

Was Weber Right?

The Effects of Pay for Ability and Pay for Performance
on Pro-Social Motivation, Ability and Effort
in the Public Sector

Sheheryar Banuri
Philip Keefer



WORLD BANK GROUP

Development Research Group
Macroeconomics and Growth Team
May 2015

Abstract

This paper examines the effects of pecuniary compensation on the ability and motivation of individuals in organizations with non-pecuniary or pro-social missions. In particular, the paper compares flat pay systems, unrelated with ability or effort, to two other systems that are considered superior: high-powered, pay for performance schemes and more traditional, “Weberian” schemes that calibrate pay to ability, independent of effort. The analysis uses a sample of future public sector workers and finds that all three pay schemes attract motivated workers into

tasks with a pro-social mission. However, flat pay schemes also attract low ability workers. In the short run, pay-for-performance schemes generate higher effort than flat pay and pay-for-ability systems, a difference driven entirely by effects on unmotivated workers. Once selection effects are accounted for, however, workers with pay for ability and pay for performance exert statistically indistinguishable levels of effort in the pro-social task. Moreover, pay for ability elicits effort at lower cost than pay for performance.

This paper is a product of the Macroeconomics and Growth Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at sbanuri@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Was Weber Right? The Effects of Pay for Ability and Pay for Performance on Pro-Social Motivation, Ability and Effort in the Public Sector

Sheheryar Banuri (World Bank)

Philip Keefer (Inter-American Development Bank)

Banuri: Development Economics Research Group, World Bank, 1818 H St NW, MC 3-356, Washington, DC, 20433 (e-mail: sbanuri@gmail.com); Keefer: Inter-American Development Bank, 1300 New York Avenue, N.W., Washington, DC 20577 (e-mail: pkeef@iadb.org).

JEL Codes: C91; H83; J45

Keywords: public sector reform, civil service, intrinsic motivation, extrinsic motivation, performance

Sector Board: Public Sector Governance (PSM)

Introduction

Substantial uncertainty surrounds the effects of pecuniary incentives on worker motivation to undertake pro-social and mission-oriented tasks, particularly in the public sector. Wide variation in public sector compensation practices reflects this uncertainty. Ability-based pay systems characteristic of “Weberian” public administrations have long been considered the ideal compensation system for the public sector (from the mandarins in East Asia to Bismarck). However, performance-based pay is increasingly common (up to half of civil servant pay in Singapore consists of performance bonuses), and flat pay systems, unrelated to either ability or performance, proliferate around the world, particularly in poor countries. We use evidence from laboratory experiments with Indonesian public sector employees to examine the tradeoffs across these pay schemes with respect to the motivation, ability and effort of individuals who choose to work in the public sector.

A central issue in comparing pay schemes in the mission-oriented sector is how they interact with intrinsic and pecuniary motivations to undertake effort. Flat pay systems may be inexpensive to implement, but drive away the most able and motivated. Pay for ability screens out low ability individuals, may attract motivated individuals, but lacks incentives for workers to exert effort, since pay is independent of performance. Pay for performance attracts higher ability individuals and gives them high-powered incentives to perform; however, it may be more costly to implement and discourage pro-social individuals from exerting effort on the job.

Our experiments shed light on these tradeoffs, using a sample of future public sector workers. First, they show that the tradeoffs are sharper with respect to ability than pro-social motivation. Though the pay systems are quite distinct, there are no significant differences among them with respect to the pro-social motivation of those who choose the mission sector. However, pay for performance and pay for ability systems are superior to flat pay systems, and similar to each other, in attracting high ability and deterring low ability individuals from joining the mission sector.

The second key concern is pay system effects on effort. In particular, do high-powered incentives – pay for performance – inspire greater effort in mission-oriented tasks? This question emerges at two key junctures, from the point of view of those seeking to reform the pay system governing a mission-oriented task. On the one hand, reformers are concerned about the effects of pay reforms on the effort of incumbent workers. On the other hand, pay reform has selection effects, as individuals re-sort themselves across the mission and non-mission sectors in response to the pay reform.

Consistent with the intuition of proponents of high-powered compensation in the public sector, pay for performance has a significant positive effect on incumbent effort relative to both flat pay and pay for ability. Moreover, these effects operate on individuals who are less pro-socially motivated. Individuals who exhibit greater pro-social motivation exert similar effort across all pay schemes.

In the “long run”, though, after selection effects have taken hold, effort is indistinguishable across pay for ability and pay for performance systems; both yield higher effort than flat pay systems. That is, after selection effects are taken into account, there is no tradeoff with respect to effort in the choice between pay for performance and pay for ability. However,

pay for ability systems, at least as parameterized in the experiments, elicit greater effort at lower cost.

These results contribute both to scholarly and policy debates concerning the effects of compensation schemes. The behavioral literature provides reason to believe that high-powered incentives in mission-oriented sectors might crowd out workers who are more pro-social. We find that pro-social individuals are no less attracted by pay for performance. At the same time, standard economics literature suggests that high-powered incentive schemes should drive greater effort than all other pay schemes. We find, on the contrary, that after selection effects are accounted for, low-powered pay for ability and high-powered pay for performance systems yield similar effort.

In addition, although prior work has shown significant effort effects of particular pay schemes, it has not been able to make head-to-head comparisons of different schemes, nor to distinguish short run effects (on incumbents, prior to selection) and long run effects (taking selection effects into account). Our experiments allow such head-to-head comparisons. For example, important research has discovered that pay for performance led to improved student test scores in India (Muralidharan and Sundaraman, 2011), relative to a control group that operated under other contractual conditions. These contractual conditions are not easily characterized, creating uncertainty about whether the introduction of pay for performance in other sectors or countries would yield similar improvements. We are able to compare three tightly defined and controlled pay systems, allowing a more precise calibration of the circumstances under which a pay reform, such as the adoption of pay for performance, might be expected to yield significant results, and over what time frame.

Finally, past research (Banuri and Keefer 2013) finds that pro-social individuals are more likely to join the mission-oriented sector. However, their work only considers flat pay systems. The analysis here finds the same result under pay for performance and pay for ability, but further finds that both of these pay schemes are more effective than flat pay at attracting high ability individuals.

These results have several implications for pay reform in the public sector. First, the contentious debate about the application of pay-for-performance to public sector workers may be misplaced: paying high-ability workers high wages, the early standard to which public administrations aspired, is no less effective than pay for performance in enticing motivated, able workers to enter the public sector and persuading them to work hard. The important gains in public sector performance seem to emerge when bureaucracies move away from systems that are unhinged from either ability or performance. Second, however, in the short run, pay for performance, but not pay for ability, can significantly raise performance of incumbent workers who were hired without regard to ability. Third, more tentatively, the results point to the cost-effectiveness of moving to pay-for-ability pay schemes, at least over a long enough period to allow for selection effects to take hold.

The model in the following section provides an analytical framework for the experiments. The sample and experimental structure are then outlined, followed by the results. We discuss the results in the context of the literature in the final section.

Model

We study the effects of different pay systems on the types of workers who select into the public sector and on the effort they exert in those systems. The model in this section identifies several ways in which pay systems, worker characteristics and worker effort can interact. To see the interaction most clearly, assume (as in Banuri and Keefer, 2013), that workers earn income in the private sector based on piece rate compensation, where income is the product of their effort and ability times the private sector wage rate ($w_{piece}a_i e_i$), and ability $a_i \in (0,1)$ determines the fraction of effort that is transformed into output. For simplicity, and to fix ideas, the private sector is assumed to be “large” relative to the public sector, so the wage rate is unaffected by employment in the public sector.

The pro-social motivation of workers is given by θ_i , $\theta_i \geq 0$. Pro-social workers (those with $\theta_i > 0$) gain utility when they exert effort on behalf of society. They may also care about how their effort actually benefits society, where the benefits to society are a function of both their ability and effort. Their utility therefore increases with effort according to $\lambda a_i e_i \theta_i$, where $\lambda \in [1, \frac{1}{a_i}]$; $\lambda = 1$ implies that workers care only about the actual contributions that their effort makes to society and $\lambda = \frac{1}{a_i}$ implies that workers value their efforts on behalf of the public, regardless of the contributions that their efforts actually make. The utility cost of effort increases in effort, and is given by $\frac{1}{2}e^2$.

Workers in the flat wage pro-social organization, where compensation is unrelated to effort and ability, earn a wage given by w_{flat} . Earnings in the pay for ability (“Weberian”) pro-social organization are more difficult to characterize. In principle, a pure ability-contingent wage could be given by $w_{Weber}a_i$. In practice, however, tests of ability demand effort by the worker, either at the time of the test or in preparation for it. Hence, the ability of workers is determined by the jointly observed product of ability and effort on the test, $a_i e_{i,test}$. Compensation in the pay for ability pro-social organization is given by $w_{Weber}a_i e_{i,test}$. Finally, those in the performance-based pro-social organization earn an output-contingent wage, $w_{pfp}a_i e_i$, recalling that worker output is a function of both effort and ability.

