unibz
Fakultät für Wirtschaftswissenschaften
Facoltà di Economia
Faculty of Economics and Management

BEMPS –
Bozen Economics & Management
Paper Series

NO 51/ 2018

Paying for what kind of performance? Performance pay and multitasking in mission-oriented jobs

Daniel Jones, Mirco Tonin, Michael Vlassopoulos

Paying for what kind of performance? Performance pay and multitasking in mission-oriented jobs*

Daniel Jones
University of Pittsburgh

Mirco Tonin
Free University of Bozen-Bolzano, IZA, CESifo, Dondena Centre

Michael Vlassopoulos
University of Southampton, IZA

July 2018

Abstract

How does pay-for-performance (P4P) impact productivity, multitasking, and the composition of workers in mission-oriented jobs? These are central issues in sectors like education or healthcare. We conduct a laboratory experiment, manipulating compensation and mission, to answer these questions. We find that P4P has positive effects on productivity on the incentivized dimension of effort and negative effects on the non-incentivized dimension for workers in non-mission-oriented treatments. In mission-oriented treatments, P4P generates minimal change on either dimension. Participants in the non-mission sector – but not in the mission-oriented treatments – sort on ability, with lower ability workers opting out of the P4P scheme.

Keywords: Prosocial motivation, Performance pay, Multitasking, Sorting
JEL Codes: C91, M52, J45

* We thank Sera Linardi and seminar and workshop participants at Birmingham, Bocconi, Erasmus, Innsbruck, and the Science of Philanthropy Initiative 2017 conference for helpful comments. Financial support from the Riegel and Emory Human Resources Center at the University of South Carolina is gratefully acknowledged. We thank Matthew Harvey and Xiaocheng Hu for excellent research assistance.

1. Introduction

Performance-related compensation has become an important personnel tool in a large segment of the labour market across many countries.[1] This type of compensation has been recognized to offer two benefits for firms that make use of it: incentive effects on productivity that result from aligning the interest of workers with those of the firm, and, sorting effects that arise as more productive workers are attracted by firms offering performance-related pay relative to fixed pay (Lazear, 1986). While bonuses, piece rates and other forms of pay-for-performance (P4P) are more widespread in the private sector, the public and broader nonprofit sector is often characterized by relatively rigid pay policies, in which seniority-based rules prevail over schemes rewarding performance. In recent years, there is a lively debate on whether or not this should change, with disagreement over the desirability of the pay of teachers, nurses, law enforcement officers and other frontline public service providers becoming more conditional on performance (Burgess and Ratto, 2003). Those who advocate for the adoption of pay-for-performance schemes extoll the potential efficiency gains,[2] while those who are more reluctant usually highlight two aspects of the public sector as being particularly problematic in regards to implementing performance-related compensation.

The first one is that performance is more difficult to measure in the public than in the private sector (Dixit, 2002). This is partly due to technological aspects, as it is easier to measure output in manufacturing (e.g. a car plant) than in services (e.g. a school), and the public-sector activity is concentrated in the (advanced) service sector. There is also a governance aspect to this, in that private, for-profit firms have a clear prominent objective, the bottom-line, to which every activity can be potentially benchmarked against. Public sector organizations are instead very often characterized by a multiplicity of stakeholders with differing objectives and this makes the jobs of workers in the sector more multidimensional and their performance more difficult to define (e.g. how much did the teacher increase test scores, as opposed to employability, as opposed to civic virtues?).

A second aspect making P4P potentially problematic in the public sector is that prosocial motivation plays a particularly important role in some occupations, e.g. teachers or nurses, in which helping beneficiaries is an important characteristic of the job (Francois and Vlassopoulos, 2008; Besley and Ghatak, 2018). In these settings, P4P may have adverse effects for two reasons: first, performance may decline due to crowding out of prosocial motivation by higher-powered financial incentives (Gneezy et al., 2011) and, second, it might lead to the attraction of less prosocially motivated workers thus causing the dilution of prosocial motivation in the workforce (Jones, 2015; Finan et al., 2017).

[1] See, for instance, Lemieux et al. (2009) and Bloom and van Reenen (2011).
[2] See Neal (2011) and Imberman (2015) for surveys of the literature on the effects of performance pay in education.

The purpose of the current paper is to contribute to the debate about the desirability of pay-for-performance in sectors with a mission, where prosocial motivation among employees is particularly important. We do so by providing experimental evidence of incentive and sorting effects of performance-related compensation in a setting characterized by prosocial elements and multitasking. To this end, we design a real-effort experiment with a novel task -- a word formation task inspired by the board game Scrabble -- characterized by two dimensions that we refer to as quantity (the number of words) and quality (the complexity of each word, as measured by points). We first observe how performance in the two dimensions differs between a treatment in which compensation is a fixed wage and one in which it is a piece rate on quantity. We also implement a treatment that induces prosocial motivation by linking the number of points to a charitable donation and, again, observe performance under a fixed wage and a piece rate on quantity. To assess sorting effects, in a second stage we allow subjects to choose whether to keep the compensation structure that was imposed upon them or switch to a fixed wage instead.

We find the following main results. In the absence of a mission, i.e. induced prosocial motivation, our experiment reproduces the classical results of the literature on incentives, with the piece rate increasing performance in the incentivized dimension, while reducing effort along the unincentivized one. Also, high ability workers self-select into the piece rate compensation scheme, therefore inducing an additional positive effect of P4P on productivity due to sorting (Lazear, 2000). In the environment with a mission, we find a much smaller incentive effect due to P4P, while sorting is also very different, as it takes place along the prosocial motivation dimension rather than the ability one. In particular, with fixed pay people with high motivation are willing to give up financial gains to stay in the mission sector, while it is not the case that self-selection into the mission sector with P4P takes place along the skill dimension. Our results imply that P4P is less effective in increasing performance when there is a motivational component, in terms of both the incentive and selection effects. On the other hand, the negative consequences on the unincentivized dimension are also not present.

Thus, to relate back to the discussion about the merit of introducing pay-for-performance in mission-oriented workplaces, our study suggests that performance pay is less successful in increasing effort on the incentivized dimension when workers are motivated, but also does less harm to other dimensions of effort. Furthermore, performance pay does not seem to attract higher quality workers in terms of ability or motivation. If we consider that we do see some sorting of motivated workers in the motivated setting under the flat wage, we may conclude that performance pay foregoes the opportunity to screen on motivation, without the benefit of screening on ability.

This paper explores the interaction between financial incentives, prosocial motivation and performance in an environment with multi-tasking. As such, it intersects with different strands of literature.

First, we build on the literature on incentives that has long recognized the effect of incentive pay in multitasking environments (Holmstrom and Milgrom, 1991). The empirical evidence is mixed, with some studies reporting findings in line with the standard neoclassical theoretical prediction (e.g. Paarsch and Shearer, 2000; Al-Ubaydli et al., 2015; Hong et al., 2018), while others finding no adverse effect due to piece rate compensation on the unincentivized dimension (Shearer, 2004; Copeland and Monnet, 2009). The issue has attracted considerable attention with regard to healthcare contracting, where multitask agent problems are indeed ubiquitous (Chalkley and Malcomson, 1998; see Scott et al., 2011, for a review). This literature has also underlined the importance of sorting effects for incentive pay (Lazear, 1986), with empirical evidence supporting the theoretical findings (Lazear, 2000). Dohmen and Falk (2011), in particular, show with a lab experiment the importance of multidimensional sorting of workers, along the risk preferences and self-assessment dimensions.

Second, we connect to the growing experimental literature that documents the importance of prosocial motivation and incentives for performance (Tonin and Vlassopoulos, 2010, 2015; Imas, 2014; Charness et al., 2016; Carpenter and Gong, 2016; DellaVigna and Pope, 2017; Cassar, 2018). Finally, a related strand has been concerned with the effect that financial incentives may have on the selection of motivated workers (Francois, 2007; Delfgaauw and Dur, 2007, 2010; Prendergast, 2007; Dal Bo et al., 2013; Banuri and Keefer, 2016, Ashraf et al., 2018; Barigozzi et al., 2018).

In a nutshell, relative to the existing literature our contribution is that we design an experiment that encompasses all of the key dimensions that have been previously studied separately: pay-for-performance incentives, multitasking, prosocial motivation and sorting. This allows us to examine the incentive effect of performance pay on the incentivized and non-incentivized dimensions, both in a context where prosocial motivation may be active and when it is not. Furthermore, we can analyze the sorting effects of performance pay both in terms of ability and the prosociality of workers drawn by it. To the best of our knowledge, this is the first study to address all of these dimensions in a common environment, allowing us to drive comparisons regarding the effects of pay-for-performance under different motivational contexts.

The rest of this paper is organized as follows: the next section describes the experimental design, while section 3 lays out a simple theoretical framework and derives behavioural predictions. In section 4 we present the experimental results. Section 5 offers some concluding remarks and the Appendix includes additional tables and the experimental instructions.

2. Experimental Design

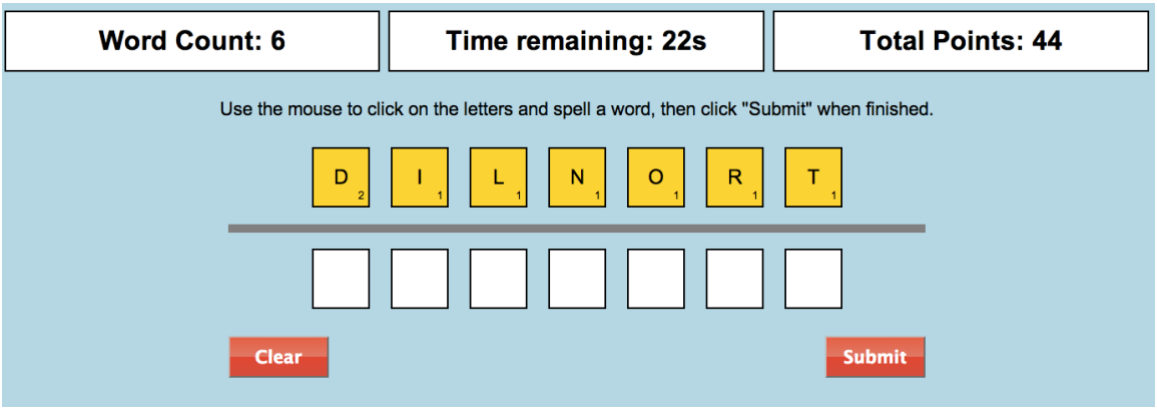
The experiment involves a real effort task performed in three stages followed by a questionnaire administered at the end. Participants were informed about the nature of the task and the structure of the experiment at the beginning but received detailed instructions about each stage at the beginning of the respective stage. The end-of-experiment questionnaire (in Appendix) included incentivised questions eliciting risk preferences based on the instrument developed by Holt and Laury (2002), a dictator game with a chosen local charity as recipient, information about demographic characteristics and a set of questions proposed by Perry (1996) measuring public service motivation (PSM). In what follows, we describe the novel word-formation task, the stages of the experiment, the treatments and some procedural details.

