless you complete the session. You will not be permitted to return to the session once you have left.

Please turn off your cell phone and put away all electronic devices or other materials you may have brought into the lab.

We expect your full attention during the session. Talking, using a cell phone, or other distractions will result in the monitor asking you to leave.

If you wish to visit the restroom or take care of other business before the session begins, please do so now.

Instructions

Your earnings in today's session come in three parts:
1. a show-up payment
2. a payment for a problem-solving task
3. a payment you earn in a final set of questions for the session

You have already earned $7 for showing up (part 1).

You will receive your earnings for 1 of the 3 problem solving tasks you complete (part 2). At the end of the session, after you have completed all 3 tasks, one task will be randomly selected by a computer. You will receive the money you earned in that task.

After you have finished the problem solving tasks, you will be given a set of choices. Depending on your choices, you may earn additional money in this final part of the session (part 3).

Instructions

Before beginning your tasks, please answer the following question:

Suppose you do 3 different problem-solving tasks during today's session. In task 1, you earn $2. In task 2, you earn $6. In task 3, you earn $0. You also earn a $7 show-up payment. You earn nothing in the third and final part of the session.

Which of the following is a possible total payment you might receive?

- a) $2
- b) $7
- c) $6
- d) $10

Continue

If 2, 6, or 10:

Dialog

That answer is incorrect. Please read the instructions carefully and try again.

OK

Instructions

Correct, the answer is $7.

You have already earned $7 for showing up. This amount does not depend on the tasks you complete.
In this example, only one of the three tasks you completed would be randomly selected to receive payment.

In task 1, you earned $2. If this task was randomly chosen, you would earn a total of $9: $7 for showing up and $2 for the task.

In task 2, you earned $6. If this task was chosen, you would earn a total of $13: $7 for showing up and $6 for the task.

In task 3, you earned $0. If this task was chosen, you would earn a total of $7: $7 for showing up and $0 for the task.

Therefore your possible payments are $9, $13, or $7, depending which task was randomly chosen for payment.

You will only be paid for ***one*** *of the problem solving tasks.*
Remember, each task you encounter could be the only one that matters!

Instructions

The session will begin once everyone is ready.

Remember, you can leave at any time and receive the show-up payment, but you will not be able to re-join the session after leaving.

Please put away any distractions and press "continue" to begin the problem-solving tasks.

Task 1 instructions: counting letters

Instructions

Your next task is counting how many times a given letter appears in a line of text. You will be told what letter to count and given a blank box. Then you will see a line of text. You solve the problem by typing the number of times the letter appears in the text into the box.

Example:

Count the letter i: [ ] Submit

Crowell attended several years of Saturday classes at the Philadelphia Museum School.

The letter i appears twice, so type "2" into the box and press "submit"

When you have finished answering a question, always press the "submit" button to make sure your answer is counted.

For your next round, you will have 3.0 minutes to answer as many questions as you can.

This round allows you to practice these types of questions. It will not count towards your earnings.

Task 2 instructions: matchsticks

Instructions

Your next task will involve moving lines to correct equations.

In each of these problems, you must drag 1 and only 1 line to a new position. You may move any line except an equals sign (=) or a minus sign (-). You may change a minus to a plus by adding a line or a plus to a minus by removing one, but you may not signify a negative number using a separate minus sign.

You must move the line to a new position in the same equation in order for your answer to be correct. Making it disappear does not count. If the line disappears, click "reset" to bring it back.

Horizontal lines may be moved to vertical positions and vice versa. The software uses the position of your mouse pointer to determine where you are placing the line.

Please examine the numbers below to see how the lines should be placed for each number.

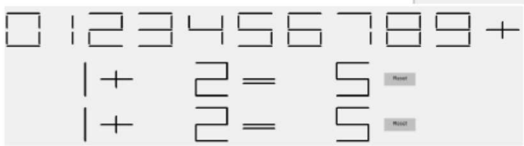
Example

Try moving the lines in the examples to the left. Click "Reset" to try again after moving a line.

When you are comfortable moving the lines, move the vertical line from the plus sign to the top of the number 1 in the first example to create the equation 7-2=5 .

For the second example, move the left vertical line in the 5 to the right to make the equation 1+2=3.

When you are done, press "Submit" to move to the next instructions. Note that both answers would count for points since both result in a correct equation.



When you have finished answering a question, always press the "submit" button to make sure your answer is counted.

For your next round, you will have 3.0 minutes to answer as many questions as you can.

This round allows you to practice these types of questions. It will not count towards your earnings.

Task 3 instructions: unusual uses

Instructions

Your next task is to list as many uses as possible for an everyday object. You will see an everyday object and a series of blank boxes. You solve the problem by writing as many uses as you can think of for that object, each in a separate box. You will receive a higher score the more different uses you suggest for the object.

You will also receive a higher score for uses that are unusual or original. Both quantity of responses and originality of responses factor into your score.

Uses that are the same as the primary use identified with the word and uses that are impossible will not be counted.