Assuming that extrinsic and intrinsic motivation and effort enter utility separably, the utility of workers who choose to work in the private or public sectors can then be described by:

Private sector: $u_i = w_{piece}a_i e_i - \frac{1}{2}e^2$

Public sector, Flat: $u_i = w_{flat} + \lambda a_i e_i \theta_i - \frac{1}{2}e^2$

Public sector, Weber: $u_i = w_{Weber}e_{i,test} + \lambda a_i e_i \theta_i - \frac{1}{2}e^2$

Public sector, Pay for Performance: $u_i = w_{pfp}a_i e_i + \lambda a_i e_i \theta_i - \frac{1}{2}e^2$

By assumption, workers undertake effort for only two reasons: the intrinsic motivation that comes from exerting effort on a pro-social task; and the pecuniary motivation of exerting effort to increase private earnings. Workers who select into any of the four options choose effort to maximize their utility from that option. In the non-mission organization, workers choose

effort to maximize $w_{piece}a_i e_i - \frac{1}{2}e^2$, giving $e_{piece} = w_{piece}a_i$. In the flat-pay pro-social organization, they maximize $w_{flat} + \lambda a_i e_i \theta_i - \frac{1}{2}e^2$, or $e_{flat} = \lambda a_i \theta_i$. In the pay-for-performance pro-social organization, maximization of utility over effort yields $e_{pfp} = \lambda a_i \theta_i + w_{pfp} a_i$.

The Weberian pro-social organization is slightly more complicated, since workers first exert effort in the test of their ability and then again, once they are on the job. In the experiments below, all subjects take the test of ability, $e_{i,test}$, knowing that it could affect their subsequent earnings. However, they have no precise information about the relationship between test performance and compensation, nor of the nature of the tasks that they will be asked to undertake. The results of the ability test are therefore a function of ability and the value that individuals place on future earnings, but are independent of the precise characteristics of the pay systems they will eventually confront.

Those who subsequently select into the Weberian public sector therefore choose task effort to maximize $u_i = w_{Weber} \bar{e}_{i,test} + \lambda a_i e_i \theta_i - \frac{1}{2}e^2$, yielding $e_{Weber} = \lambda a_i \theta_i$. Effort under the pay for ability pay scheme is therefore the same as in the flat pay scheme for all subjects. However, if – as is the case in our experiments – performance on the ability test is positively correlated with actual ability, more of the high ability individuals select into the Weberian mission sector than into the flat pay sector, yielding higher effort among the Weberian mission workers.

Table 1: Sector and pay system effects on worker utility, ability and mission-orientation

	Utility at optimal effort	$\frac{\partial u}{\partial a}$ – Ability	$\frac{\partial u}{\partial \theta}$ – Pro-social motivation
Piece rate, non-pro-social task	$\frac{1}{2}(w_{piece}a_i)^2$	$a_i w_{piece}^2$	0
Flat pay, pro-social task	$w_{flat} + \frac{1}{2}(\lambda a_i \theta_i)^2$	$a_i (\lambda \theta_i)^2$	$\theta_i (\lambda a_i)^2$
Weberian, pro-social task	$w_{Weber} a_i \bar{e}_{i,test} + \frac{1}{2}(\lambda a_i \theta_i)^2$	$w_{Weber} a_i \frac{\partial \bar{e}_{i,test}}{\partial a} + a_i (\lambda \theta_i)^2$	$w_{Weber} a_i \frac{\partial \bar{e}_{i,test}}{\partial \theta} + \theta_i (\lambda a_i)^2$
Pay for performance, pro-social task	$\frac{1}{2}(a_i)^2 (\lambda \theta_i + w_{pfp})^2$	$a_i (\lambda \theta_i + w_{pfp})^2$	$\lambda (a_i)^2 (\lambda \theta_i + w_{pfp})$

Table 1 summarizes the relative attractiveness of work in the mission sector under the three pay systems and work in the piece rate, non-mission sector. Column one displays utility when subjects exert optimal effort in the respective task and pay system. Columns two and three indicate how utility – the attractiveness of each sector-pay combination – changes with the ability and pro-social motivation of workers.

Comparing the last three cells of the third column of Table 1 to the first cell immediately reveals that the more pro-social are workers, the more likely they are to prefer the pro-social task, regardless of pay system. This is intuitive: the piece-rate task has no mission and the

utility from the piece-rate task does not change as workers become more pro-social. The question, then, is which of the three mission-oriented pay systems are most attractive to pro-social individuals.

The additional utility that more mission-oriented individuals receive from the flat pay mission task is $\theta_i(\lambda a_i)^2$. This is unambiguously less than the extra utility that they receive from the pay for performance task, $\lambda(a_i)^2(\lambda\theta_i + w_{pfp})$. Hence, pay for performance should unambiguously attract more pro-social individuals than flat pay systems. However, it is ambiguous whether more pro-social individuals derive greater utility from the pro-social task under the Weberian versus the flat pay systems, since performance on the ability test need not vary systematically with pro-sociality (the sign of $\frac{\partial \bar{e}_{i,test}}{\partial \theta}$ is ambiguous). For $\frac{\partial \bar{e}_{i,test}}{\partial \theta} = 0$, for example, pay for ability and flat pay systems are equally attractive to pro-social individuals. For the same reason, it is also ambiguous whether the pay-for-performance or Weberian tasks are more attractive to pro-social individuals.

As workers become more able, their utility from all tasks and pay schemes increases, but the increases differ across pay schemes and tasks. The question is therefore how, as ability increases, preferences for the pro-social task change under the different pro-social pay schemes.

The comparative statics in column two of Table 1 make clear that the attractiveness of the pro-social task increases faster with ability under pay for ability and pay for performance than under flat pay. Considering the flat pay regime, more able workers prefer the pro-social task when they are sufficiently pro-social and the piece rate in the non-mission sector sufficiently low, such that $\lambda\theta_i > w_{piece}$. The attractiveness of the Weberian pay scheme rises faster with ability, however, as long as $\frac{\partial \bar{e}_{i,test}}{\partial a} > 0$ – as long as higher ability individuals perform better on the ability test, increasing their potential earnings in the pro-social task. This is the case in our experiments. Finally, both the pecuniary and non-pecuniary payoffs in pay for performance systems increase with ability.

The experimental results are consistent with these predictions. First, under all pay schemes, the pro-social task attracts more pro-social individuals; the magnitude of the effect, however, is greatest under pay for performance, where greater pro-social effort also receives a pecuniary reward. Second, the flat pay pro-social task is significantly less attractive to high ability individuals and more attractive to low ability individuals: individuals who select the flat pay pro-social task, but not those who select the pro-social task under the other two pay schemes, are significantly less able than those who select the piece rate task. Third, among all subjects, effort is no different in the flat pay and pay for ability regimes, and significantly greater under pay for performance. However, among subjects who choose to work in the mission sector under the respective pay schemes, effort under pay for ability and pay for performance is higher than effort under the flat pay system.

Experimental Design and Subject Pool

The model yields a number of hypotheses that we examine in the experiments, related to the effects of pay systems on the types of individuals who select into mission sectors and on effort that individuals undertake on mission tasks. Both Weberian, ability-based pay and

performance-based pay systems should attract higher ability individuals into mission work than flat pay systems. In addition, because highly-motivated, high ability individuals should prefer the mission sector under the pay-for-ability and pay-for-performance regimes, the ability of individuals who select the mission sector under these pay schemes will be higher than in the piece-rate non-mission sector. All pay systems should attract more pro-social individuals, but pay-for-performance, which provides both pecuniary and non-pecuniary rewards for greater effort on the mission task, should attract more mission-oriented individuals than the pay-for-ability and flat pay schemes. Finally, among the population at large, flat pay and pay-for-ability should elicit similar effort and pay-for-performance more effort. However, among those individuals who select into the mission sector, pay-for-performance and pay-for-ability should both yield greater effort than among those who select into the flat pay mission sector.

The research questions require measurement of both pro-social motivation and effort. The purpose of the paper is to analyze the motivation of workers to undertake a particular type of mission task, one associated with the “public sector” and therefore benefitting a large group of anonymous individuals, spread across the entire country. We therefore sought a mission that would have the same characteristics. In Indonesia, one organization fits this description: the Indonesian Red Cross Society, a general, nation-wide charity that assists with disaster-relief, ambulance services, climate change, disaster preparedness, water, sanitation, HIV/AIDS, Avian FLU and blood donation, among other activities.¹ All pro-social tasks engaged in during the experiment generated benefits for the Indonesian Red Cross Society.

Theory emphasizes that worker motivation depends on the degree of the match between the mission orientation of the worker and the actual mission of the organization. We therefore adopt a measure of motivation that exactly matches the mission in the effort measures using a version of the dictator “game.”² Subjects were asked to donate as much as they liked out of an endowment of 2000 tokens (equal to 16,666 IDR or \$1.78) to the Indonesian Red Cross.³ In companion work (Banuri and Keefer, 2013), we show that greater motivation (as measured by donations to the Indonesian Red Cross) significantly predicts the effort that subjects exert on behalf of the IRC.

To measure effort, we utilize the “slider task” adapted from Gill and Prowse (2011). Subjects are shown 48 sliders on a computer screen. Each slider is set on the left, and the task for subjects is to move the slider to the center of the slider bar. The task demands real effort, but is sufficiently dull so as to minimize intrinsic motivation to engage in the task itself. In each round, subjects are given two minutes to try and complete as many sliders as they can (it is extremely rare for any subject to complete all 48 sliders in two minutes). The number of sliders

¹ Previous research has also used charitable organizations in dictator games. Eckel and Grossman (1996) find, for example, that subjects give substantially more when the anonymous recipient is replaced with a charity (in their case, the American Red Cross). See also Carpenter et al. (2008) and Li et al. (2011).

² A large literature in behavioral economics uses the dictator game as its core measure of altruism and pro-sociality (Forsythe et al 1994; Eckel and Grossman, 1996; Whitt and Wilson, 2007; among many others). Previous research has also replaced the recipient of the dictator game from a student to a charitable organization (Eckel and Grossman, 1996; Li et al, 2010; Carpenter et al. 2008, among others).