2.1 Real Effort Task

We employ a novel word-formation real effort task. Participants are presented with a set of seven letters, which they must use to spell a word using all or a subset of the letters (at least two). When a participant spells and submits a word a new set of letters appears on the screen. Letters have points attached to them that are indicated on the screen in the lower right-hand corner of each letter square, like the real-life word game Scrabble (see Figure 1 for a Screenshot). This opens the possibility to measure participants' performance on the task both in terms of quantity (number of words completed) and of quality (points per word). Only words that are grammatically correct are permissible and if the submitted word is incorrectly spelled the chosen letters are cleared and the participant is asked to try again. Participants receive feedback about the number of words spelled, the total points earned and the time remaining at the top of the screen.

The main feature of the task that is crucial for our purposes is that it allows us to measure both quantity and quality produced, with a worker facing a meaningful trade-off between the two. Note that to ensure that participants would face a meaningful trade-off between spelling a large number of words and accumulating points, we placed the restriction that the points associated with a certain word would be credited only if the total points of the word are at least 5.

Figure 1: Screenshot of Word Formation Task



2.2 Stages

The experiment involves 3 stages aimed at measuring ability, the incentive effect and the sorting effect of the incentive schemes.

- Stage 1 (Ability measurement): The purpose of this stage is to familiarize participants with the task and to elicit measures of ability on the task. We use a series of 5 different steps to capture different aspects of ability on the task: (i) complete 5 word formation tasks in 2 minutes (no compensation), (ii) complete 5 tasks in 60 seconds (participants start with an endowment of experimental currency units and are penalized the longer it takes to complete the 5 tasks), (iii) complete as many word formation tasks as possible in 3 minutes (receive piece rate for each word), (iv) complete five word formation tasks (compensation based on score accumulated for words worth five or more points), (v) complete as many word formation tasks as possible in 3 minutes (compensation based on score accumulated for words worth five or more points).
- Stage 2 (Incentive effect measurement): Participants complete the word-spelling task for 8 minutes. Payment differs by treatment (described below).
- Stage 3 (Sorting effect measurement): Participants spend another 8 minutes doing the word task. Before starting, they can choose to continue with the treatment-assigned payment scheme (as they saw in phase 2) or move to an outside option (common across treatments) in which they would receive a flat payment of $9. In this stage, we are primarily interested in the choice of payment scheme that participants made.

Final payment for participants consisted of a $5 show-up fee, plus their payoff from the first stage and the payoff of a randomly chosen of the two other stages.

2.3 Treatments

We implement a 2*2 between-subject design, depicted in Figure 2, in which we manipulate two aspects of stage 2: the incentive scheme (flat or piece rate) and whether there is an opportunity to raise money for a charity (we refer to the presence of a charity as *mission* and absence as *non-mission*). In particular, the treatments are as follows:

- F-NM (flat-rate, non-mission): participants are paid $7 for working during Stage 2 regardless of the number of words spelled (or points accumulated).
- P-NM (piece-rate, non-mission): participants receive $0.10 per word spelled.
- F-M (flat-rate, mission): participants are paid $7 for working during Stage 2 regardless of the number of words spelled (or points accumulated). In addition, participants generate $0.02 for a local charity of their choosing for each *point* that they accumulate, with points associated with each word credited only if they are five or greater.
- P-M (piece-rate, mission): participants receive $0.10 per word spelled. In addition, participants generate $0.02 for a local charity of their choosing for each point that they accumulate, with points associated with each word credited only if they are five or greater.

It is noteworthy, that to dispel any potential confusion about the purpose of having points assigned to letters in the two non-mission treatments, we opted to tell participants in all treatments: "We are hoping to learn how many points people can accumulate in a fixed period of time".[3] This aligns expectations about the objective of the principal and mimics real workplace conditions involving quantity-quality trade-offs where while quantity may be directly incentivized because of easier measurement, it is usually understood by workers that the employer is also concerned about quality.

We summarize the timeline of the experiment across these four treatments in Figure 3.

[3] A full set of Experimental Instructions is reproduced in Appendix B.

Figure 2: Treatment Conditions

		Sector: Non-mission	Sector: Mission
Incentive Scheme	Flat-rate	F-NM flat-rate, non-mission (N=29)	F-M flat-rate, mission (N=31)
	Piece-rate	P-NM piece-rate, non-mission (N=31)	P-M piece-rate, mission (N=31)

Figure 3: Summary of the timeline of the experiment

TREATMENT	STAGE 1	STAGE 2: Do word tasks for 8 minutes, earning…	Just before Phase 3…	STAGE 3: Do word tasks for 8 more minutes, earning…
F-NM	Ability measurement: Series of word formation tasks with incentives common across treatments	$7	Choose outside option ($9 flat, no charity) - or - Incentives faced in Phase 2	$9 (no charity), or… $7
F-M		$7, plus money for charity as a function of quality		$9 (no charity), or… $7, plus money for charity as a function of quality
P-NM		$0.10/word		$9 (no charity), or… $0.10/word
P-M		$0.10/word, plus money for charity as a function of quality		$9 (no charity), or… $0.10/word, plus money for charity as a function of quality

2.4 Procedures

The experimental sessions were conducted in the Behavioral Lab at the University of South Carolina. Payments were expressed in terms of experimental currency units (ECUs), which had an exchange rate of 10 ECUs to $1. We conducted 15 sessions total, with roughly 8 participants per session on average.[4] The mean earnings across all participants was roughly $16. The typical session lasted roughly 75 minutes. Participants were recruited largely from introductory (freshman/sophomore-level) economics courses from the business school at the University of South Carolina. Participation was voluntary and was not linked to grades in their introductory courses.

3. Predictions

This section presents a simple theoretical framework to generate theoretical predictions. Consider an agent who exerts effort on two dimensions $e = (e_1, e_2)$, where e_1 can be thought of producing quantity, while e_2 can be thought of producing quality. For simplicity, we assume that effort leads to outcomes linearly. Let us further assume that the cost of effort for the agent $c(e_1, e_2)$ is increasing and strictly convex with $\frac{\partial^2 c(e_1,e_2)}{\partial e_1 \partial e_2} > 0$, that is, the two efforts are substitutes. Furthermore, we assume that $c(e_1, e_2)$ attains a local minimum at an interior point $e^* = (e_1^*, e_2^*)$, so that, even without incentives the agent will exert some positive effort in the two tasks.

Agents are heterogeneous in two dimensions: ability on the task, denoted by γ, and prosociality, denoted by θ. Both characteristics are drawn from continuous distributions. We assume that ability enters the cost function multiplicatively, $\frac{1}{\gamma} c(e_1, e_2)$.

Agents facing flat-rate compensation therefore maximize:

$$w + m\theta e_2 - \frac{1}{\gamma} c(e_1, e_2)$$

where *w* is the flat monetary rate received regardless of effort and *m* is an indicator variable equal to 1 if the worker is employed in the mission-oriented sector.

Similarly, agents facing piece-rate compensation maximize:

[4] Because there was no strategic interaction between participants, the sessions did not require a particular number of participants to run. The number of participants varied across sessions, but the average number of participants per session is relatively constant across treatments.

$$pe_1 + m\theta e_2 - \frac{1}{\gamma}c(e_1, e_2)$$

where p is the piece-rate that dictates how effort on the quantity dimension is translated into payment.

We can use this simple framework to derive some predictions as to the relative levels of effort exerted in the various treatments.

3.1 Stage 2

Flat-rate Non-Motivated (FNM)

In the absence of any pay-for-performance in the non-motivated condition, the agent will choose $e^* = (e_1^*, e_2^*)$, given by $\frac{\partial c(e_1,e_2)}{\partial e_1} = \frac{\partial c(e_1,e_2)}{\partial e_2} = 0$. Effort on either dimension is not expected to vary with ability or prosociality. This treatment forms a baseline against which we can compare the other treatments below to form predictions.

Piece-rate Non-Motivated (PNM)

In this treatment, the agent is offered a piece rate p, in the quantity dimension. The first-order conditions are given by $\frac{\partial c(e_1,e_2)}{\partial e_1} = \gamma p$ and $\frac{\partial c(e_1,e_2)}{\partial e_2} = 0$. Given that e_1 and e_2 are substitutes, it follows that the agent will increase effort in the quantity dimension at the expense of quality and that effort in the quantity dimension will be increasing in ability (can be shown formally by total differentiation of F.O.C.s).[5] We thus predict that:

Prediction 1a: $e_1^{FNM} < e_1^{PNM}$ *and* $e_2^{FNM} > e_2^{PNM}$

Prediction 1b: $\frac{de_1^{PNM}}{d\gamma} > 0$

[5] Totally differentiating the F.O.C.s with respect to p yields: $\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_1}\frac{de_1}{dp} + \frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_2}\frac{de_2}{dp} - \gamma = 0$ and $\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_1}\frac{de_1}{dp} + \frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}\frac{de_2}{dp} = 0$. Solving the second equation for $\frac{de_2}{dp}$ and substituting into the first gives after rearrangement $\frac{de_1}{dp} = \frac{\gamma \frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}}{\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_1}\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2} - \left(\frac{\partial c(e_1,e_2)}{\partial e_1 \partial e_2}\right)^2} > 0$, because of the convexity of the cost function. Also, $\frac{de_2}{dp} = -\frac{\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_1}\frac{de_1}{dp}}{\frac{\partial c(e_1,e_2)}{\partial e_2 \partial e_2}} < 0$. Thus, relative to the case of a flat wage where the agent chooses $e^* = (e_1^*, e_2^*)$, given by $\frac{\partial c(e_1,e_2)}{\partial e_1} = \frac{\partial c(e_1,e_2)}{\partial e_2} = 0$, introducing a piece rate p increases e_1 and decreases e_2.

Flat-rate Motivated (FM)

In this treatment, we assume that a prosocial agent will derive an intrinsic benefit from producing quality, which we denote by θe_2. The agent thus maximizes $\theta e_2 \cdot c(e_1, e_2)$, with first-order conditions given by $\frac{\partial c(e_1,e_2)}{\partial e_1} = 0$ and $\frac{\partial c(e_1,e_2)}{\partial e_2} = \gamma\theta$. Given the substitutability of effort we thus predict that (can be shown formally by total differentiation of F.O.C.s as above):

Prediction 2a: $e_1^{FM} < e_1^{FNM}$ *and* $e_2^{FM} > e_2^{FNM}$

Prediction 2b: $\frac{de_2^{FM}}{d\theta} > 0$ *and* $\frac{de_2^{FM}}{d\gamma} > 0$

Piece-rate Motivated (PM)

In this treatment, the agent receives a piece-rate (p) for quantity and derives an intrinsic benefit θ from producing quality. The agent thus maximizes $pe_1 + \theta e_2 - \frac{1}{\gamma}c(e_1, e_2)$, with first-order conditions given by $\frac{\partial c(e_1,e_2)}{\partial e_1} = \gamma p$ and $\frac{\partial c(e_1,e_2)}{\partial e_2} = \gamma\theta$. We thus predict that:

Prediction 3a: $e_1^{PM} > e_1^{FM}$ *and* $e_2^{PM} < e_2^{FM}$

Prediction 3b: $\frac{de_1^{PM}}{d\gamma} > 0$ *and* $\frac{de_2^{PM}}{d\theta} > 0$ *and* $\frac{de_2^{PM}}{d\gamma} > 0$

3.2 Stage 3

We next generate predictions about choice of payment scheme in stage 3.