Example:

List uses for a coffee mug.
Primary use: holding drinks such as coffee.

1: hold coffee — Submit

2: — Submit

Instructions:

Type "hold coffee" into the first box, "weapon" into the second, and "decoration" into the third.

Press "Submit" to get a new box after typing each answer.

Instructions

Well done!

Your responses in this task will earn points if they are possible and different from the primary use.

In this example, the answers "weapon" and "decoration" would earn points because they are possible and different from the primary use.

"Hold coffee" would not earn points because it is the same as the primary use.

The response "wear as shoes" would not earn points because it is not possible.

If you responded "hit someone" in addition to "weapon," you would only receive points for 1 answer because they represent the same type of use. Similarly, "hit someone" and "throw at someone" would both count as the same use,

weapon, and only earn 1 set of points even if you did not include the word "weapon."
The uses must be clearly different to count separately.

For your next round, you will have 3.0 minutes to answer as many questions as you can.

This round allows you to practice these types of questions. It will not count towards your earnings.

Payment instructions

Flat rate

For your next round, you will have 3.0 minutes to find as many answers as you can.

You will earn $10 for participating in this round. Your earnings do not depend on how many answers you find.

If this task is selected as your paid task, you will receive these earnings in cash at the end of the session.

Piece rate

For your next round, you will have 3.0 minutes to answer as many questions as you can.

You will earn $0.70 for each problem you solve correctly.

If this task is selected as your paid task, you will receive these earnings in cash at the end of the session.

Timed goal

For your next round, you will have 3.0 minutes to answer as many questions as you can.

Your earnings in this task will depend on whether you meet the following goals:
- If you complete fewer than 14 problems in 3.0 minutes, you will earn **$0.**
- If you complete 14 or more problems in 3.0 minutes, you will earn **$10.00.**
- If you complete 18 or more problems in 3.0 minutes, your payment will increase to **$20.00.**

If this task is selected as your paid task, you will receive these earnings in cash at the end of the session.

Payment instructions continued

Competition

For your next round, you will have 3.0 minutes to find as many answers as you can.

You have been placed in a group with three other participants. The participant in your group who earns the most points will earn $40.00 for this task. If you do not win, you will earn $0.00 for this task. If there is a tie then the computer will randomly choose a winner from the top scorers in the group.

•If you win, you earn $40.00 for this task.
•If you do not win, you earn $0.00 for this task.

If this task is selected as your paid task, you will receive these earnings in cash at the end of the session.

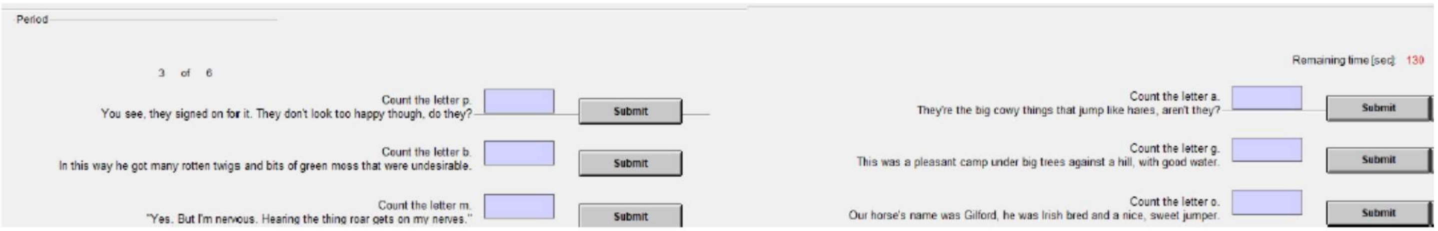
High stakes

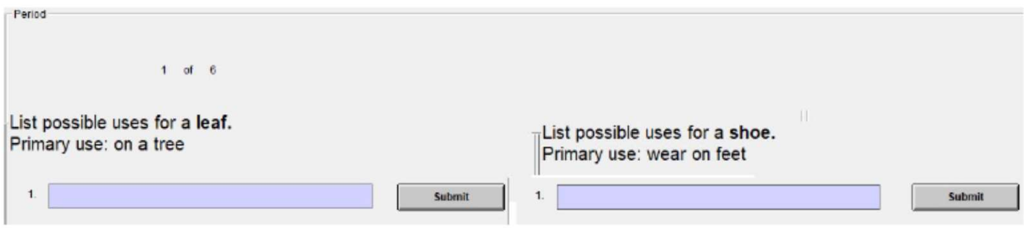
For your next round, you will have 3.0 minutes to answer as many questions as you can.

Your earnings in this task will depend on whether you meet the following goals:
- If you complete fewer than 14 problems in 3.0 minutes, you will earn **$0.**
- If you complete 14 or more problems in 3.0 minutes, you will earn **$100.00.**
- If you complete 18 or more problems in 3.0 minutes, your payment will increase to **$200.00.**

Please note the high possible payments for this round. If this task is selected as your paid task, you will receive these earnings in cash at the end of the session..