³ Income per capita in Indonesia is approximately \$3,000; this amount is approximately 20 percent of daily income per capita. The average cost of lunch at the local cafeteria was approximately 15,000 IDR, so we can be confident that the stakes were not trivial for the subjects.

completed in two minutes is the measure of effort. The use of this computerized version of the “envelope-folding effort task” to simulate effort costs is increasingly common in the literature (Breuer, 2013; Georganas et al. 2013; Ibanez and Schaffland, 2013; among others). The one disadvantage of this task is that the dispersion of subject effort in the task is tightly distributed; this works against finding significant treatment effects, however.

While the task is simple, subject performance could still exhibit heterogeneous learning effects that would inject noise into our estimates. To minimize learning noise, subjects completed four rounds of practice with the slider task (called the “Practice Block”).

The final practice round was used to assess subject ability. Subjects were informed that they would engage in another round of the slider task, but their score would be recorded and reported back to them, and their performance could influence their earnings at a later stage. However, they were not compensated for this task. Thus, this task mimics an entrance exam into the civil service. This round was conducted in all treatments reported in the paper. Effort in this round was used to slot subjects into one of three pay grades in the Weberian, pay-for-ability pro-social organization.

In the next round (immediately following the practice block) we informed subjects that they would now be using the slider task to raise money for charity. In this round (referred to here and below as the “Effort for Charity” round)⁴, subjects were informed that for each slider they successfully completed, 100 tokens would be (and actually were) donated to the Indonesian Red Cross. The subjects themselves did not earn anything during this round. We use the results from this round to estimate whether those who give more to charity in the dictator game also exert more effort on behalf of the charity (as reported in Banuri and Keefer, 2013).

When subjects completed the effort-for-charity round, they were told that they would engage in four tasks in the remainder of the experiment. Subjects were told that one of the four tasks would be chosen at random at the end of the experiment and that they would be paid according to the results of the chosen task (in addition to the payouts associated with the dictator and charity tasks described above). This was done to promote independence in decisions across the tasks.

The first task faced by the subjects (Task 1 in Figure 1) was a “pay-for-effort” or “piece rate” task in which subjects were asked to complete the slider task under a piece rate pay scheme designed to mimic the private sector. Each slider earned subjects 100 tokens for themselves. They engaged in this task for three rounds of two minutes each. If this task was chosen for payment, the sum total of all sliders completed in the three rounds (times 100 tokens) was paid to the subject.

The piece-rate task has two features that are important. First, a significant difficulty in research on the motivation of public officials is to establish the reservation wages of officials. As the model underscores, however, the reservation wage is a key parameter in establishing the effects of pay systems. Task 1 precisely establishes this for all subjects.

Second, we necessarily compare effort under different pro-social organization pay schemes by measuring subjects’ performance in the task, but their performance is a product of both their motivation and ability. As in most such comparisons in the literature, therefore, we

⁴ To avoid priming the subjects, we referred to this round as part of the practice rounds in the instructions.

control for ability in order to isolate the effect of motivation on differences in performance. IQ and previous salary history are two common measures of ability in the literature. These, however, may correspond more or less closely to the actual abilities that the task calls for. The most precise measure of ability is one that directly reflects the task. In addition, as in Prior and Lupia (2008), the least noisy measures of ability are those for which there are stakes attached to demonstrating greater ability. The piece rate task offers more immediate and concrete rewards than the ability-grading task (since there are no direct incentives, making it a noisy measure of ability), and so we use the piece rate task.

Once the pay-for-effort task was completed, all subjects engaged in Task 2 in Figure 1, a pro-social task in which additional effort yielded greater donations for the Indonesian Red Cross. The main focus of the experiments is to assess the selection and effort effects of the three different pay schemes. The assessments are based on a between-subjects design in which subjects are randomly assigned to one of the three mission sector pay treatments in Task 2. That is, in Task 2, subjects received compensation according to one of the three pay schemes for which they were randomly selected (flat pay, pay for ability, or pay for performance). Under the flat pay scheme, subjects were paid a flat salary of 6,600 tokens (i.e. 2,200 tokens per round for a total of three rounds).

Under the pay for ability (Weberian) scheme, subjects were paid a salary that depended on their performance in the final round of the practice block (the ability-grading round).⁵ We implemented two ability thresholds. If their performance was below the lower threshold (as defined as 23 sliders out of 48 sliders: 48% completion), they were in the low ability grade and entitled to a total flat pay level of 4,200 tokens (i.e. 1,400 tokens per round for a total of three rounds). If their performance was equal to or above the lower threshold but lower than the higher threshold (defined as 28 out of 48 sliders: 58% completion), then they were in the medium ability grade and entitled to a total flat pay level of 6,600 tokens (i.e. 2,200 tokens per round for a total of three rounds). Finally, if their performance was equal to or above the higher threshold, they were in the high ability grade and entitled to a total flat pay level of 8,100 tokens (i.e. 2,700 tokens per round for a total of three rounds).

Salary levels under pay for ability were chosen to ensure that subjects would earn considerably more in the pay-for-effort task (piece rate: 100 tokens per slider). Thus, even if subjects maintained the level of effort exerted in the ability-grading round (most subjects' effort level improved under piece rate), they would still earn more under piece rate. For example, if a subject exerted effort in the piece rate at the lower threshold (23 sliders), he would be paid 2,300 tokens per round for three rounds (for a total of 6,900), which is higher than the flat salary of the medium ability worker (6,600).

⁵ The piece rate task was not used to slot subjects to ability grades because, although it is the most precise measure of ability, it was designed to mimic a private sector task, with incentives to exert effort in the current task, independently of other tasks to which the subjects could be exposed. If we had added language such that effort in this task would influence earnings in subsequent tasks, effort in this task would no longer have been independent and no longer a pure measure of individual effort under piece rate systems. Instead, therefore, we used an uncompensated earlier round to slot subjects to ability grades so that we could inform them that we would do so. We also did not incentivize this round so as to keep the motivation based on future (and not current) earnings, much like entrance exams into the civil service.

Finally, under the pay-for-performance scheme, subjects were instructed that their pay would depend on their performance in the following way: subjects were provided with performance targets, using the same thresholds and pay levels as in the pay for ability scheme. In each round during the task, if subjects were able to manipulate less than 23 sliders, they would earn a flat salary of 1,400 tokens for that round. If subjects successfully manipulated between 23 and 27 slides (inclusive), they would earn a flat salary of 2,200, and if they manipulated 28 or more sliders, they would earn 2,700 tokens for the round. Thus in each round, subjects had an incentive to manipulate as many sliders as they could, up to the threshold. As in the pay for ability scheme, higher ability individuals could always earn more in the piece rate task, consistent with relative pay between the private and public sectors in most countries.

Once subjects completed the pay for effort and pro-social tasks, they were then asked to choose between the two pay schemes (piece rate or mission). Subjects knew that there were at least two additional tasks remaining in the session, and were informed that their choice of pay would impact the final two tasks (therefore, the choice between pay scheme in task 1 and pay scheme in task 2 was incentivized and meaningful). Subjects knew that at least task 3 would be identical to either task 1 or 2 (depending on their choice). Furthermore, subjects were provided with all information relevant to this decision, including their performance in each of the two preceding tasks, as well as the amounts generated for themselves and for the Indonesian Red Cross. They were also informed that only one of the four tasks would be selected for payment. Task 4 (not analyzed for this paper) added an additional shock to the pro-social task by raising or lowering the wage level. No substantial differences in effort were found as a result of these shocks (possibly due to data limitations, as each shock further reduced the number of subjects per treatment).

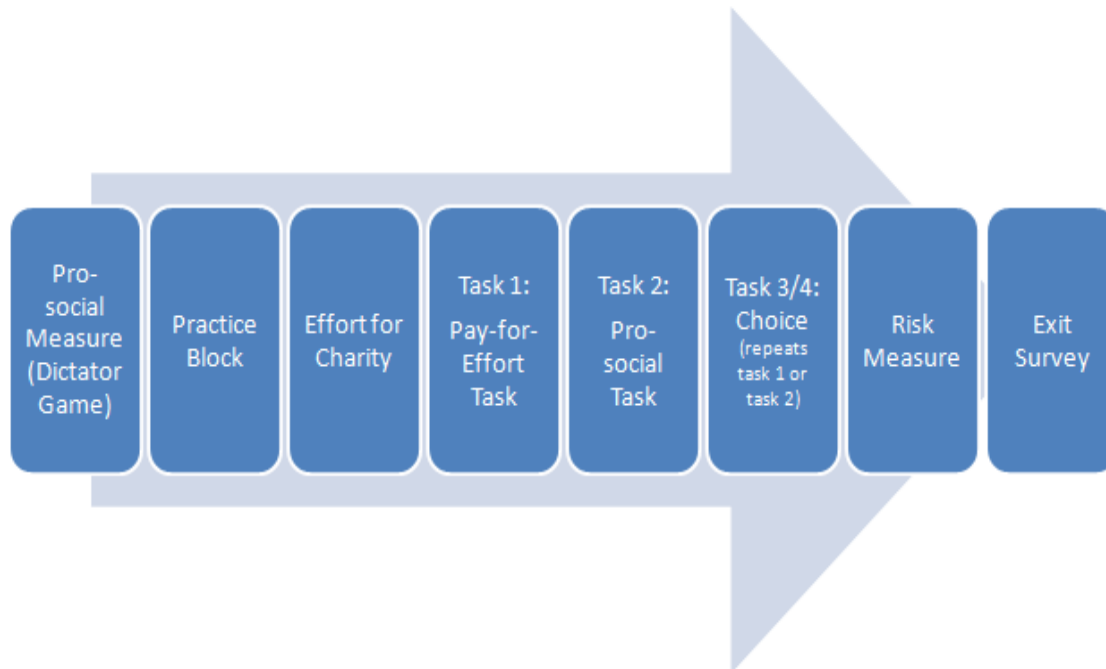


Figure 1: Structure of the Experiment

After completing all of the games involving sliders, subjects engaged in two additional tasks. One was a risk measure (using the Eckel and Grossman elicitation method, Eckel and Grossman, 2002). The other was an extensive survey recording subject demographics. Once these were completed, subjects were asked for a volunteer to oversee the payment to the charity. This volunteer would roll a four-sided die to determine the task that would be paid out at the end of the session, and would verify payment by accompanying an experimenter to the closest bank and make the cash donation directly to the bank account of the charity.