Flat-rate Motivated (FM)

The agent chooses FM when the utility it furnishes is higher than or equal to the utility from the outside option:

$w + \theta e_2^{FM} - \frac{1}{\gamma}c(e_1^{FM}, e_2^{FM}) > w^O - \frac{1}{\gamma}c(e_1^O, e_2^O)$, *where* $w < w^O$.

$$\theta > \frac{w^O - w + \frac{1}{\gamma}[c(e_1^{FM}, e_2^{FM}) - c(e_1^O, e_2^O)]}{e_2^{FM}}.$$

This suggests that for a given level of ability γ sorting on prosociality type will take place, as agents with sufficiently high levels of prosocial motivation would choose FM.

Piece-rate Non-Motivated (PNM)

The agent chooses the piece rate scheme if:

$$pe_1^{PNM} - \frac{1}{\gamma} c(e_1^{PNM}, e_2^{PNM}) > w^O - \frac{1}{\gamma} c\left(e_1^O, e_2^O\right), \quad with\ c(e_1^{PNM}, e_2^{PNM}) > c\left(e_1^O, e_2^O\right)$$

$$\gamma > \frac{c(e_1^{PNM}, e_2^{PNM}) - c\left(e_1^O, e_2^O\right)}{pe_1^{PNM} - w^O}$$

This suggests that there will be sorting on ability. The proposed scheme is more likely to be chosen by high ability types.

Piece-rate Motivated (PM)

In the mission case, the agent chooses the piece rate scheme if:

$$pe_1^{PM} + \theta e_2^{PM} - \frac{1}{\gamma} c(e_1^{PM}, e_2^{PM}) > w^O - \frac{1}{\gamma} c\left(e_1^O, e_2^O\right), with\ c(e_1^{PM}, e_2^{PM}) > c\left(e_1^O, e_2^O\right)$$

$$\theta > \frac{w^O - pe_1^{PM} + \frac{1}{\gamma}[c(e_1^{PM}, e_2^{PM}) - c(e_1^O, e_2^O)]}{e_2^{PM}} = \hat{\theta}$$

or

$$\gamma > \frac{c(e_1^{PM}, e_2^{PM}) - c\left(e_1^O, e_2^O\right)}{pe_1^{PM} + \theta e_2^{PM} - w^O} = \hat{\gamma}$$

The above inequalities suggest that the proposed scheme is more likely to be chosen by high θ and/or γ agents. Note also that $\frac{d\hat{\theta}}{d\gamma} < 0$ and $\frac{d\hat{\gamma}}{d\theta} < 0$, suggesting that ability and motivation are substitutes for sorting, that is, a higher ability individual requires less motivation to sort and similarly a higher motivated individual requires less ability to find sorting in optimal.

4. Results

This section reports results from our experiment. We start by providing simple summary statistics. Table 1 provides a summary of participant characteristics. Notably, given the recruiting approach, a large majority of participants have declared a business field as their major (marketing, management, economics, finance, etc.) and skew towards earlier years in college (see "Year in college" which is 1 for freshmen and 4 for seniors; the average participant was a freshman or sophomore). We also report other survey and decision task responses in the table. Importantly for the sake of our real effort task, which hinges on spelling words, almost all participants are native English speakers. It is also worth noting that females are slightly overrepresented in our sample.

Table 1: Summary statistics – Participant characteristics

Female	0.61	Any monetary donations? (past year)	0.75
	(0.49)		(0.43)
Age	19.56	Any goods donations? (past year)	0.83
	(2.13)		(0.38)
Year in college	1.89	Any volunteering? (past year)	0.80
	(0.95)		(0.41)
Business major	0.73	Any blood donations? (past year)	0.37
	(0.45)		(0.48)
Native English?	0.92	Holt-Laury: Share of risky choices	0.53
	(0.28)		(0.20)
Currently works?	0.43	Dictator: Share contributed	0.46
	(0.50)		(0.32)
Current wage (conditional on current work)	$12.18	PSM Survey score (0-1, higher = more PSM)	0.69
	(19.75)		(0.09)
Observations: 122			

Standard deviations in parentheses

Recall that Stage 1 of our experiment consisted of a set of five tasks aimed at measuring participants' ability. Importantly, the tasks faced in this stage were constant across treatments, and any differences across treatments in the incentives participants would face in the following stages were not yet described. It is therefore a pure measure of ability, free of any potential treatment effects, which is directly comparable across all participants.

The first of the five tasks in Stage 1 was non-incentivized and was primarily meant to provide participants with practice using the interface. Tasks 2 and 3 incentivized participants to spell many words in a short period of time, thereby measuring participants' ability in the quantity dimension. The remaining two tasks incentivized participants to accumulate as many points as possible in a short period of time, thereby measuring participants' ability in the quality dimension. From these tasks, we construct a measure of participants' ability in each of these dimensions, by standard-normalizing each participant's earnings in each task. We then sum the standard-normalized earnings from Tasks 2 and 3, and standard-normalize again, to construct a measure of ability in the quantity dimension. We do the same for Tasks 4 and 5 to construct a quality ability measure. We also construct a composite ability measure, summing tasks 2-5 and standard normalizing.

Figure 4 graphically depicts the relationship between the quantity ability measure and the quality ability measure. Notably, there is a substantial positive correlation between the two measures, making it unclear whether they are genuinely picking up distinct forms of ability in the task. For this reason, and

also for ease of interpretation of later empirical results and comparison with our theory model (which features a single variable to capture ability), we use the composite measure of ability – which sums all four incentivized tasks from Stage 1 – throughout the remainder of the paper.

Figure 4: Scatterplot depicting the relationship between the quantity and quality ability measures

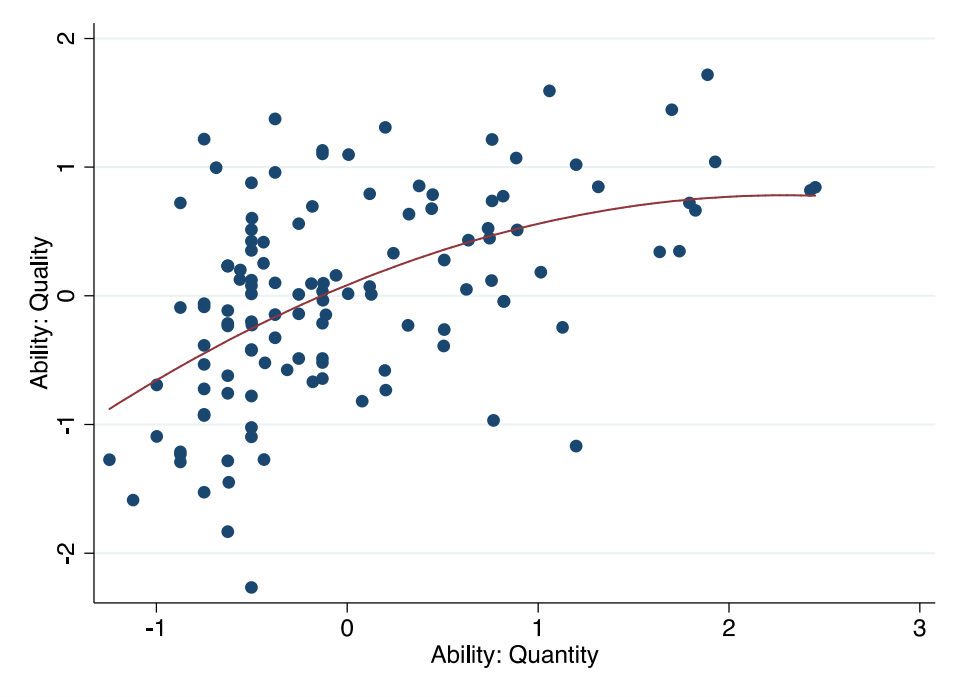


Figure note: The solid red line is a quadratic fit.

Finally, as we test for heterogeneity in participants' response along both the ability dimension (using the composite ability measure just described) and along the motivation dimension (primarily using the public service motivation (PSM) survey score), it is worth documenting the relationship between these two measures in our setting. Figure 5 graphically depicts the relationship between the measures, with Ability on the x-axis and the PSM Survey Score on the y-axis. The average score, 0.69, on the PSM survey is reported in Table 1; as noted there, higher scores indicate more public service motivation. Figure 5 shows that there is substantial variation in the PSM Score across participants. Importantly, there does not appear to be any clear relationship between the PSM Score and Ability in our setting. Any potential correlation between ability and motivation obviously impacts the degree to which we may expect, and the degree to which it is desirable to observe, high ability workers to sort into the mission-oriented job with performance pay. The fact that these measures are independent in our setting will help us separately identify sorting of workers along these separate dimensions.

Figure 5: Scatterplot depicting the relationship between the composite ability measure and the Public Service Motivation survey score

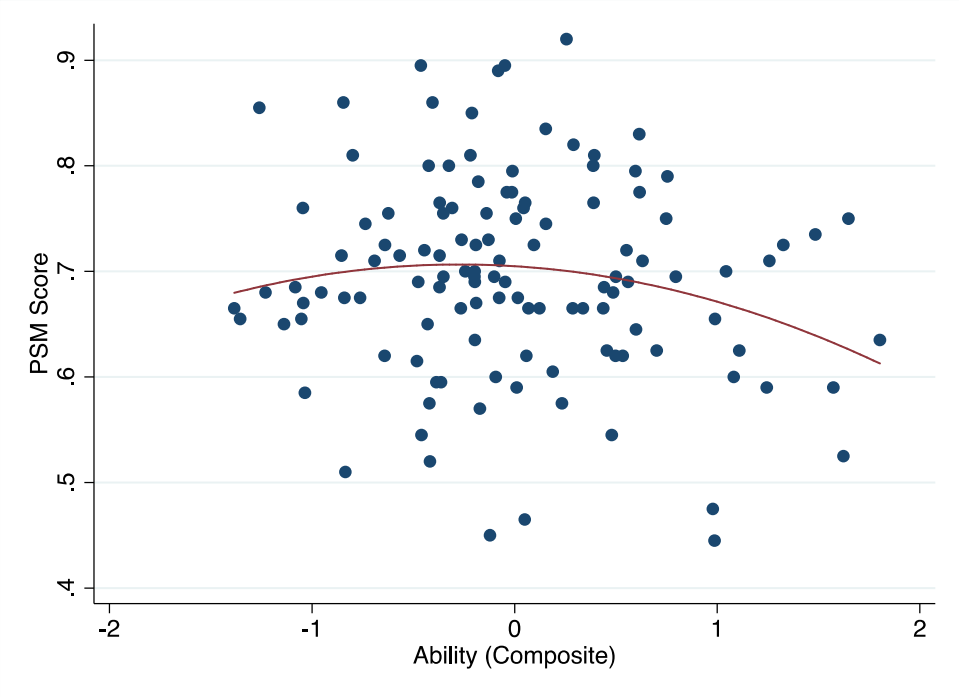


Figure note: The solid red line is a quadratic fit.