Task screens





NB: these screenshots show the top of the screen. The questions continue down the full page.

After each round (practice or paid)

Period

1 of 6

The round has ended.
Please press "continue" to move to your next task.

You answered 0 questions correctly in 3.0 minutes.

Final task: risk preferences

Instructions

Your next task is to make a series of 10 choices. Each decision is a between an "Option A" and "Option B." You will make all ten choices, but only one of them will be randomly selected to determine your earnings for this task. Each choice has an equal chance of being used.

In the first choice, Option A is a 1 in 10 chance to earn $4.00 and a 9 in 10 chance to earn $3.10. Option B is a 1 in 10 chance to earn $7.80 and a 9 in 10 chance to earn $0.20. Which option would you prefer?

When you have made your choices, the computer will select a choice and a random number between 1 and 10 to determine your earnings. If the chance of a high payment is 1/10, you have to roll a 10 to get the high payment. If the chance is 2/10, you have to roll a 9 or 10, and so on.

Suppose the choice above was selected and the computer rolled a 5. Since the chance of a high payment in this choice is 1/10, you would have to roll a 10 to get the high payment. In this case, you would earn $3.10 if you had chosen Option A and $0.20 if you had chosen Option B.

Press continue to make your choices.

	Choice A	Choice B	A or B?
1.	1/10 chance to earn $4.00, 9/10 chance to earn $3.10	1/10 chance to earn $7.80, 9/10 chance to earn $0.20	
2.	2/10 chance to earn $4.00, 8/10 chance to earn $3.10	2/10 chance to earn $7.80, 8/10 chance to earn $0.20	
3.	3/10 chance to earn $4.00, 7/10 chance to earn $3.10	3/10 chance to earn $7.80, 7/10 chance to earn $0.20	
4.	4/10 chance to earn $4.00, 6/10 chance to earn $3.10	4/10 chance to earn $7.80, 6/10 chance to earn $0.20	
5.	5/10 chance to earn $4.00, 5/10 chance to earn $3.10	5/10 chance to earn $7.80, 5/10 chance to earn $0.20	
6.	6/10 chance to earn $4.00, 4/10 chance to earn $3.10	6/10 chance to earn $7.80, 4/10 chance to earn $0.20	
7.	7/10 chance to earn $4.00, 3/10 chance to earn $3.10	7/10 chance to earn $7.80, 3/10 chance to earn $0.20	
8.	8/10 chance to earn $4.00, 2/10 chance to earn $3.10	8/10 chance to earn $7.80, 2/10 chance to earn $0.20	
9.	9/10 chance to earn $4.00, 1/10 chance to earn $3.10	9/10 chance to earn $7.80, 1/10 chance to earn $0.20	
10.	10/10 chance to earn $4.00, 0/10 chance to earn $3.10	10/10 chance to earn $7.80, 0/10 chance to earn $0.20	

Thank you! You have completed the second portion of this session.

Please enter your station number below, then press "continue" to answer some demographic questions while we prepare your payment.

We will use your answers to help us test our methods and understand our results. Your responses will be pooled with the responses of other participants so that each individual's answer is confidential.

Questions

Expected graduation year:

Major field 1:
- Economics or business
- Psychology
- Other social sciences
- Math, computer science, engineering
- Natural sciences
- Humanities or languages
- Arts
- Other

Major field 2:
- Economics or business
- Psychology
- Other social sciences
- Math, computer science, engineering
- Natural sciences
- Humanities or languages
- Arts
- Other

Gender:
- male
- female
- other/prefer not to state

Number of college-level math or statistics classes taken:

Age range:
- 18-19
- 20-21
- 22 or older

Are you a native English speaker?
- Yes
- No

Do you speak another language besides English?
- Yes
- No

Take a moment to think about your spending patterns.
On average, how much do you think you spend on discretionary items per month? This amount does not include spending on basics like eating at the cafeteria, printing, tuition and your cell phone bill. Instead, think of spending on things like coffee or meals out, social outings, travel, clothes or sports gear.
- $0-$49
- $50-$99
- $100-$299
- $300-$499
- $500-$999
- $1000 or more

Are you able to save anything at the end of the month?
- Never
- Rarely
- Sometimes
- Always

Which of the following best describes you?
- Gay/lesbian/homosexual
- Bisexual
- Heterosexual/Straight
- Prefer Not To State

Do you have any other comments?

Results

Thank you for completing the session!

Your show-up payment is $7.

The **task** randomly selected for your payment is **counting letters.**

The payment system for your randomly-selected task is money per point.

Your earnings for this task are $0.00 **unless your task type is unusual uses,** in which case your earnings are being calculated by the monitor.

For the second part of the session, row number 1 was selected and your roll was 5. Your earnings for the second part of the session are $3.10.

Your total earnings is $10.25.

Please copy your total payment and station number onto the receipt. **If your task was unusual uses,** your total may be incorrect. In this case fill out the station number and wait to find out your payment. When you have finished copying everything down, press "continue" to keep your payment amounts private.