Students received 25,000 IDR as their show-up fee. Average earnings from the experiment were around 120,000 IDR. All sessions were conducted during March 2012 and March 2013 and took about 2 weeks to complete. In addition, since subjects used a mouse to manipulate the sliders, care was taken to utilize identical mice and screens during both rounds of data collection and to use the same screen resolution on the computers to minimize differences. Since this was an individual task, multiple treatments took place within the same session. Subjects were randomly assigned to seats within the computer lab and subjects in adjoining seats were given alternating treatments. Experimental sessions were conducted in March 2012 and March/April 2013, with a total of 431 subjects.⁶

All earnings were expressed in tokens, with an exchange rate of 8.33 IDR per token, and all subjects were paid in cash at the end of each session. At the end of a session, experimentalists asked for a volunteer from the session, who stayed behind to verify payment to the charity. Once all subjects were paid, the volunteer added up the total donation to the charity from the session, and filled out a cash deposit slip. Deposits were made in the presence of the volunteer once per day in cash at the closest bank location. All subjects were informed of this procedure in the instructions at the beginning of the experiment.⁷

Students from a prominent and highly competitive college in Jakarta, Indonesia, the State College of Accountancy (Sekolah Tinggi Akuntansi Negara, or STAN), participated in the experiments. Study at STAN is tuition-free in exchange for a commitment to join either the Ministry of Finance or to assume an accounting role at one of the other ministries, should a position be offered to the students. Nearly all do so. Students who are offered a position in government and turn it down are required to repay their tuition.

STAN students therefore constitute a sample of public sector officials who have not been socialized by work in the public sector. They are not representative of all public officials: the competitiveness of STAN and the career tracks of STAN graduates place them in the upper echelons of government employment. They are, furthermore, dissimilar from the public

⁶ Sessions were conducted in one month over two years. The overall research project contained a number of treatments using the same sample and always began with the dictator game. We pool the data together for the purposes of this paper so as to maximize the number of subjects per treatment. To check for systematic differences between the samples across the two years, we test for significant differences in survey responses across the two years (since the survey questions were nearly identical). With the exception of age (subjects in the second wave were 4-5 months older), no significant differences are found in the data.

⁷ Payments were carried out once a day in the presence of a volunteer from the session. In the case of multiple sessions in a day, participants were informed when the donation was to take place, and were invited to come verify payments at that time.

employees who feature in other research, who typically work in the “caring” professions (health or education) rather than the technical positions to which STAN students are slotted.

Subjects were recruited using a combination of an information session, postings on the school website and on social media. These communicated three messages: that the students would play games, have the opportunity to earn money, and would earn money simply for showing up. They did not reveal the nature or purpose of the experiments. Students were encouraged to sign up for pre-specified sessions. No effort was made to limit participation. Student assistants were hired to assist in the conduct of the experiments. In the overarching study, 1,073 students at STAN participated in the experiments.

Results

We examine results that bear on the two sets of hypotheses, those relating to the selection effects of the three pay systems, and those relating to the effort effects. These are reviewed in turn in this section. The section concludes with a discussion of the efficiency of each pay scheme in terms of generating effort.

Pay system effects on selection into the mission sector

After they complete Task 2, subjects are told that they will engage either in the piece rate task (non-mission) or the mission task under the particular pay scheme given in Task 2. We analyze subjects’ decision to choose the mission sector and its associated contract (pay system) over the non-mission sector and its (more lucrative) piece rate contract, estimating the following logit equation:

$$Public\ Organization_i = \alpha + \beta Motivation_i + \gamma Ability_i + \delta CONTROLS_i + \varepsilon_i$$

The *Public Organization* variable equals one if subject i chooses the pro-social, mission task (where effort increases donations to the charity) and zero if they choose the piece rate task (which does not benefit the charity). To assess whether pro-social individuals are attracted to the pro-social task, we use the variable *Motivation*, the amount subject i donated to the Indonesian Red Cross in the dictator game. The variable *Ability* controls simultaneously for the ability of the subjects and their reservation wage: it is the effort exerted in task 1 (the earlier piece-rate task).

Since subjects were randomly assigned to the different pay schemes, the appropriate test of the effects of motivation and ability on the decision to join the mission sector is without controls. We also report estimates with an exhaustive set of controls, gender, age, wealth, risk preferences (using an incentivized risk measure from Eckel and Grossman, 2005), religiosity (frequency of attendance at religious services), subject belief that the charity was paid in accordance with the instructions, and subjects’ own private information about the charity (measured using a 7 point Likert scale asking subjects to rate the effectiveness of the Indonesian Red Cross). Model 1 performs the regression for the flat pay treatment, model 2 for pay for ability, and model 3 for pay for performance.

Table 2 shows that, consistent with the earlier arguments, under all pay schemes, pro-social individuals are attracted to the pro-social task ($p < 0.05$). Under the flat pay scheme, a 10% increase in motivation of the subject increases the probability of choosing the pro-social contract

by 4.9% ($p < 0.01$). This impact is nearly identical for the pay for ability (6.5%; $p < 0.01$). The coefficient estimate is largest for pay for performance contracts (6.8%; $p < 0.05$). This is consistent with the fact that pay for performance contracts yield larger payoffs (pecuniary and non-pecuniary) for pro-social effort. Moreover, the difference in coefficients between the flat pay and pay for performance regressions is significant.

All pro-social pay systems do not, however, attract more able individuals. Under flat pay, *low* ability individuals are significantly more likely to choose the pro-social task ($p < 0.01$). A 1-slider decrease in the ability of the subject is associated with a 1.0% increase in the probability of choosing a pro-social contract. However, under both the pay for ability and pay for performance treatments, ability is not significantly related to task choice ($p = 0.66$ and 0.25 respectively), and the two coefficients are not significantly different from each other.

This result is key: average ability in the mission sector, whether under the low-powered, pay for ability or high-powered pay for performance compensation systems, is no different than in the non-mission sector, even though pecuniary compensation in the non-mission sector is always higher. We verify this result by pooling the pay-for-ability and pay-for-performance treatments together and adding interaction terms that multiply a dummy for the pay system with ability and motivation variables (not shown). We find no significant differences by treatment. Motivated subjects are equally likely to join the mission sector under pay-for-ability and pay-for-performance ($p = 0.79$). Similarly, high ability subjects are equally likely to join the mission sector under either pay scheme ($p = 0.86$). Thus, both pay schemes attract a similar profile of workers, though high ability workers are more likely to join either pay scheme relative to flat pay.

Table 2: Who chooses the mission sector under different pay schemes?

Dependent Variable: Sector Choice (1 = Join Mission Sector)			
Treatment	Flat	Pay-for-Ability	Pay-for-Performance
	I	II	III
Amount Sent in Dictator Game	0.987*** (0.35)	1.315*** (0.31)	1.498** (0.59)
Ability (Effort Exerted in Piece Rate)	-0.039*** (0.01)	0.011 (0.01)	0.006 (0.02)
Constant	2.624** (1.13)	-1.770* (0.95)	-0.964 1.70
Log Likelihood	-111.3	-123.7	-35.3
Pseudo R-squared	0.0854	0.0799	0.105
P-value	0.000	0.000	0.016
Observations	176	195	60

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Logit specification, standard errors in parentheses. The dictator game variable is divided by 1000 for presentational convenience. Please see table A1 for the same table but with a full set of controls.

Pay system effects on effort in the mission sector

Effort in the mission sector can be measured at two junctures. The first is pre-selection, when all subjects are assigned to a particular mission pay system (Task 2). This group matters because many pay reforms occur in the presence of incumbent workers whose recruitment occurred under quite different pay systems. However, in the long run, the effort effects of any particular pay scheme are a product of the efforts of individuals who actually choose to work in the mission sector under that pay scheme. We are also able, though, to assess effort among subjects at a second juncture, looking at those who have *chosen* to work under the mission pay system, after having been randomly assigned to that pay system's treatment group.

Reviewing the earlier expressions for optimal effort, in the non-pro-social (private) organization, workers exert effort $e_{piece} = w_{piece}a_i$. In the flat-pay pro-social organization, they maximize $w_{flat} + \lambda a_i e_i \theta_i - \frac{1}{2}e^2$, or $e_{flat} = \lambda a_i \theta_i$. Similarly, optimal effort under the Weberian pay scheme is given by $e_{Weber} = \lambda a_i \theta_i$. In the pay-for-performance pro-social organization, maximization of utility over effort yields $e_{pfp} = \lambda a_i \theta_i + w_{pfp}a_i$. Among all subjects, therefore, prior to selection, we expect subjects to exert higher effort under pay for performance than under the flat pay and pay for ability, due to the variable wage component of effort: $w_{pfp}a_i$. Moreover, since greater motivation affects optimal effort in the same way across the three pay systems in the pro-social task (in all three, effort is a function of $\lambda a_i \theta_i$), the additional effects of pay for performance on effort should emerge because of its influence on unmotivated individuals, which we find. However, we expect no differences between pay for ability and flat pay systems.