The remainder of the section assesses differences in participant behaviour across treatments. First, we will present the basic averages of outcomes. Second, we will use regression analysis to statistically test for treatment effects in Phase 2 of the experiment in response to randomly assigned payment schemes; that is, we will study the "incentive effects" of performance pay in mission-oriented vs. non-mission-oriented tasks. Finally, we will turn to exploring (both graphically and with regressions) the types of participants that opt for the randomly assigned payment scheme over a common outside option, assessing "sorting effects" of performance pay in our setting.

4.1 Summarizing results: Averages of outcomes across treatments

Table 2 reports the averages of our main outcome variables across the four treatments. Figures 6a, 6b, 6c, and 6d graphically depict the same averages for a subset of the outcomes reported in Table 2. Three broad findings are readily observable from Table 2 and the figures.

First, effort on the incentivized dimension (quantity of words spelled) is much higher in the piece-rate non-motivated treatment (PNM) than any other treatment. The number of words spelled is somewhat higher in the piece-rate motivated (PM) treatment than the comparable flat-rate (FM) treatment, but the difference is much smaller than the difference between the two non-motivated treatments. We explore

the statistical significance of these differences in the next subsection, but it is already clear that performance pay has a larger positive impact on the incentivized dimension of effort in the non-mission-oriented setting.

Table 2: Averages of outcome variables by treatment

TREATMENT:	Flat-rate Motivated (FM)	Flat-rate Non-Motivated (FNM)	Piece-rate Motivated (PM)	Piece-rate Non-Motivated (PNM)
Phase 2: Words spelled	31.16	33.76	38.71	59.32
	(10.25)	(13.13)	(19.51)	(30.90)
Phase 2: Points per word	6.94	6.69	6.54	5.65
	(0.57)	(0.61)	(0.68)	(1.13)
Phase 2: Counted points per word	6.57	6.21	6.03	4.69
	(0.71)	(0.82)	(0.96)	(1.66)
Phase 2: Counted points per counted word	7.36	7.16	7.06	6.64
	(0.54)	(0.53)	(0.47)	(0.59)
Phase 2: Share of words that are counted (>4 pts.)	0.89	0.87	0.85	0.69
	(0.07)	(0.08)	(0.11)	(0.20)
Phase 3: Chose treatment-assigned payment?	0.61	0.03	0.39	0.35
	(0.50)	(0.19)	(0.50)	(0.49)
Observations	31	29	31	31

Standard deviations in parentheses

Figure 6a: Total words spelled (Phase 2)

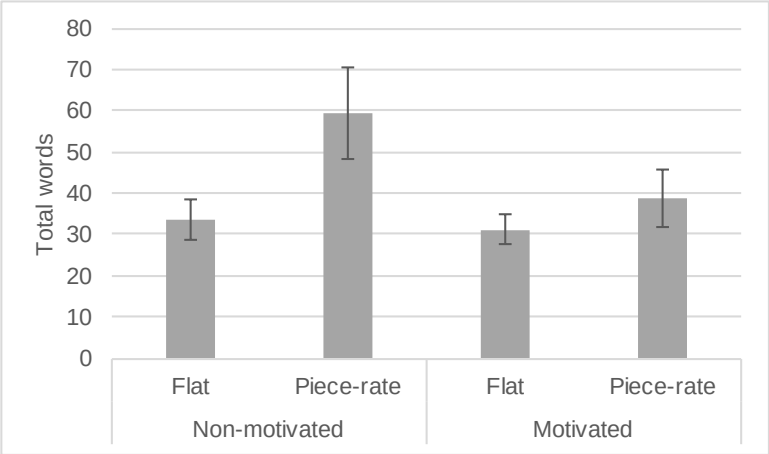


Figure 6b: Points per word (Phase 2)

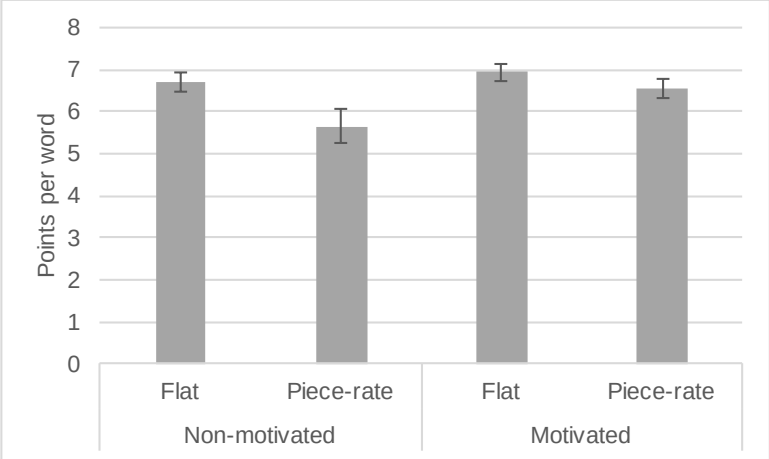


Figure 6c: Counted points per word (Phase 2)

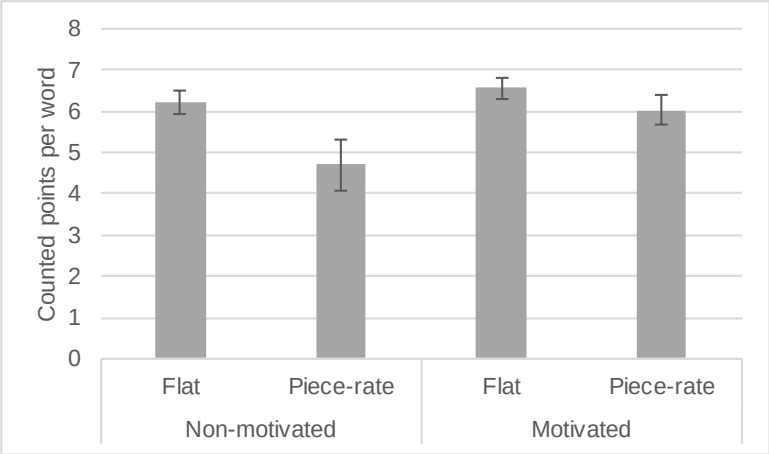
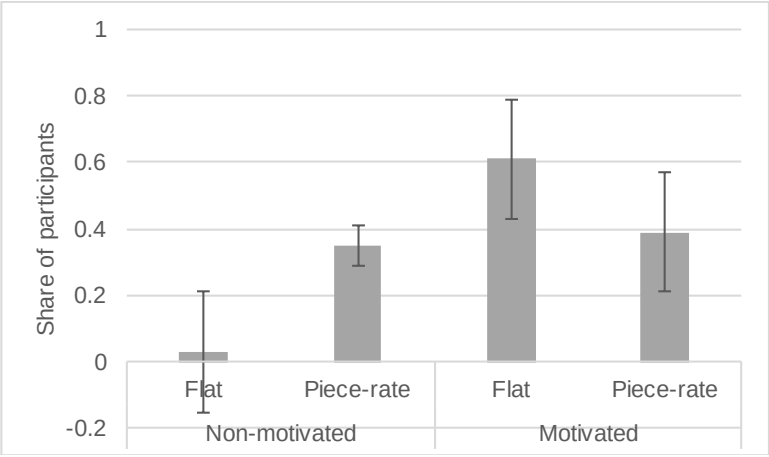


Figure 6d: Share of participants choosing treatment-assigned pay scheme (over common outside option)



Second, the quality of words (as measured by the average number of points associated with each word) spelled is lower in the PNM treatment than in any other treatment (and is relatively similar across the remaining three treatments). This can be seen in Figures 6b and 6c, or the second through fifth rows of Table 2, which present several different ways of measuring the "quality" of words.

"Points per word" is the simplest method of measuring average quality: we divide total points accumulated (even for words worth less than 5 points, which are worth 0 points from the participant's perspective) by total words spelled. "Counted points per word" divides total points accumulated (counting words worth less than 5 points as 0 points, as would be the experience of the participant) by total words spelled. This, then, is closer to the measure the participants would experience. The final two measures in Table 2 are aimed at assessing the margin along which participants changed their behavior to increase points (and therefore money for charity, in the motivated treatments). Participants might have focused on spelling higher quality words (an intensive margin response), or they might have focused on making sure more of their words were at least five points (an extensive margin response). To get at this, we construct "Counted points per counted word", which divides total points accumulated from words worth at least five points by the number of words spelled that were worth at least five points, and "Share of words that are counted", which is the number of words worth at least five points divide by the total words spelled. The first of these would reveal the intensive margin response, the second would reveal the extensive margin response. In practice, we find that both follow similar patterns, so participants respond both by increasing the average quality of words (even conditional on the word already achieving five points) and by increasing the fraction of words that count. Because patterns of results are similar across these different measures of quality, we mostly report the simplest measure ("Points per word") moving forward.

Recall that, in the "motivated" treatments, more money for charity is generated by spelling words worth more points. This non-incentivized quality dimension of effort appears to be lowest in PNM, where effort on the incentivized quantity dimension is highest, indicating that participants faced a trade-off between quantity and quality. They acted on this trade-off to a greater degree in the PNM treatment than in the PM treatment, where quantity is slightly higher than in FM, but with no large drop in quality.

Finally, the last row of Table 2 (and Figure 6d) reports the choice of payment scheme that participants opted for in Phase 3. In particular, we report the share of participants who opt for the treatment-assigned payment scheme (e.g., piece-rate with money raised for charity) over the common outside option (flat-rate $9, with no opportunity to raise money for charity). Note that the $9 flat rate from the outside option is higher than the $7 flat rate from the FM and FNM treatment-assigned payment schemes. It is not

surprising then to see that all but one of the participants in the FNM treatment-assigned payment scheme preferred the outside option. More interesting is the fact that a large share of participants (61%) in the FM treatment opted for the treatment-assigned payment scheme in Phase 3; that is, a majority of participants in that treatment were willing to accept a lower flat rate for the opportunity to raise money for charity. Lower rates of participants opted for the treatment-assigned payment schemes in the two piece-rate treatments. The rate of participants choosing the treatment-assigned scheme is relatively similar across PM and PNM, which may seem surprising given the differences in behavior, discussed above, in Phase 2 across the two treatments. In a later subsection, we explore *who* is opting for the treatment-assigned option and see that there are differences across the two treatments that run deeper than simply the numbers of participants choosing each option.

Based on these initial comparisons of averages, we see preliminary evidence consistent with Theoretical Prediction 1: that is, relative to FNM, PNM leads to higher quantity but lower quality. The evidence that speaks to Theoretical Prediction 2 is less clear; at least based on simple means comparisons, both quantity and quality are relatively similar across the FM and FNM treatments. In the next section, we test for statistical differences across these treatments. Similarly, Theoretical Prediction 3 (predicting higher quantity and lower quality in PM vs. FM) will require statistical tests of differences. Finally, the simple means comparisons do not allow us to speak to our theoretical predictions regarding Phase 3 sorting choices; subsection 4.3 will address those predictions directly.

4.2 Assessing incentive effects in Stage 2

Table 3 presents a series of regressions which allow us to test for statistical differences in effort on the incentivized (quantity) and non-incentivized (quality) dimensions across treatments. We simply regress an outcome (e.g., total words spelled in Column 1) on treatment indicator variables. FNM is the omitted treatment, so all tests are relative to that treatment. The bottom of the table reports noteworthy tests of differences between other coefficients in the table (e.g., whether the PM coefficient is significantly different than the FM coefficient). In Appendix Table A1, we report a similar series of regressions that include participant-level controls (gender, the Phase 1 ability measure, etc.). Appendix Table A2 reports results from nonparametric tests of differences across treatments for our two main outcome variables (total words spelled and points per word). Results across both of these tables are consistent with the results reported in Table 3.