We test these predictions using effort exerted in task 2 (the pro-social task, where subjects are paid according to treatment, and effort generated earnings for the Indonesian Red Cross). All subjects in our sample participated in this task, with 176 subjects in the flat pay treatment, 195 subjects in the pay-for-ability treatment, and 60 subjects in pay-for-performance.⁸

Specifically, we pool all subjects and then regress dummy variables for the treatments (pay for ability and pay for performance) to which subjects were randomly assigned in task 2 on their effort in that task. Since subjects vary in their ability to do the task, injecting noise into our estimates of the effects of pay schemes on effort, we focus on the effort that they exert in the mission task relative to (divided by) their effort in the piece rate, non-mission task (task 1). For example, instead of directly examining effort in the flat pay scheme, given by $e_{flat} = \lambda a_i \theta_i$, we look at $\frac{\lambda a_i \theta_i}{w_{piece}a_i} = \frac{\lambda \theta_i}{w_{piece}}$, an ability-adjusted measure of effort.

As before, our preferred specification has no controls, since the subjects were randomly assigned to the different pay schemes. Nevertheless, we also present results using the same controls as before, preference for the Indonesian Red Cross (as measured by the amount donated in the dictator game), gender, age, religiosity (frequency of attendance at religious services),

⁸ One subject (in the flat treatment) exerted no effort in the practice and piece rate tasks (ability measure) yielding their ability-adjusted effort to be undefined. Hence this subject is dropped for the subsequent analyses.

wealth, area of study, and the subject's belief that the charity was paid in accordance with the instructions.⁹

Table 3: Treatment effects on effort in pro-social organizations, pre-selection (all subjects)

Dependent Variable: Effort for Charity Relative to Ability			
	Ability- adjusted Effort I	Ability- adjusted Effort II	Ability- adjusted Effort III
Pay-for-Ability (D)	0.021 (0.03)	0.020 (0.03)	0.021 (0.03)
Pay-for-Performance (D)	0.086** (0.04)	0.087** (0.04)	0.093** (0.04)
Amount Sent in Dictator Game		-0.015 (0.02)	-0.013 (0.02)
Gender (D) (1 = Female)		0.016 (0.03)	0.015 (0.03)
Age (in years)		-0.002 (0.01)	0.000 (0.01)
Religious Attendance (5 = More than Once a Week)			0.004 (0.01)
Family Income (Relative to Others) (5 = Much Above Average)			0.027* (0.02)
Accounting Major (D)			0.017 (0.03)
Tax Major (D)			0.023 (0.03)
Belief that Charity was Paid (5 = Complete Confidence)			-0.023 (0.02)
Constant	0.999*** (0.02)	1.048*** (0.21)	0.991*** (0.24)
R-squared	0.013	0.015	0.029
P-Value	0.061	0.268	0.257
Observations	430	430	430

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable is ability-adjusted effort in the Pro-social task (task 2). OLS specification, standard errors in parentheses. The dictator game variable is divided by 1000 for presentational convenience.

Table 3 presents the results of the analysis. The first two rows, displaying the coefficients of the pay for ability and pay for performance dummy variables, are the results of greatest

⁹ Wealth is measured by responses to the question "Relative to other students at your institution, would you say your family income is:" where 1 = "Much below average" and 5 = "Much above average"

interest. These coefficients give the effort under the indicated pay scheme relative to effort under the flat pay scheme, recalling that “effort” is ability-adjusted: effort in Task 2 (the mission or pro-social task) divided by effort in Task 1 (the piece-rate, non-mission task).

Consistent with the theory, among all subjects and prior to selection, there is no significant difference in ability-adjusted effort between the flat pay and pay for ability contracts ($p=0.42$). However, again consistent with the theory, ability-adjusted effort is significantly greater in the pay for performance treatment ($p<0.05$): ability-adjusted effort in the pro-social organization increases by 10% compared to the flat pay scheme. Effort exerted under pay for performance is also significantly higher than effort exerted under the pay for ability treatment ($p<0.1$). These differences are robust to controlling for social preferences, age, gender, and income.

One aspect of interest on the effects of pay systems across all subjects concerns crowding out. The literature identifies the possibility that high-powered incentives could lead to lower effort among motivated subjects than they would otherwise exert. A corollary indication of crowding out is simply that motivated subjects are less susceptible to high powered incentive schemes in mission tasks than are less motivated subjects. Our experiments provide evidence of this latter effect: motivated subjects exhibit no differences in effort across pay schemes. Unmotivated subjects, in contrast, exert significantly more effort in the high-powered pay for performance pay scheme than they do in the low-powered pay for ability and flat pay schemes.

To see this, we first identify subjects with high and low motivation by splitting the sample according to whether subjects gave more or less than the median of all donations in the dictator game (median donations for this sample amount to 25 percent of the endowment). Effort by the motivated group of subjects (those who gave more than 25 percent of their endowment) turns out to be statistically indistinguishable across pay schemes, including the flat pay scheme. Their effort is not significantly different between the flat and pay for ability treatments ($p=0.44$), nor between flat and pay for performance ($p=0.33$), nor between pay for ability and pay for performance ($p=0.71$). The unmotivated subjects (defined as those that donated, at most, 25 percent on their endowment) also exert similar effort in the flat and pay for ability treatments ($p=0.60$). However, unmotivated subjects in the pay for performance treatment exert significantly more effort than either the flat pay treatment ($p<0.05$) or the pay for ability treatment ($p<0.05$).¹⁰ Table 4 presents these results.

¹⁰ These results are robust to using the mean of donations, rather than the median.

Table 4: Treatment effects on effort in pro-social organizations, pre-selection (Unmotivated vs. motivated subjects)

Dependent Variable: Effort for Charity Relative to Ability		
	Unmotivated subjects (Dictator ≤ 500)	Motivated subjects (Dictator > 500)
	I	II
Pay-for-Ability (D)	0.027 (0.04)	0.016 (0.03)
Pay-for-Performance (D)	0.153** (0.06)	0.022 (0.04)
Constant	0.625 (0.39)	1.234*** (0.24)
Controls	Yes	Yes
R-squared	0.057	0.062
P-Value	0.136	0.404
Observations	259	171

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable is ability-adjusted effort in the Pro-social task (task 2). OLS specification, standard errors in parentheses. The dictator game variable is divided by 1000 for presentational convenience. Please see table A2 for the same table but with a full set of controls.

Among all subjects (and, importantly, unmotivated ones), pay for performance elicits greater effort than all other pay systems. However, in the long run, pay systems affect the types of individuals who select into the mission sector. The earlier analysis indicates that flat pay systems, in particular, encourage low ability individuals to enter the sector, suggesting that the choice of pay scheme may have long run effects on effort, even when no effects emerge among incumbent workers in the short run. The final piece of analysis in this section therefore examines effort exerted by subjects who choose to work in the mission sector under the respective pay schemes.

Prior to selection (i.e., among incumbent mission sector workers) only pay for performance elicits greater ability-adjusted effort than the flat pay scheme. However, motivated workers are more likely to select into the mission sector. This raises the possibility that there may be no differences in effort after selection has occurred. To test this, we run the same specifications as in table 3, except with ability-adjusted effort with those subjects that have chosen the Mission sector. Table 5 presents these results.

From our original sample of 430 subjects, 219 elected to join the mission sector (51 percent). The results in table 5 are treatment effects on ability-adjusted effort of just these subjects who have selected into the mission sector, in contrast to table 3, which looks at all subjects prior to selection. The first row indicates that subjects who selected into the pay-for-ability mission sector exert significantly higher effort than those who select the flat pay mission sector ($p < 0.10$). Effort effects for the pay-for-performance treatment are of the same magnitude as pay for ability, but are not significantly different ($p = 0.15$), reflecting the lower sample size of the pay-for-performance treatment. Importantly, however, effort between the pay-for-ability and pay-for-performance are not different from each other ($p = 0.97$). Using a log-linear specification

(to ease interpretation), subjects in the pay-for-ability and pay-for-performance exert 5 percent more effort than the flat pay treatment. This is expected since, as shown in the earlier section, both pay-for-ability and pay-for-performance attract statistically identical ability-motivation profiles, and are more likely to attract motivated subjects. Thus, effort under the low-powered pay for ability system and the high-powered pay for performance system is statistically indistinguishable.