Table 3: Testing differences in Phase 2 outcomes across treatments

VARIABLES	(1) Total words	(2) Points per word	(3) Counted points per word	(4) Counted points per counted word	(5) Share of counted words (5 pts or greater)
Treatment: F-NM	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
Treatment: F-M	-2.60	0.25	0.36*	0.20	0.03
	(3.05)	(0.15)	(0.20)	(0.14)	(0.02)
Treatment: P-NM	25.56***	-1.04***	-1.52***	-0.51***	-0.17***
	(6.06)	(0.23)	(0.33)	(0.15)	(0.04)
Treatment: P-M	4.95	-0.15	-0.17	-0.09	-0.01
	(4.27)	(0.17)	(0.23)	(0.13)	(0.02)
Observations	122	122	122	122	122
R-squared	0.24	0.29	0.30	0.20	0.28
PM vs. FM p-val.	0.06	0.01	0.01	0.02	0.09
PM vs. PNM p-val.	0.00	0.00	0.00	0.00	0.00

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

We frame this discussion around the theoretical predictions. Theoretical Prediction 1 speaks to the comparison between piece-rate and flat-rate payment in a non-mission-oriented setting. In our experiment, that is the comparison between treatments PNM and FNM. The theoretical prediction suggests that quantity should be higher in PNM than FNM, while quality (which is sacrificed to increase quantity) should be lower. We find clear evidence of both pieces of this prediction. Column 1 shows that participants in the PNM treatment spell an additional 25.56 words relative to the FNM treatment. Recall that the mean of "words spelled" in the FNM treatment was 31.16, so piece-rate payment nearly doubles the quantity of words spelled. The simple comparison of averages suggested that this increase in quantity comes at the expense of quality, and the statistical tests for differences in Table 3 bear that out: regardless of which measure of quality is used in Columns 2-5, quality is significantly lower in PNM.

Theoretical Prediction 2 focused on the difference between flat-rate payment in mission-oriented vs. non-mission-oriented settings, or FM and FNM. We predicted higher quantity in FNM, but higher quality in FM. The basic intuition was that, in order to increase effort on quality, workers in FM would have to reduce effort on quantity. We ultimately find very little evidence of differences across the two treatments. There is no statistical difference across treatments in Total Words Spelled (column 1), nor in the quality of words (Columns 2-5, with the exception of a marginally significant increase in quality in Column 3).

Theoretical Prediction 3 focused on the impact of performance pay in a mission-oriented setting. There, as in the non-mission-oriented setting, we predicted that performance pay would positively impact quantity but not quality. Here, the relevant test is the difference between the PM and FM coefficients, the p-value of which is reported in the bottom portion of the table. We do find evidence that quantity (Column 1) is higher in PM than FM, with PM participants spelling 7.55 additional words on average (p-value = 0.06). While this is consistent with our theoretical prediction, it is worth noting that this response is substantially more muted than in the non-mission-oriented treatments. There, as we noted above, performance pay led to an additional 25.56 words. We observe a similar phenomenon with respect to the quality of the words: PM participants spell words worth 0.4 fewer points than FM participants. Although this difference is significant (p-value=0.01), the magnitude of the effect of performance pay is much smaller than the parallel comparison between PNM and FNM, where points per word were 1.04 points lower with performance pay. Thus, while there is evidence in favor of the theoretical prediction, the more notable finding is that the effects of performance pay are much smaller in the mission-oriented setting than in the non-mission-oriented setting.

Table 4 tests for heterogeneity in response to performance pay by worker ability. Recall that Phase 1 of the experiment consists of a variety of tasks aimed at measuring the ability of workers, and that we have constructed an ability measure by summing earnings from across the tasks in Phase 1 and then standard-normalizing the resulting sum.

In Table 4, we modify the specification reported in Table 3, interacting treatment indicators with our constructed ability measure. Column 1 shows, unsurprisingly, that higher ability workers spell more words. This is true across all treatments, which is consistent with theoretical predictions 6b, 7b, and 3b, but runs counter to our prediction for the F-NM treatment which stated that effort would not be a function of ability. More importantly, the results suggest that there is no statistical difference in the relationship between ability and total words spelled across the F-NM, F-M, and P-NM treatment.[6] However, the relationship between ability and words spelled is significantly different in the P-NM treatment; there, higher ability treatment has roughly twice the marginal impact than it does in other treatments. A one standard deviation increase in ability leads to 9.73 additional words in the F-NM treatment. The same increase in ability leads to 19.25 additional words in the P-NM treatment.[7] In other words, in the non-mission-oriented setting, performance pay pushes the most able participants to dramatically increase their output beyond the higher level they would be producing anyway. This echoes theoretical predictions from Lazear (1999) who notes that "a piece rate allows the more able to work harder and receive more from the job". The fact that the same stark increase in the marginal impact

[6] The "F-M X Ability" and "P-M X Ability" coefficients are not statistically different than zero.
[7] Here, we take the linear combination of the coefficients "Phase 1 Ability" and "P-NM X Ability".

of ability is *not* true of workers in the P-M treatment provides some of the first evidence that the way performance pay is working in the mission-oriented setting is quite different than what is typically expected.[8]

Table 4: Heterogeneity by ability on incentivized and non-incentivized dimensions of productivity

VARIABLES	(1) Total words	(2) Points per word
Treatment: F-NM	(omitted)	(omitted)
Treatment: F-M	-1.74	0.25
	(2.02)	(0.16)
Treatment: P-M	8.76***	-0.17
	(3.32)	(0.18)
Treatment: P-NM	23.86***	-0.96***
	(4.50)	(0.22)
Phase 1 Ability Measure	9.73***	0.04
(Standard-normalized)	(1.76)	(0.13)
Treatment: F-M X Ability	-1.40	-0.15
	(2.28)	(0.16)
Treatment: P-M X Ability	3.66	-0.13
	(3.11)	(0.17)
Treatment: P-NM X Ability	9.52**	-0.57***
	(3.97)	(0.20)
Observations	122	122
R-squared	0.59	0.38
PM vs. FM p-val.	0.00	0.01
PM vs. PNM p-val.	0.00	0.00

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

Column 2 shows that higher ability participants in PNM are more responsible for the large decline in points per word in that treatment than lower ability participants; this is likely driven by these workers shifting effort away from quality, instead taking advantage of their ability to increase quantity (and receive piece-rate payment for doing so). Interestingly, there is no relationship between ability and points per word in any other treatment. That is noteworthy because it suggests that, unlike in the P-NM treatment, higher ability workers are spelling more words without sacrificing the quality of the words they spell.

[8] The relevant comparison here is the difference between the "F-M X Ability" coefficient and the "P-M X Ability" coefficient, which measures the difference in the marginal impact of ability when performance pay is turned on conditional on being in a mission-oriented setting. These coefficients are statistically different at only the 10% level. In alternative specifications where we include additional controls (not reported), these coefficients are not statistically different while "P-NM X Ability" remains significant at the 5% level.

To summarize, we have documented that – at least with respect to incentive effects – performance pay has a substantially smaller positive effect on the incentivized dimension of effort (quantity) when workers are in our experimental mission-oriented sector compared to the case without mission. We also document that the large positive effect of performance pay on the incentivized dimension in the non-mission-oriented treatments comes at the expense of quality (a non-incentivized dimension of effort). While a similar result is found in the mission-oriented treatments, this effect is also much smaller. Thus, performance pay is less successful in increasing effort on the incentivized dimension when workers are motivated, but also does less harm to other dimensions of effort.

4.3 Assessing sorting effects in Stage 3

We now turn to assessing the types of participants who opt for performance pay when given the choice and how these sorting decisions differ depending on whether the work is in the mission-oriented or non-mission-oriented sector. Recall that Phase 3 consists of an additional eight minutes of the real effort task, but – unlike Phase 2 – we give participants the option to either: continue to work under the treatment-assigned payment scheme, or switch to a common outside option ($9 flat-rate payment with no funds raised for charity).

We begin by presenting a series of figures, which compare the average characteristics of workers choosing the treatment-assigned option to those choosing the outside option. In each of the figures that follow, we omit average characteristics for workers choosing the treatment-assigned option in the FNM treatment – as there was only one such worker. Figure 7a graphs the average words spelled in Phase 2 by treatment and by payment scheme choice. The darker red bars report average words spelled amongst workers who chose the common outside option, while the lighter (pink) bars report the average words spelled amongst workers who remained with the treatment-assigned payment scheme. First consider the PNM treatment. Aside from the multitasking element, that treatment most closely mirrors the type of setting that performance pay has typically been studied, as in Lazear (2000) for instance. Lazear (2000) theoretically predicts and empirically documents that the most productive workers will be attracted to performance pay, which provides the beneficial "sorting effect" of performance pay. Our theoretical model makes the same prediction (predicting that higher ability workers are more likely to choose the performance pay scheme if given the choice) and the averages depicted in Figure 7a support that prediction. The workers who choose the treatment-assigned option (in this case, piece-rate with no charity) spelled roughly 70 words in Phase 2, while those who opt for the outside option spelled roughly 50 words in Phase 2. Interestingly, no other treatment reveals sorting on the basis of past productivity.

Figure 7b allows us to consider whether workers sort based on the quality of work they had done in Phase 2. Again, the only clear difference (though we will test for statistical differences below) appears

in the PNM treatment. The workers who chose to remain with performance-pay put forth significantly less effort on the quality of their words than the workers who opted for the outside option. Of course, both Figures 7a and 7b report measures that are dependent on the treatment-assigned scheme workers faced in Phase 2. For instance, the difference in points per word across the two groups of participants in the PNM treatment is likely a reflection of the difference in total words spelled revealed in Figure 7a; Table 4 had of course already suggested that the two measures are related.

With this in mind, Figures 7c and 7d instead assess how participants choosing different payment schemes differ on characteristics that are not linked to the treatment they were assigned to, namely: our constructed ability measure from Phase 1 (Figure 7c) and a standard-normalized score from the Public Service Motivation survey (Figure 7d). Both of these measures were collected from tasks that were identical across treatments. Nonetheless, Figure 7c leads to a conclusion that is similar to that of Figure 7a: the participants in the PNM treatment who opted for performance pay were significantly higher ability (which in turn explains how they spelled more words) than the workers who opted for the outside option. This is direct evidence in favour of our theoretical prediction. Our theoretical prediction on the types of workers who would opt for performance pay in the mission-oriented sector was more ambiguous. Here we see no evidence that the workers opting for performance pay in the PM treatment are higher ability than the workers opting for the outside option. Thus, while the "incentive effects" of performance pay are present in both the mission-oriented and non-mission-oriented treatments (albeit much smaller in the mission-oriented treatments), the "sorting effects" of performance pay may only be present in the non-mission-oriented treatments.