Table 5: Post-selection treatment effects on effort in pro-social organizations

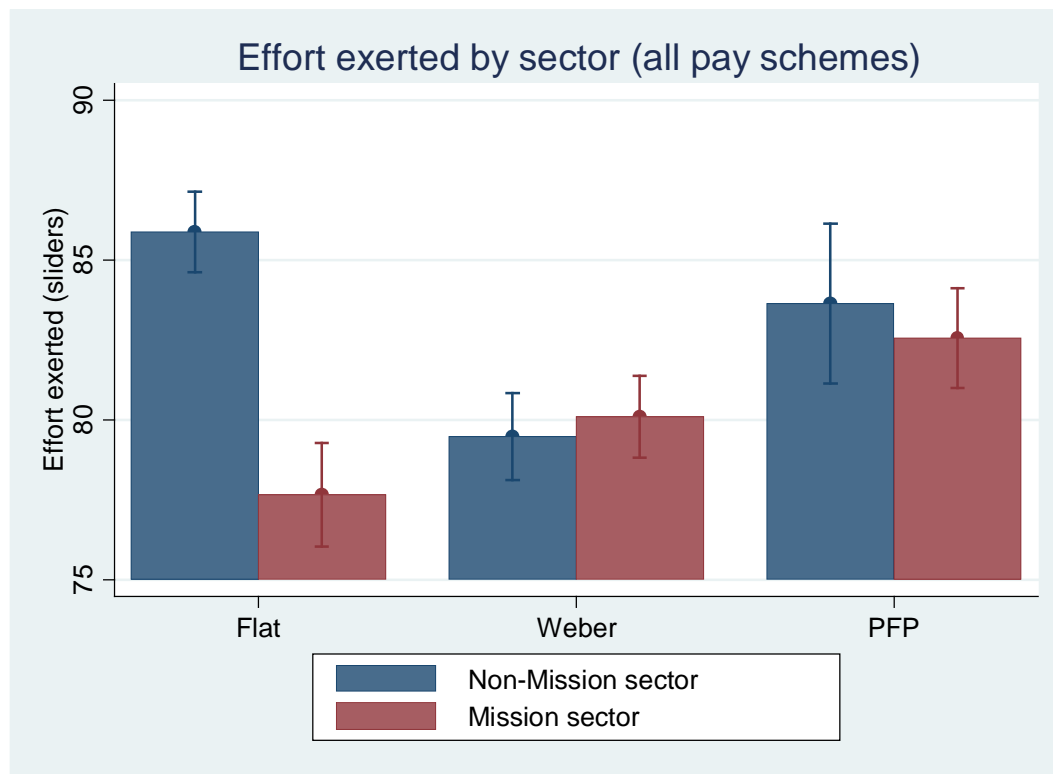
Dependent Variable: Effort for Charity Relative to Ability			
Dependent Variable:	Ability-adjusted Effort I	Ability-adjusted Effort II	Ability-adjusted Effort III
Pay-for-Ability (D)	0.033* (0.02)	0.032* (0.02)	0.031* (0.02)
Pay-for-Performance (D)	0.032 (0.02)	0.034 (0.02)	0.033 (0.02)
Amount Sent in Dictator Game		0.009 (0.02)	0.006 (0.02)
Gender (D) (1 = Female)		-0.007 (0.02)	-0.005 (0.02)
Age (in years)		0.009 (0.01)	0.008 (0.01)
Religious Attendance (5 = More than Once a Week)			0.012 (0.01)
Family Income (Relative to Others) (5 = Much Above Average)			0.001 (0.01)
Accounting Major (D)			0.011 (0.02)
Tax Major (D)			0.007 (0.02)
Belief that Charity was Paid (5 = Complete Confidence)			0.014 (0.01)
Constant	1.037*** (0.01)	0.851*** (0.15)	0.751*** (0.17)
R-squared	0.020	0.030	0.050
P-Value	0.115	0.263	0.363
Observations	219	219	219

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable is ability-adjusted effort in the choice task, restricted to subjects that selected the Mission Sector. OLS specification, standard errors in parentheses. The dictator game variable is divided by 1000 for presentational convenience.

Not only is post-selection effort under pay for ability equivalent to effort under pay for performance, both pay systems elicit greater (raw, not ability-adjusted) effort among those who select into the mission sector than the flat pay treatment. In the flat pay system, we find subjects successfully manipulate 77.67 sliders on average, while in the pay for ability treatment, subjects manipulate 80.10 sliders on average, which is higher, but not significantly different (two-tailed t-test: $p=0.24$). For the pay for performance treatment, subjects exert significantly greater effort than flat pay, with 82.55 sliders on average ($p<0.10$). However, this effort is not significantly different from the pay for ability treatment ($p=0.27$). Thus, the combination of sorting and incentives yield an improvement by pay for performance systems over flat pay systems, but not over pay for ability systems.

Figure 2 displays these results. It compares, within each of the three pay treatments, the effort exerted by those who select into the mission sector and those who select into the non-mission (piece-rate) sector. Under the flat pay treatment, consistent with the earlier logic, subjects who select into the piece rate non-mission sector exert significantly greater effort (two tailed t-test: $p<0.01$).

Figure 2: Effort by pay treatment: mission vs. non-mission workers



Note: Each pair of columns compares the effort of subjects who chose the pro-social task with the effort of those who chose the piece rate task under each pay treatment (flat, pay for ability or Weberian, and pay for performance or PFP).

Under the Weberian (pay-for-ability) pay system, no difference emerges. Effort by those who preferred the piece-rate, non-mission task over the Weberian, mission task was essentially the same as the effort by those who chose the mission task ($p=0.74$). Weberian systems attract sufficiently able, motivated workers that – despite the absence of high-powered incentives – they

are able to elicit the same effort as the non-mission task with high-powered (piece-rate) incentives.

Results are similar under the pay for performance system: comparing effort by workers who select into the mission sector under a high-powered pay system, pay for performance, with those who select into the piece-rate non-mission sector, no significant differences in effort emerge ($p=0.70$). Those who select into the pay for performance mission sector exert significantly more effort than those who select into the flat pay mission sector, however. Like pay for ability, pay for performance systems also succeed in attracting sufficiently able and motivated individuals into the mission sector that their performance rivals that of those who prefer the non-mission task. However, Weberian pay systems accomplish this even in the absence of high-powered incentives.

Which wage systems elicit effort at lowest cost?

Pay for performance seems to have the most systematic effect on effort, increasing effort (significantly) in the short term (with incumbents) and in the long term (after selection). However, the wage cost of effort also varies widely across pay systems, and is highest under pay for performance. Under the flat pay system, all subjects that selected the mission sector earned the same pay (6,600 tokens). Under pay for ability, subjects that selected the mission sector earned one of three possible pay levels, 4,200 tokens (30% of subjects), 6,600 tokens (46%) and 8,100 tokens (24%).¹¹ Average pay under pay for ability was therefore 6,226 tokens and less than under flat pay ($p<0.05$), even though pay for ability elicited significantly greater effort. For pay-for-performance, we used the same pay thresholds as in pay-for-ability, but this time conditional on subjects achieving the specific output targets associated with the thresholds. For this pay scheme, average pay was 7,163 tokens, which was significantly higher than the flat pay scheme ($p<0.01$) and higher than the pay for ability pay scheme ($p<0.01$). While both the pay-for-ability and pay-for-performance pay schemes elicited greater effort post-selection (relative to flat pay), average wage costs were significantly lower for pay-for-ability.

Using the effort and earnings data, we construct an output-earnings ratio, the effort exerted by the subject in the mission sector (multiplied by 100: the implied value of each slider) divided by the subject's wage cost (earnings) in the post-selection task. If subject output is exactly equal to subject compensation, the ratio is 1.0. Figure 3 presents the output-earnings ratio for subjects choosing the mission sector in the three treatments. We first note, that in all three treatments, the ratio is significantly higher than 1 ($p<0.01$ for all three treatments): subjects generate greater output than their compensation.

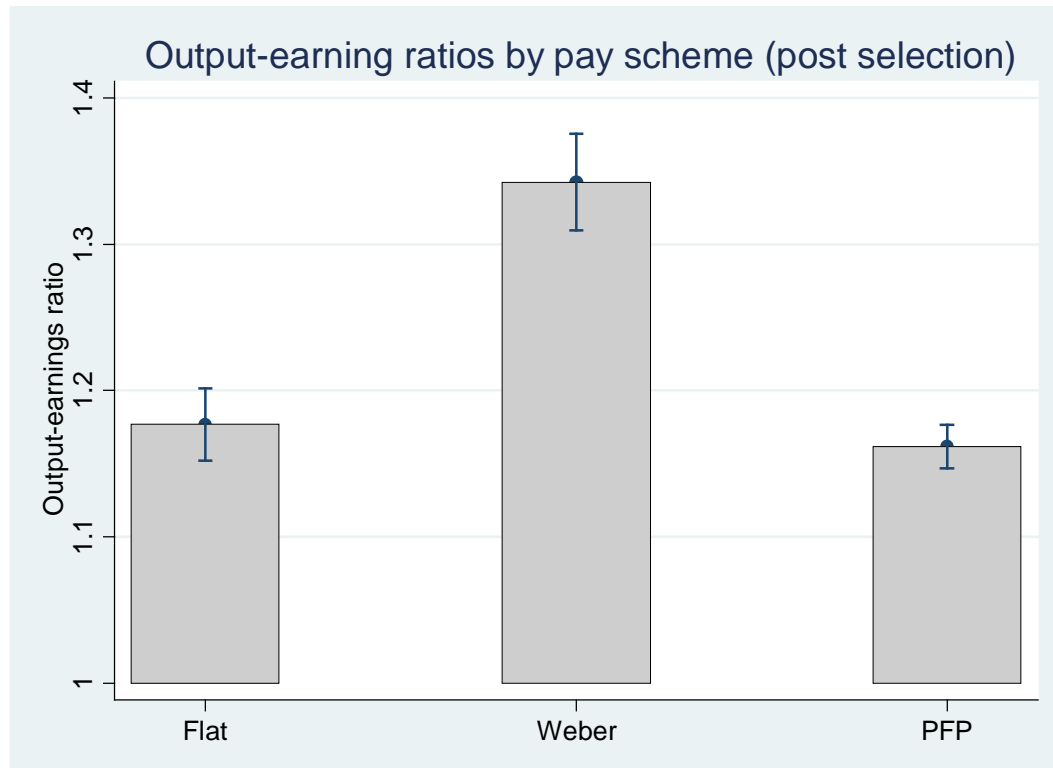
More importantly, we find no significant differences in the ratio between the flat pay and pay-for-performance treatments ($p=0.70$). In both treatments, the value that subjects generate is between 116 and 118 percent of their compensation. The ratio is significantly higher for the

¹¹ Subjects were locked in to the pay level based on the ability grading task.

Weberian pay-for-ability treatment relative to the flat pay ($p < 0.01$) and pay-for-performance ($p < 0.01$) treatments: subjects generate 134 percent of their compensation.¹²

In sum, we see that the pay for performance scheme increases output for the entire set of public officials, and is also able to effectively retain able subjects (relative to flat). Pay-for-ability is also able to retain the same subjects, and is the least expensive in terms of wage costs, while pay-for-performance is the most expensive.

Figure 3: Output-earnings ratio by treatment for mission workers



Note: Each bar is the average, within each treatment group (the flat pay, Weberian or pay for ability, and PFP or pay for performance groups), of the ratio of subjects' total output, equal to the number of sliders times 100, divided by the compensation they received, focusing only on those subjects who chose to enter the mission or pro-social sector under the respective pay schemes. If subject output is exactly equal to subject compensation, the ratio is 1.0.

Discussion

We report four main findings. First, all three pay schemes tested here are equally effective in attracting motivated workers. However, both pay for ability and pay for performance are more likely to attract high ability workers, relative to flat pay systems. Importantly, however, we find no significant differences in the ability-motivation profiles of workers attracted by pay for ability and pay for performance. Second, pay for performance schemes generate significantly more effort among incumbents, but this scheme operates through its effects on

¹² The output earnings ratio results are nearly identical for the entire sample in the pro-social task (task 2).

unmotivated workers; motivated workers exert equivalent effort in all three pay schemes. Third, once selection effects are accounted for, we find no significant differences in effort between the pay-for-ability and pay-for-performance systems. Fourth and finally, pay for ability contracts attract better workers at lower cost, yielding a significantly greater ratio of output to earnings.

No prior empirical work has compared the selection effects of flat pay, pay for ability and pay for performance schemes, though flat pay and pay for performance have, individually, been the subject of substantial attention. Finan, dal Bo and Rossi (2014) conclude that higher salary offers attract more able and not less motivated individuals; Banuri and Keefer (2013) show that in a context with important differences with respect to the reservation wage of mission-oriented workers and the social motivation of applicants, higher flat pay attracts significantly less motivated individuals than lower flat pay, but has no significant effect on ability.

Two concerns have triggered a large debate around pay for performance in the public sector (for thorough reviews, see Perry, Mesch and Paarlberg 2006 and Hasnain, Manning and Pierskalla 2012). One is the possibility, standard in principal-agent models, that high-powered incentives can have perverse effects in the face of mis-measurement and gamesmanship. Holmstrom and Milgrom (1991), for example, argue that low-powered contracts proliferate, even in environments where worker output can be accurately and easily measured, when worker tasks are multidimensional, only some tasks can be measured, and greater effort on measurable tasks comes at the expense of measurable output. The other concern is the possibility that pay for performance might crowd out intrinsic motivation to serve the public sector.

Our experimental setting excludes the first possible difficulty and allows us to focus entirely on the second: we find that pro-social organizations that employ pay for performance continue to attract more mission-motivated individuals. At the same time, contrary to the intuitive assumption in Holmstrom and Milgrom (1991), that high-powered pay is necessary to generate greater effort, we show that one low-powered pay system – pay for ability – generates similar effort on public sector tasks as a high-powered pay system – pay for performance – after accounting for the effects of entry on the type of individuals who enter the sector.

The central question in empirical investigations of pay for performance is simply whether it yields greater effort relative to baseline pay systems, whatever those might be. The introduction of incentive pay into the Brazilian tax collection agency increased fines collected by 75 percent (Kahn, Silva, et al. 2001). Glewwe, Ilias and Kremer (2010) and Muralidharan and Sundaraman (2011) look at performance pay in education (higher pay for higher test scores) in Kenya and Andhra Pradesh, respectively. Both studies report significant effects on outcomes (higher student test scores), though additional evidence in Glewwe, et al. (2010) suggests that these results could be driven by teacher efforts to “teach to the test”. The analysis below examines the effects of pay for performance, using unambiguous measures of effort, relative to rigorously-identified comparator pay systems, and accounting for longer-term selection effects.

We do not compare pecuniary and non-pecuniary incentives, like Ashraf, et al. (2012), who examine the effects of high-powered (performance-linked) pecuniary incentives and non-pecuniary incentives in the case of a somewhat unique task (the sale of female condoms by hair stylists in Lusaka). Non-pecuniary incentives (the awarding of gold stars) encouraged more effort.

Our analysis also contributes to research investigating the potentially heterogeneous effects of pay systems across types of workers. Past research focuses most on worker

heterogeneity with respect to ability and risk aversion. For example, in their model of managerial pay in the private sector, Goldmanis and Ray (forthcoming) conclude that firms seeking both to attract high-ability managers and to encourage them to work hard will use less performance-sensitive pay schemes than firms seeking only to encourage effort; this effect is exacerbated when managers are risk-averse. In their laboratory experiments, Dohmen and Falk (2012) show that, in fact, higher ability (also less risk-averse and male) subjects are more likely to prefer performance-based pay. In contrast to this work, we focus on pro-sociality and the decision to enter into mission versus non-mission sectors.

The research in Dohmen and Falk (2011) comes closest to investigating the questions we investigate here. They ask subjects to multiply one-digit numbers by two-digit numbers and to choose whether to undertake this task in return for a fixed payment, or in one of three variable pay schemes, including a piece rate scheme. Output is higher in all variable-payment schemes than under a fixed payment. As in our results, the effects are mostly due to selection (individuals sorting into the pay schemes that offer them the highest reward, given their ability). The research reported here differs in several ways, however. We examine effects on self-selection into a pro-social organization; our comparator pay schemes in the pro-social organization include a very common pay system that is overlooked in prior research, ability-based pay; and we examine the pay effects on the pro-social motivation of individuals who select into pro-social organizations.

Our analysis also relates to a broader literature on the measurement of motivation and pro-social behavior and on the effects of salaries on the ability and intrinsic motivation of new recruits and retained incumbents into organizations with a pro-social mission. Rather than rely on widely-used personality questionnaires, we continue a recent trend to use dictator games to measure the mission-orientation of subjects (Banuri and Keefer, 2013; Hanna and Wang, 2014; Deserranno, 2014). Our evidence is consistent with a theoretical literature that predicts that pro-social individuals are more likely to act pro-socially when the mission of the organization to which they belong is pro-social. However, this literature predicts that they will also work harder if the organization chooses a compensation scheme that assumes that they will act pro-socially (see especially Ellingsen and Johannesson 2008, but also Bénabou and Tirole, 2006 and Andreoni and Bernheim, 2009). In fact, we find that high-powered incentive schemes in the mission sector yield as much effort as low powered, pay for ability schemes, though both yield more effort than flat pay systems, detached from both ability and effort.

This paper also contributes to the literature examining the consequences for selection and optimal wage systems of matching – or not – the mission-orientations of workers and organizations (Frey, 1997; Francois, 2000; Besley and Ghatak, 2005; Brewer and Selden, 1998; Crewson, 1997; Perry, 1996; Perry and Wise, 1990; Sheehan, 1996; Tirole, 1994; Wilson, 1989). We find some evidence that organizations that can attract individuals who share the organization's mission will undertake activities at lower cost and be subject to less shirking (Besley and Ghatak, 2003, 2005; Delfgaauw and Dur, 2007; Dixit, 2002; Francois, 2007; among others).

Conclusion

We attempt to shed light on the public sector pay debate with the finding that, both in terms of selection into and subsequent effort in pro-social tasks, pay for performance and pay for ability perform equivalently: they attract higher ability than flat pay systems, with no loss of pro-social motivation, and they yield equivalent and high effort. We conclude that reform efforts should focus on encouraging public sector and other mission organizations to move away from compensation systems that reward neither ability nor performance to systems that do one or the other.

The choice is likely to be dictated by the capacity of organizations to measure either ability or performance in a way that corresponds to the outputs that the organization cares about, and the credibility of its commitment to stick to its announced pay scheme. If, for example, performance on a job (designing appropriate regulations or increasing student learning) is difficult to measure accurately, but abilities that are strongly correlated with achieving those tasks are easier to measure, then our experiments point to greater emphasis on implementing pay for ability rather than pay for performance schemes.

Acknowledgements: The authors are grateful for financial support from the World Bank, and acknowledge helpful conversations with Ghazala Mansuri and very helpful comments from David Stasavage and participants at seminars in the NYU Wagner School of Public Service. In addition, the authors are grateful to Riatu Qibthiyyah at the University of Indonesia, Dr. Muhammad Taufiq at STIA, Mr. Ridwan Galela at STAN, and Maria Tambunan at the World Bank, for arranging for access to the three institutions where we conducted the experiments; and to Eric McLester for his invaluable help in running the experiments. Funding for the experiments was provided by the Knowledge for Change Program (KCP).