Figure 7d tests whether there is any sorting by level of public service motivation (as elicited by our public service motivation (PSM) survey). The most striking difference occurs in the FM treatment: the workers who prefer the $9 flat-rate with no charity to the $7 flat-rate with charity score much lower on the survey. This lends some confidence to the idea that the PSM score we have is indeed related to participants' behaviour and preferences in the main portion of the experiment. There is some evidence that participants in the PM treatment who prefer to continue under performance pay (with charity) have higher PSM scores than those who opt for the outside option. Below, we test whether this difference is significant.

Figure 7a: Average words spelled in Phase 2 of workers choosing treatment-assigned vs. outside option

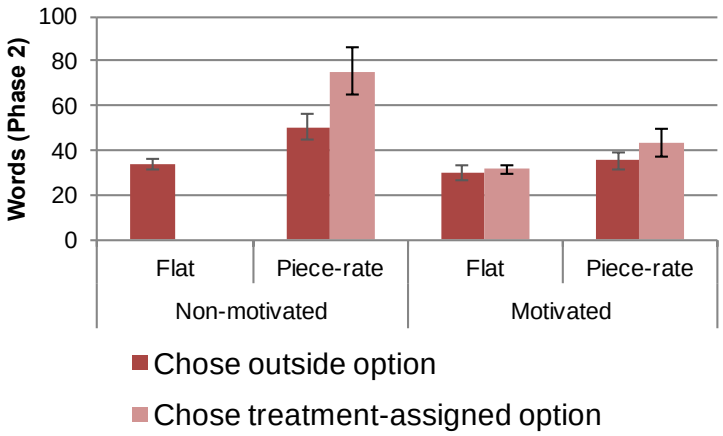


Figure 7b: Average points per word in Phase 2 of workers choosing treatment-assigned vs. outside option

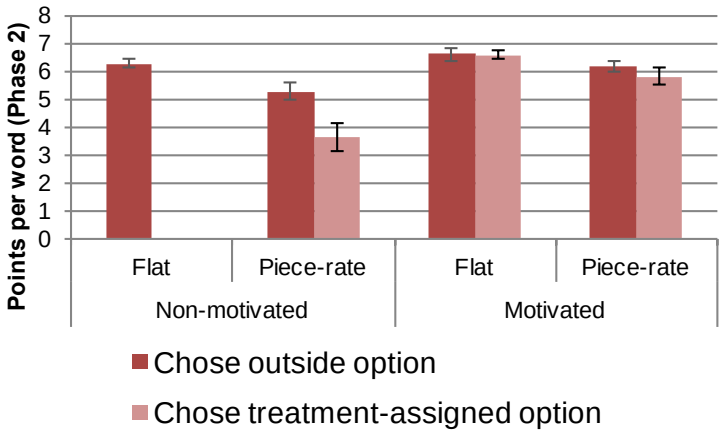


Figure 7c: Phase 1 Ability Measure (standard-normalized) of workers choosing treatment-assigned vs. outside option

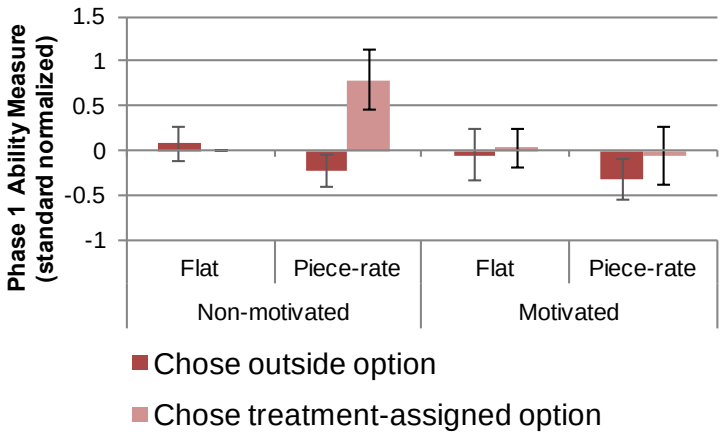


Figure 7d: Public Service Motivation score (standard-normalized) of workers choosing treatment-assigned vs. outside option

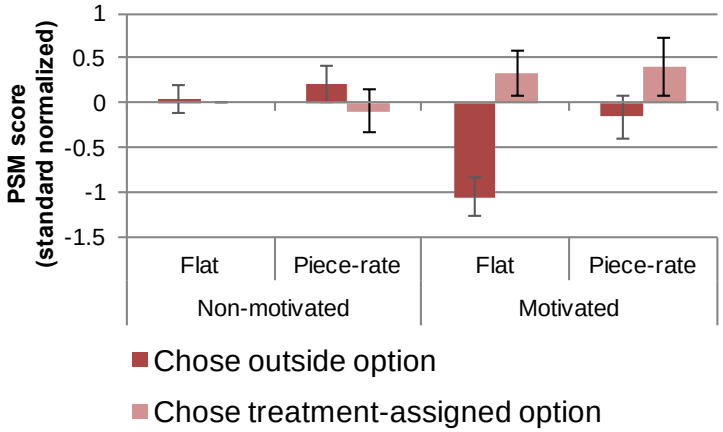


Table 5 reports results from regression analyses, where we test for differences across treatments in the characteristics of workers who prefer the treatment-assigned payment scheme over the common outside option. The simplest regression (Column 1) regresses a dummy variable indicating that the participant chose the treatment-assigned payment scheme on a set of treatment indicator variables (with FNM again serving as the omitted category). This reveals little new information, as it essentially reports the rates of participants choosing the treatment-assigned option, but serves as a baseline for the remaining regressions in the table.

Because we are interested in the types of workers who sort into the performance pay, our primary interest is in understanding how ability and motivation (as measured by the Phase 1 ability score and PSM score respectively, both standard-normalized) dictate the choice to choose the treatment-assigned option, especially in the PM and PNM treatments. In Columns 2 and 3, both of these measures enter as controls. In Column 2, we fully interact the treatment indicators with the ability measure to test how ability differentially impacts the decision to opt for the treatment-assigned payment scheme across treatments. Notably, the decision to opt for the treatment-assigned scheme is only meaningfully impacted by ability when the treatment-assigned scheme is PNM (or piece-rate with no charity). In that treatment, workers who are one standard deviation higher in ability are 20 percentage points more likely to choose to work under the piece-rate scheme. This is consistent with the patterns in Figures 7a and 7c. Interestingly, there is *no* relationship between ability and likelihood of choosing to work under performance pay in the PM treatment. That is, as in Figures 7a and 7c (and very much unlike the PNM treatment), workers in the mission-oriented treatment do not sort into performance pay on the basis of ability. There is no evidence of sorting on ability in FM or FNM, but there is also no reason to expect that there should be.

Column 3 allows for an interaction between treatment and the PSM score. PSM is only a significant determinant of choosing the treatment-assigned option in the FM treatment. Workers who are one standard deviation higher in their PSM score are 24 percentage points more likely to opt for the $7 flat-rate with charity than the $9 flat-rate without. While the PM X PSM coefficient is positive, it is not significant. Thus, just as there is no evidence of sorting into performance pay on the basis of ability when workers are motivated, there is also little evidence of sorting on the basis of public service motivation. On the other hand, there is sorting based on motivation when the choice is between a fixed pay without mission and a (lower) fixed pay with mission.

Table 5: Heterogeneity in ability and motivation of workers choosing treatment-assigned vs. outside option

VARIABLES	(1) Pr(Chooses treatment-assigned payment)	(2) Pr(Chooses treatment-assigned payment)	(3) Pr(Chooses treatment-assigned payment)
Treatment: F-NM	(omitted)	(omitted)	(omitted)
Treatment: F-M	0.58***	0.61***	0.64***
	(0.10)	(0.09)	(0.08)
Treatment: P-M	0.35***	0.37***	0.38***
	(0.10)	(0.10)	(0.10)
Treatment: P-NM	0.32***	0.29***	0.33***
	(0.09)	(0.09)	(0.09)
PSM score		0.11***	0.02
(standard normalized)		(0.03)	(0.03)
Phase 1 Ability Measure		0.05	0.10***
(Standard-normalized)		(0.03)	(0.04)
Treatment: F-M X Ability		0.01	
		(0.10)	
Treatment: P-M X Ability		0.01	
		(0.09)	
Treatment: P-NM X Ability		0.19**	
		(0.09)	
Treatment: F-M X PSM			0.24***
			(0.06)
Treatment: P-M X PSM			0.08
			(0.08)
Treatment: P-NM X PSM			-0.12
			(0.09)
Observations	122	122	122
R-squared	0.18	0.29	0.34
PM vs. FM p-val.	0.07	0.04	0.02
PM vs. PNM p-val.	0.80	0.51	0.67

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

To summarize the results from this subsection, we have essentially revisited the classic theoretical and empirical finding from Lazear (2000), which suggests that performance pay benefits firms, in part, by attracting higher ability workers. In our non-mission-oriented treatment, we replicate this result. However, in our mission-oriented setting, there is no evidence that performance pay attracts higher quality workers.

Pairing this with the results from previous subsections, our experiment replicates existing findings from non-mission-oriented settings more generally: we provide evidence of beneficial incentive and sorting effects. We do document that this comes at a cost: workers put forth less effort on the non-incentivized dimensions of their job. However, in comparing these findings to a mission-oriented setting, where workers generate money for charity by doing "high quality" work, we find some evidence of a more limited incentive effect and no evidence of a sorting effect.

5. Conclusions

We carry out a real effort lab experiment to address the following two questions: (i) How does performance pay impact productivity on incentivized and non-incentivized dimensions when workers are in a sector with a mission? (ii) How does performance pay impact the composition of workers in jobs with a prosocial dimension? We first show how, in absence of a mission, performance pay has strong positive effects on productivity on the incentivized dimension (quantity) and negative effects on the non-incentivized dimension (quality). When a mission is present, however, the effect of performance pay is much more subdued, with much smaller changes on either dimension. Finally, when workers can choose to remain in the experimentally-assigned payment scheme, or opt for an outside option with flat payment and no mission, we find that workers in the no mission sector sort on ability, with lower ability workers opting out of the P4P scheme. There is, instead, no evidence of sorting into performance pay on the basis of ability for workers in the sector with a mission, while sorting takes place along the motivation dimension when the choice is between a fixed wage without mission and a (lower) fixed wage with mission, where people with high public service motivation are willing to give up financial gains to be able to contribute to the mission.

How does this inform the debate on whether or not to introduce performance pay in sectors with a strong mission, like education or healthcare? On one side, our results suggest that it may be misleading to extrapolate the experience about high-powered incentives coming from standard sectors of the economy to sectors where motivation plays a more important role. For instance, those pushing for the adoption of P4P for teachers on the basis of its beneficial effects on productivity in the manufacturing sector may miss their target. On the other hand, our results also counter the argument that introducing pay-for-

performance may have ruinous effects in a mission environment characterized by multi-tasking. Returning to the example of the educational sector, our findings do not support arguments that with bonuses based on test scores teachers will be teaching solely to the test. We find that, when mission is present, the reallocation of effort away from the unincentivized dimension into the incentivized one is subdued. Regarding selection, by introducing pay for performance a mission sector may give up the selection on motivation we observe with fixed pay, without selecting particularly high-skilled workers (but, also, without adversely selecting low motivation workers). These results are of course related to the fact that, as shown in figure 5, motivation and ability are uncorrelated in our sample and this may not be the case in all contexts.

More generally, what we have shown is that the effects of pay-for-performance are not independent of mission. Therefore, whenever prosocial motivation matters, these two types of incentives should be studied together.

REFERENCES

Al-Ubaydli, O., Andersen, S., Gneezy, U. and List, J.A., 2015. Carrots that look like sticks: Toward an understanding of multitasking incentive schemes. *Southern Economic Journal*, *81*(3), pp.538-561.

Ashraf, N., Bandiera, O. and Lee, S., 2018. *Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services* (No. 88175). London School of Economics and Political Science.

Banuri, S. and Keefer, P., 2016. Pro-social motivation, effort and the call to public service. *European Economic Review*, *83*, pp.139-164.

Barigozzi, F., Burani, N. and Raggi, D., 2018. Productivity crowding-out in labor markets with motivated workers. *Journal of Economic Behavior & Organization*, *151*, pp.199-218.

Besley, T. and Ghatak, M., 2018. Prosocial Motivation and Incentives. *Annual Review of Economics*, *10*, pp.411-38.

Bloom, N. and Van Reenen, J., 2011. Human resource management and productivity. In *Handbook of Labor Economics* (Vol. 4, pp. 1697-1767). Elsevier.

Burgess, S. and Ratto, M., 2003. The role of incentives in the public sector: Issues and evidence. *Oxford Review of Economic Policy*, *19*(2), pp.285-300.

Carpenter, J. and Gong, E., 2016. Motivating Agents: How much does the mission matter?. *Journal of Labor Economics*, *34*(1), pp.211-236.

Cassar, L., 2018. Job Mission as a Substitute for Monetary Incentives: Benefits and Limits. *Management Science*.

Chalkley, M. and Malcomson, J.M., 1998. Contracting for health services with unmonitored quality. *The Economic Journal*, *108*(449), pp.1093-1110.

Charness, G., Cobo-Reyes, R. and Sanchez, A., 2016. The effect of charitable giving on workers' performance: Experimental evidence. *Journal of Economic Behavior & Organization*, *131*, pp.61-74.

Copeland, A. and Monnet, C., 2009. The welfare effects of incentive schemes. *The Review of Economic Studies*, *76*(1), pp.93-113.

Dal Bó, E., Finan, F. and Rossi, M.A., 2013. Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, *128*(3), pp.1169-1218.

Delfgaauw, J. and Dur, R., 2007. Signaling and screening of workers' motivation. *Journal of Economic Behavior & Organization*, *62*(4), pp.605-624.

Delfgaauw, J. and Dur, R., 2010. Managerial talent, motivation, and self-selection into public management. *Journal of Public Economics*, *94*(9-10), pp.654-660.

DellaVigna, S. and Pope, D., 2017. What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies*, *85*(2), pp.1029-1069.

Dixit, A., 2002. Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, pp.696-727.

Dohmen, T. and Falk, A., 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, *101*(2), pp.556-90.

Finan, F., Olken, B.A. and Pande, R., 2017. The personnel economics of the developing state. In *Handbook of Economic Field Experiments* (Vol. 2, pp. 467-514). North-Holland.

Francois, P., 2007. Making a difference. *The RAND Journal of Economics*, 38, pp.714-732.

Francois, P. and Vlassopoulos, M., 2008. Pro-social motivation and the delivery of social services. *CESifo Economic Studies*, *54*(1), pp.22-54.

Gneezy, U., Meier, S. and Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, *25*(4), pp.191-210.

Holmstrom, B. and Milgrom, P., 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, *7*, pp.24-52.

Holt, C. and Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, 92: 1644–1655.

Hong, F., Hossain, T., List, J.A. and Tanaka, M., 2018. Testing the theory of multitasking: Evidence from a natural field experiment in Chinese factories. *International Economic Review*, *59*(2), pp.511-536.

Imas, A., 2014. Working for the "warm glow": On the benefits and limits of prosocial incentives. *Journal of Public Economics*, *114*, pp.14-18.

Imberman, S.A., 2015. How effective are financial incentives for teachers?. *IZA World of Labor*.

Johnson, R.M., Reiley, D.H. and Muñoz, J.C., 2015. "The War for the Fare": How Driver Compensation Affects Bus System Performance. *Economic Inquiry*, *53*(3), pp.1401-1419.

Jones, D.B., 2015. The supply and demand of motivated labor: When should we expect to see nonprofit wage gaps?. *Labour Economics*, *32*, pp.1-14.

Lazear, E.P., 1986. Salaries and piece rates. *Journal of Business*, pp.405-431.

Lazear, E.P., 2000. Performance pay and productivity. *American Economic Review*, *90*(5), pp.1346-1361.

Lemieux, T., MacLeod, W.B. and Parent, D., 2009. Performance pay and wage inequality. *The Quarterly Journal of Economics*, *124*(1), pp.1-49.

Neal, D. 2011 The design of performance pay in education. In: Hanushek, E. A., S. J. Machin, and L. Woessmann (eds). *Handbook of the Economics of Education*: Vol. 4. Amsterdam: Elsevier, pp. 495–550.

Paarsch, H.J. and Shearer, B., 2000. Piece rates, fixed wages, and incentive effects: Statistical evidence from payroll records. *International Economic Review*, *41*(1), pp.59-92.

Perry, J. L. (1996). Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory*, *6*(1), 5-22.

Prendergast, C., 2007. The motivation and bias of bureaucrats. *American Economic Review*, *97*(1), pp.180-196.

Scott, A., Sivey, P., Ouakrim, D.A., Willenberg, L., Naccarella, L., Furler, J. and Young, D., 2011. The effect of financial incentives on the quality of health care provided by primary care physicians. Cochrane Database of Systematic Reviews.

Shearer, B., 2004. Piece rates, fixed wages and incentives: Evidence from a field experiment. *The Review of Economic Studies*, *71*(2), pp.513-534.

Tonin, M. and Vlassopoulos, M., 2010. Disentangling the sources of pro-socially motivated effort: A field experiment. *Journal of Public Economics*, *94*(11-12), pp.1086-1092.

Tonin, M. and Vlassopoulos, M., 2015. Corporate Philanthropy and Productivity: Evidence from an Online Real Effort Experiment. *Management Science*, *61*(8), pp.1795-1811.

Appendix A: Additional Tables & Figures

Appendix Table A1: Averages of Phase 3 outcome variables by treatment

TREATMENT:	Flat-rate Motivated (FM)	Flat-rate Non-Motivated (FNM)	Piece-rate Motivated (PM)	Piece-rate Non-Motivated (PNM)
Phase 3: Words spelled	34.90	35.72	42.58	56.71
	(11.09)	(13.44)	(20.00)	(29.27)
Phase 3: Points per word	6.76	6.57	6.27	5.04
	(0.61)	(0.99)	(0.96)	(1.09)
Phase 3: Counted points per word	6.22	5.93	5.43	3.74
	(0.82)	(1.35)	(1.31)	(1.56)
Phase 3: Counted points per counted word	7.35	7.16	7.12	6.38
	(0.56)	(0.77)	(0.68)	(0.52)
Phase 3: Share of words that are counted (>4 pts.)	0.84	0.82	0.75	0.57
	(0.08)	(0.12)	(0.13)	(0.21)
Observations	31	29	31	31

Standard deviations in parentheses

Appendix Table A2: Nonparametric tests of differences in primary Phase 2 outcomes

	(1) Total words (Phase 2)	(2) Points per Word (Phase 2)
Treatment Comparison	*Diff. in means (in X vs. Y, reporting X minus Y)* *[p-value from Wilcoxon test]*	
F-M vs. F-NM	-2.6	0.24
	[0.464]	[0.137]
P-NM vs. F-NM	25.57	-1.04
	[0.001]***	[0.0002]***
P-M vs. F-NM	4.96	-0.15
	[0.534]	[0.652]
P-NM vs. F-M	28.16	-1.29
	[0.0001]***	[0.000]***
P-M vs. F-M	7.55	-0.40
	[0.176]	[0.039]**
P-NM vs. P-M	20.61	-0.89
	[0.007]***	[0.001]***

Each cell reports the difference in means between an outcome (noted in the column header) between two treatments. We conduct a Wilcoxon Rank-Sum Test for each treatment comparison and report the p-value in brackets beneath the differences in means.

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

Appendix Table A3: Testing differences in Phase 2 outcomes across treatments (with some additional controls)

VARIABLES	(1) Total words	(2) Points per word	(3) Counted points per word	(4) Counted points per counted word	(5) Share of counted words (5 pts or greater)
Treatment: F-NM	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
Treatment: F-M	-1.35	0.23	0.32	0.21	0.02
	(2.52)	(0.16)	(0.21)	(0.14)	(0.02)
Treatment: P-NM	23.99***	-0.98***	-1.45***	-0.47***	-0.17***
	(4.59)	(0.22)	(0.30)	(0.15)	(0.03)
Treatment: P-M	8.70**	-0.12	-0.18	-0.01	-0.03
	(3.49)	(0.19)	(0.27)	(0.14)	(0.03)
Observations	122	122	122	122	122
R-squared	0.57	0.36	0.40	0.24	0.44
PM vs. FM p-val.	0.00	0.04	0.03	0.09	0.05
PM vs. PNM p-val.	0.00	0.00	0.00	0.00	0.00

All specifications include the following additional controls: main Phase 1 ability measure, gender, years in college, business major indicator, native English speaker indicator, current work status.

Robust standard errors in parentheses

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

Appendix Table A4: Testing differences in Phase 3 outcomes across treatments

VARIABLES	(1) Total words	(2) Points per word	(3) Counted points per word	(4) Counted points per counted word	(5) Share of counted words (5 pts or greater)
Treatment: F-NM	(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
Treatment: F-M	-0.82	0.19	0.29	0.20	0.02
	(3.19)	(0.21)	(0.29)	(0.14)	(0.03)
Treatment: P-NM	20.99***	-1.52***	-2.19***	-0.51***	-0.25***
	(5.82)	(0.27)	(0.38)	(0.15)	(0.04)
Treatment: P-M	6.86	-0.30	-0.50	-0.09	-0.07**
	(4.37)	(0.25)	(0.34)	(0.13)	(0.03)
Observations	122	122	122	122	122
R-squared	0.17	0.35	0.37	0.20	0.37
PM vs. FM p-val.	0.06	0.02	0.01	0.02	0.00
PM vs. PNM p-val.	0.03	0.00	0.00	0.00	0.00

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

Appendix B: Experimental Instructions

This appendix reproduces the instructions that were provided to participants during our experimental sessions. Sections of the instructions that are specific to the flat-rate treatment are preceded by "F-NM/{M}". Sections of the instructions that are specific to the piece-rate treatment are preceded by "P-NM/{M}". Within those sections, text within curly brackets ("{…}") are specific to mission-oriented treatments. All other sections of the instructions are common across treatments. Instructions prior to "Phase 1" were presented on participants' screens, but also read aloud by the experimenter. All instructions from "Phase 1" onward were presented on participants' screens as they preceded through the experiment at their own pace and were not read aloud.