References

- Andreoni, James and B. Douglas Bernheim. 2009. "Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects." *Econometrica* 77(5):1607–1636.
- Ashraf, Nava, Oriana Bandiera, and Kelsey Jack (2012). "No Margin, No Mission? A Field Experiment on Incentives for Pro-Social Tasks." CEPR Discussion Paper 8834 (February).
- Banuri, Sheheryar and Philip Keefer (2013). "Intrinsic Motivation, Effort and the Call to Public Service." World Bank Policy Research Working Paper Series 6729 (December).
- Bénabou, Roland and Jean Tirole. 2006. "Incentives and prosocial behavior." *American Economic Review* 95(5):1652–1678.
- Besley, Timothy and Maitreesh Ghatak. 2003. "Incentives, choice, and accountability in the provision of public services." *Oxford Review of Economic Policy* 19(2):235–249.
- _____. 2005. "Competition and incentives with motivated agents." *American Economic Review* 95(3):616–636.
- Brewer, Gene A. and Sally Coleman Selden. 1998. "Whistle blowers in the federal civil service: New evidence of the public service ethic." *Journal of Public Administration Research and Theory* 8(3):413–440.
- Carpenter, Jeffrey, Cristina Connolly and Caitlin Myers. 2008. "Altruistic behavior in a representative dictator experiment." *Experimental Economics* 11:282–298.
- Carpenter, Jeffrey, and Erick Gong. 2013. "Motivating Agents: How Much Does the Mission Matter?" *IZA Discussion paper series* 7602.
- Crewson, Philip E. 1997. "Public-service motivation: Building empirical evidence of incidence and effect." *Journal of Public Administration Research and Theory* 7(4):499–518.
- Dal Bo, Ernesto, Federico Finan and Martin A. Rossi (2012). "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." Mimeo, (May).

- Delfgaauw, Josse and Rober Dur. 2007. "Signaling and screening of workers' motivation." *Journal of Economic Behavior and Organization* 62(4):605–624.
- Deserrano, (2014).
- Dixit, Avinash. 2002. "Incentives and organizations in the public sector: an interpretative review." *The Journal of Human Resources* 37(4):696–727.
- Dohmen, Thomas and Armin Falk (2011). "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101:2, 556-590 (April).
- _____, David Huffman, and Uwe Sunde. (2012). "The Intergenerational Transmission of Risk and Trust Attitudes." *The Review of Economic Studies* 79(2): 645-677.
- Eckel, Catherine C. and Phillip J. Grossman. 1996. "Altruism in anonymous dictator game." *Games and Economic Behavior* 16:181–191.
- Eckel, Catherine C. and Phillip J. Grossman. 1998. "Are women less selfish than men? Evidence from dictator experiments." *The Economic Journal* 108(448):726–735.
- Eckel, Catherine C., and Philip J. Grossman. 2008. "Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices." *Journal of Economic Behavior and Organization* 68 (1): 1-17.
- Ellingsen, Tore and Magnus Johannesson. 2008. "Pride and prejudice: The human side of incentive theory." *American Economic Review* 98(3):990–1008.
- Forsythe, Robert, Joel L. Horowitz, N. E. Savin and Martin Sefton. 1994. "Fairness in simple bargaining experiments." *Games and Economic Behavior* 6:347–369.
- Francois, Patrick. 2000. "'Public service motivation' as an argument for government provision." *Journal of Public Economics* 78:275–299.
- Francois, Patrick. 2007. "Making a difference." *The RAND Journal of Economics* 38(3):pp. 714–732.
- Frey, B.S. 1997. *Not Just for the Money: An Economic Theory of Personal Motivation*. Cheltenham: Edward Elgar Press.

- Gill, D. and Prowse, V. (2011) “A Structural Analysis of Disappointment Aversion in a Real Effort Competition.” *American Economic Review*.
- Glewwe, Paul, Nauman Ilias and Michael Kremer (2010). “Teacher Incentives.” *American Economic Journal: Applied Economics* 2, 205-27.
- Goldmanis, Maris and Korok Ray (forthcoming). “Sorting Effects of Performance Pay.” *Management Science*.
- Hanna, Rema, and Shing-Yi Wang. Dishonesty and Selection into Public Service. No. w19649. National Bureau of Economic Research, 2013.
- Hasnain, Zahid, Nicholas Manning and Jan Henryk Pierskalla (2012). “Performance-related Pay in the Public Sector.” World Bank Policy Research Working Paper 6043(April).
- Holmstrom, Bengt and Paul Milgrom (1991). “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design.” *Journal of Law, Economics and Organization* 7 (January): 24-52.
- Kahn, Charles M., Emilson C. D. Silva, and James P. Ziliak (2001). “Performance-Based Wages in Tax Collection: The Brazilian Tax Collection Reform and Its Effects.” *The Economic Journal* 111 (January): 188-205.
- Li, Sherry Xin, Catherine C. Eckel, Phillip J. Grossman and Tara Larson Brown. 2011. “Giving to government: Voluntary taxation in the lab.” *Journal of Public Economics* 95(9-10):1190–1201.
- Muralidharan, Karthik and Venkatesh Sundararaman (2011). “Teacher Performance Pay: Experimental Evidence from India.” *Journal of Political Economy* 119:1, 39-77.
- Perry, James L. and Lois Recascino Wise. 1990. “The motivational bases of public service.” *Public Administration Review* 50(3):367–373.
- Perry, James L. (1996). “Measuring Public Service Motivation An Assessment of Construct Reliability and Validity.” *Journal of Public Administration Research and Theory* 6(1): 5 -22.

- Perry, James L., Debra Mesch and Laurie Paarlberg (2006). "Motivating Employees in a New Governance Era: The Performance Paradigm Revisited." *Public Administration Review* (July/August).
- Prior, Markus and Arthur Lupia (2008). "Money, Time, and Political Knowledge: Distinguishing Quick Recall and Political Learning Skills." *American Journal Political Science* 52: 168-182.
- Sheehan, R. (1996) "Mission Accomplishment as Philanthropic Effectiveness: Key Findings from the Excellence in Philanthropy Project." *Non-profit and Voluntary Sector Quarterly*, 25 (1): 110–23.
- Tirole, Jean. 1994. "The internal organization of government." *Oxford Economic Papers* 46(1):1–29.
- Whitt, Sam and Rick K. Wilson. 2007. "The dictator game, fairness and ethnicity in postwar Bosnia." *American Journal of Political Science* 51(3):655–668.
- Wilson, James Q. 1989. *Bureaucracy: What Governments Do and Why they Do It*. New York: Basic Books.

Table A1: Who chooses the mission sector under different pay schemes?

Dependent Variable: Sector Choice (1 = Join Mission Sector)			
Treatment	Flat	Pay-for-Ability	Pay-for-Performance
	I	II	III
Amount Sent in Dictator Game	0.964*** (0.35)	1.349*** (0.32)	1.686** (0.73)
Ability (Effort Exerted in Piece Rate)	-0.048*** (0.02)	0.006 (0.01)	0.038 (0.03)
Gender (D) (1 = Female)	-0.591 (0.45)	-0.213 (0.37)	0.674 (0.93)
Age (in years)	0.025 (0.19)	-0.054 (0.11)	0.883** (0.39)
Family Income (Relative to Others) (5 = Much Above Average)	0.212 (0.23)	0.084 (0.22)	-0.359 (0.44)
Risk Preferences (6 = Risk Seeking)	-0.150 (0.12)	-0.052 (0.11)	0.039 (0.26)
Religious Attendance (5 = More than Once a Week)	-0.107 (0.19)	0.031 (0.16)	0.797* (0.42)
Belief that Charity was Paid (5 = Complete Confidence)	0.109 (0.19)	0.223 (0.23)	-0.541 (0.51)
Charity Rating (7 = Most effective)	0.146 (0.13)	0.227** (0.12)	0.061 (0.25)
Constant	2.035 (4.30)	-2.721 (2.80)	-21.61** (9.15)
Log Likelihood	-108.2	-120.4	-29.2
Pseudo R-squared	0.111	0.104	0.259
P-value	0.001	0.001	0.016
Observations	176	195	60

Note: * p<0.1, ** p<0.05, *** p<0.01. Logit specification, standard errors in parentheses. The dictator game variable is divided by 1000 for presentational convenience.

Table A2: Treatment effects on effort in pro-social organizations, pre-selection (Unmotivated vs. motivated subjects)

Dependent Variable: Effort for Charity Relative to Ability		
	Unmotivated subjects (Dictator ≤ 500)	Motivated subjects (Dictator > 500)
	I	II
Pay-for-Ability (D)	0.027 (0.04)	0.016 (0.03)
Pay-for-Performance (D)	0.153** (0.06)	0.022 (0.04)
Amount Sent in Dictator Game	0.020 (0.09)	-0.026 (0.03)
Gender (D) (1 = Female)	-0.003 (0.04)	0.037 (0.03)
Age (in years)	0.018 (0.02)	-0.013 (0.01)
Religious Attendance (5 = More than Once a Week)	0.010 (0.02)	-0.006 (0.01)
Family Income (Relative to Others) (5 = Much Above Average)	0.039 (0.02)	0.004 (0.02)
Accounting Major (D)	0.048 (0.05)	-0.048 (0.04)
Tax Major (D)	0.033 (0.05)	-0.004 (0.04)
Belief that Charity was Paid (5 = Complete Confidence)	-0.043* (0.02)	0.019 (0.02)
Constant	0.625 (0.39)	1.234*** (0.24)
R-squared	0.057	0.062
P-Value	0.136	0.404
Observations	259	171

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable is ability-adjusted effort in the Pro-social task (task 2). OLS specification, standard errors in parentheses. The dictator game variable is divided by 1000 for presentational convenience.