Introduction

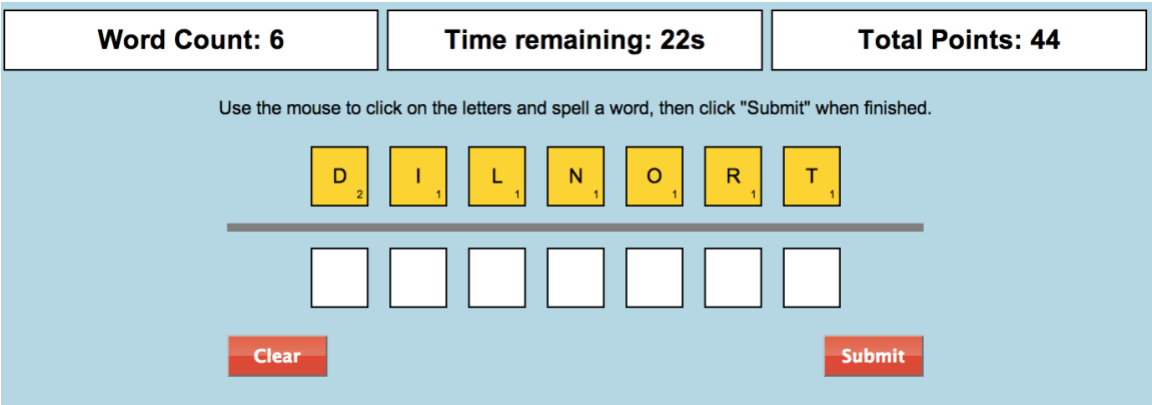
This experiment is a study of decision-making. Your earnings will depend on the actions that you take during the experiment. At the end of the experiment, you will be paid $5 for showing up plus whatever you earn during the course of the experiment. Throughout the experiment, your earnings will be reported in "Experimental Currency Units" or "ECUs". At the end of the experiment, we will convert however many ECUs you have earned into dollars at a rate of X ECU's = $1. Payments will be made privately and in cash. All decisions are made anonymously. Please do not talk to other participants during the experiment. If at any point you have a question, raise your hand and an experimenter will come to you to provide an answer.

The experiment will consist of three task phases and a questionnaire. In the three task phases, you will perform a *word formation task*. At the end of the experiment, we will pay you based on your performance in the first phase and your performance in *either* the second or third phase (plus the $5 show-up fee). We will randomly select whether you (and other participants in the room) are paid for the second or third phase, both of which are equally likely to be selected. We will discuss the specific procedures of each of these phases and how they may impact your earnings as they occur. First, we will describe the word formation task and software interface in detail.

*** screen break ***

The word formation task

During the experiment, you will be asked to complete a number of "word formation tasks." In the word formation task, you will be presented with a set of seven letters as in the figure below.



Your task is to form a word using at least two of the letters available to you. The word must be spelled as it is in the dictionary. For example, if you were provided the set of letters "D I L N O R T", you might spell the word "ON" or the word "LORD". To spell a word, click each letter in order and each letter's tile will move to the lower row. To clear the letters you have entered, click "Clear." Click "Submit" when you are done. If the word you have submitted is not an acceptable word or is incorrectly spelled, your chosen letters will be cleared and you will be asked to try again.

Each letter has a certain number of points associated with it, which is noted in the lower right hand corner of each letter square. When you submit a word, these points will be added up to determine your score for that word. If the total number of points associated with a word is less than five, your score for that word is zero. If the total number of points associated with a word is five or greater, your score for that word is given by the number of points. For example, the word "ON" is made up of two one-letter tiles and as such is worth zero points. The word "LORD" is a total of five points (1+1+1+2), and as such is worth five points.

So that you can practice working with the software, form a word with the seven letters on the next screen, then click "Submit".

*** screen break ***

Phase 1

We will now begin the first phase. This phase consists of five "sections", each of which has slightly different instructions. Follow the instructions on each screen.

*** screen break ***

Phase 1, Section 1

In this first section, you will be asked to complete 5 word formation tasks. Spell a word for each set of letters. You will have 2 minutes to do so. Remember, after spelling a word click "Submit" to move on to the next set of letters.

((task proceeds))

*** screen break ***

Phase 1, Section 2

In this section, the task is again to complete 5 word formation tasks. This time (and in all of the remaining sections of this phase), you have the opportunity to earn ECUs.

In this section, you will start with 20 ECUs. Once you click "Start" the first task will appear. You then have 60 seconds to complete all 5 tasks. You lose 1/3 ECU for each second that it takes you to complete and submit all five tasks. For example, if it takes you 6 seconds to complete *each* of the five tasks, you will have used a *total* of 30 seconds. As a result, your earnings would be 20 ECUs – (1/3)30, or 10 ECUs.

((task proceeds))

*** screen break ***

Phase 1, Section 3

This section is slightly different. You will again have the opportunity to earn money by completing word formation tasks. This time, you will have 3 minutes to complete as many word formation tasks as you can. You will earn 1 ECU for each word formation task you complete.

((task proceeds))

*** screen break ***

Phase 1, Section 4

In this section, you will complete five word formation tasks. You have 25 seconds to complete each one. In this section, the amount of money you earn will depend on the score you accumulate rather than the number of tasks you complete. For each point, you will earn 0.2 ECUs. Remember, words with point totals less than five receive a score of zero points. There is no penalty for taking the full 25 seconds for each task, and no advantage from not doing so, so take the time to spell the best word you can.

((task proceeds))

*** screen break ***

Phase 1, Section 5

In this section, you will have 3 minutes to complete word formation tasks. There is no limit to the number of words you can form during the 3 minute period. This time, your payment again depends on the number of points you accumulate. Every time you spell a word worth five or more points, you will accumulate points from that word. You will be paid 0.2 ECUs for the total number of points you have accumulated at the end of the three minutes. For example, if you spell 9 words that are worth 4 points and 10 words that are worth 8 points, you will be paid for the points accumulated on the 10 words worth at least 5 points. That is, you would receive 8 pts * 10 words * 0.20 ECUs, or 16 ECUs.

((task proceeds))

*** screen break ***

Phase 2

We will now begin the second phase of the experiment. Unlike the previous phase, there is only one section. We are hoping to learn how many points people can accumulate in a fixed period of time.

The rules are as follows:

In this phase, you will have 8 minutes to complete word formation tasks. There is no limit to the number of words you can form during the 8 minute period. Again, you can earn money. {In this phase, your efforts can also benefit a charity of your choosing. You will be allowed to choose a charity in a moment.} You {and a charity} will be paid for the outcome of this phase only if this phase is randomly chosen instead of the third phase. You will not know which phase is being randomly selected until the end of the experiment, so you should proceed as though you may receive payment. There is a 50% chance this Phase will be selected, and a 50% chance Phase 3 will be selected.

If this phase is selected for payment:

F-NM/{M}: [Regardless of how many word formation tasks you complete in the next eight minutes, you will receive 70 ECUs. {Additionally, when you score 5 or more points while spelling a word, you will generate some money for charity. For every point that you score on that word, you will generate 0.2 ECUs for charity. For example, a 5 point word will generate the monetary equivalent of 1 ECU for charity. A 6 point word will generate 1.2 ECUs for charity. A 4 point word will generate 0 ECUs for charity (because a word must be worth at least 5 points to generate money for charity). }

For example, if you complete 40 word formation tasks {and accumulate 150 points on words worth at least 5 points}, you will receive 70 ECUs {and your charity will receive the equivalent of 30 ECUs}. If you complete 100 word formation tasks {and accumulate 250 points on words worth at least 5 points}, you will receive 70 ECUs {and your charity will receive 50 ECUs}.]

P-NM/{M}: For each word formation task that you complete, you will receive 1 ECU. {Additionally, when you score 5 or more points while spelling a word, you will generate some money for charity. For every point that you score on that word, you will generate 0.2 ECUs for charity. For example, a 5 point word will generate the monetary equivalent of 1 ECU for charity. A 6 point word will generate 1.2 ECUs for charity. A 4 point word will generate 0 ECUs for charity (because a word must be worth at least 5 points to generate money for charity). }

For example, if you complete 40 word formation tasks {and accumulate 150 points on words worth at least 5 points}, you will receive 40 ECUs {and your charity will receive the equivalent of 30 ECUs }. If you complete 100 word formation tasks {and accumulate 250 points}, you will receive 100 ECUs {and your charity will receive 50 ECUs}.]

{We will donate money generated for charities after the experiment has ended. So that you can verify that this has happened, we will post receipts from the charities on the experimenter's website. Before we begin this Phase, please indicate which of the following charities you would like to benefit.}

((Participants presented with list of charities with radio buttons.))

We will now ask you a few questions to make sure the instructions are clear. After you have answered the questions, raise your hand and the experimenter will come check your answers.

((Comprehension Check (below) passed out on paper))

1. This phase will end after:
    a. 50 words are spelled
    b. 8 minutes

c. 20 minutes
d. The end is randomly determined.

2. Suppose: At the end of this phase you have spelled a total of 25 words. For the sake of this example, suppose that 10 of the words were worth 2 points apiece and the remaining 15 words were all worth 10 points apiece. If this phase is randomly selected for payment, your earnings from this phase are _____ {and charity receives _____}. (Fill in the blanks.)
3. Suppose: At the end of this phase you have spelled a total of 50 words. For the sake of this example, suppose that 49 of the words were worth 4 points apiece and the remaining word was worth 15 points. If this phase is randomly selected for payment, your earnings from this phase are _____ {and charity receives _____}. (Fill in the blanks.)
4. Which of the following is accurate:
    a. You will be paid for Phase 1 OR Phase 2 OR Phase 3.
    b. You will be paid for Phase 1 AND, either Phase 2 OR Phase 3.
    c. You will be paid for Phase 1 AND Phase 2 AND Phase 3.
    d. You will be paid for Phase 1 OR Phase 2, AND Phase 3.

Phase 3

We will now begin the third phase of the experiment. This is the final phase that will require you to complete word formation tasks. After this phase, you will complete a questionnaire and then the experiment will be over.

This phase is similar to Phase 2. Once again you will have 8 minutes to complete as many word formation tasks as you want. The main difference is that this time you can choose how you are paid. You have two options:

(Previous Option) [Repeat treatment specific payment information.] {The money generated for charity will go to the charity that you chose in the previous phase.}
(New Option) Regardless of how many word formation tasks you complete in the next eight minutes, you will receive 90 ECUs.

Recall that you will receive payment {and benefit charity} as according to your choice only if this phase is the one that is randomly selected for payment at the end of the experiment.

Please respond to two quiz questions to ensure that these instructions are clear. After you have answered the questions, raise your hand and the experimenter will come check your answers.

((Comprehension Check (below) passed out on paper))

1. Suppose you have selected the "Previous Option". At the end of this phase you have spelled a total of 25 words. {For the sake of this example, suppose that 10 of the words were worth 2 points apiece and the remaining 15 words were all worth 10 points apiece.} If this phase is randomly selected for payment, your earnings from this phase are _____ {and charity receives _____}. (Fill in the blanks.)
2. Suppose you have selected "New Option". Again suppose that at the end of this phase you have spelled a total of 25 words. {For the sake of this example, suppose that 10 of the words were worth 2 points apiece and the remaining 15 words were all worth 10 points apiece.} If this phase is randomly selected for payment, your earnings from this phase are _____ {and charity receives _____}. (Fill in the blanks.)

Now, please select the option you prefer. After you have made your selection, this phase will begin.