Performance-related pay, fairness perceptions, and effort in public management tasks: a parallel encouragement design

Paolo Belardinelli
London School of Economics and Political Science, UK

Nicola Belle
Sant'Anna School of Advanced Studies, Italy

Paola Cantarelli (iD)
Sant'Anna School of Advanced Studies, Italy

Paul Battaglio (iD)
University of Texas at Dallas, USA
Sant'Anna School of Advanced Studies, Italy

Abstract

This randomized study explores the causal mechanisms linking contingent pay to individual performance on a series of tasks mimicking real public management activities. Employing a *parallel encouragement design* in a laboratory setting, we disentangle the overall, direct, and indirect performance effects of perceived fairness as well as a pay scheme that reproduces the merit system provisions adopted by the Italian government. The overall performance effect of that contingent pay scheme turned out to be insignificant when averaged across the four experimental tasks. However, a significant pay-for-performance effect was detected for the most routine task. Moreover, we observed heterogeneity in the treatment effect depending on the participants' relative positioning in

Corresponding author:
Paul Battaglio, University of Texas at Dallas, 800 W Campbell Rd, Richardson, Texas 75080, USA.
Email: battaglio@utdallas.edu

the performance ranking. Overall, the data do not provide support for a mediation model linking contingent pay-for-performance through perceived fairness.

Points for practitioners

- Workers tend to perceive pay-for-performance as fairer than equal pay.
- The effectiveness of pay-for-performance seems to be greater for more routine tasks.
- Public organizations and their managers should be aware that the effects of pay-for-performance may be unpredictable because they depend on a multitude of factors.

Keywords
human resources management, performance, public management

Introduction

The debate about the use of pay-for-performance or performance-related pay (PRP) is not new in public management scholarship. Perry et al. (2009) conclude a review of 57 studies published between 1977 and 2008 by stating that "performance-related pay in the public sector consistently fails to deliver on its promise" (p. 43). Indeed, extant evidence now suggests that the effectiveness of financial rewards depends on such factors as the size of the incentives (e.g. Belle and Cantarelli 2015; Gneezy and Rustichini 2000), their visibility or transparency (e.g. Ariely et al., 2009; Belle, 2015), the time at which they are distributed (e.g. Fryer et al., 2012), the type of task (e.g. Weibel et al., 2010), as well as individual characteristics of the subjects performing the task (e.g. Dal Bó et al., 2013; Forest, 2008).

Yet, 80% of OECD countries have adopted some form of PRP since the 1980s (Lah and Perry, 2008). Italy is no exception, with a variety of PRP schemes introduced in the early 1990s across public organizations operating at the national, regional, and local levels. The implementation of PRP provisions has typically led to a lack of differentiation among employees, resulting in most employees assigned to the highest performance category (e.g. Corte dei Conti, 2003; Ministero per la Pubblica Amministrazione, 2008). Subsequently, the Italian government issued Legislative Decree No. 150 of 2009 introducing a forced-ranking system, whereby PRP is based on a ranking of public employees within each agency. Employees are assigned to one of three ranking clusters: high merit, intermediate merit, and low merit. Allocations for each ranking are 25%, 50%, and 25% of the employees respectively, based on their measured performance. Each agency has a predetermined amount of money for the payment of merit-based bonuses, which is split among the three clusters: 50%, 50% and 0% to the high, intermediate, and low merit clusters respectively.[1] Our study explores the effect of introducing and actually enforcing such a contingent pay system as opposed to introducing it formally and failing to enforce it by equally distributing monetary incentives while accounting for perceived fairness. Specifically, we test the causal relationship between enforcing a PRP scheme and performance across task types and the mediating role of perceived fairness in a laboratory setting adopting a parallel encouragement design (Imai et al., 2013).

We adopt three theoretical lenses that have been previously used in public administration research, namely PRP (Lah and Perry, 2008), psychological contract (Kellough and Nigro, 2002), and equity theory (Rainey, 2009). We find that enforcing the PRP scheme has a significant effect only in the most clerical oriented and least interesting task performed by participants, suggesting that the effectiveness of merit pay schemes may be conditional not only on the correct implementation of the scheme, but also on the complexity of job tasks performed by public employees. We do not detect any significant mediating role of perceived fairness of the supervisor in the relationship between the PRP scheme and performance.

Research about the effectiveness of monetary incentives within the public sector is still inconclusive, mostly relying on observational studies that focus on attitudes or perceptions from survey instruments, with few exceptions (e.g. Belle, 2015; Belle and Cantarelli, 2015; Dal Bo et al., 2013). We contribute to this literature by providing a new piece of experimental evidence on the relationship between pay schemes and individuals' work effort. Methodologically, our study contributes to the broader public administration, management, and policy literature by adopting a parallel encouragement design (Imai et al., 2013) that is specifically suited for identifying not only causal effects, but also causal mechanisms.

Moreover, experimental evidence increasingly offers a practical means for taking on difficult public problems (Tummers 2019). Our performance-related pay experiments here offer practical solutions for managers interested in employee perceptions of fairness toward reward distribution and the impact of job task type.

Literature review and hypotheses

Performance-related pay and its theoretical mechanisms

Compensation systems are an integral part of decision-making within organizations (Pfeffer, 1998). Implementing and developing compensation systems involves motivating and reinforcing desired behaviors among employees that ultimately provide support for an organization's overall mission and objectives (Gupta and Shaw, 2014). Of particular interest is how we attribute psychological meaning of financial compensation to employees. Since employee interpretations of compensation often drive their subsequent behavior, we should be engaged in analyzing said behavior and its impact on organizational performance (Munyon et al., 2016: 119). To this end, scholars have argued that compensation systems predicated upon psychological mechanisms are likely to be key drivers of organizational performance (Munyon et al., 2016). These mechanisms employ theoretical constructs, such as psychological contract theory and equity theory, as a means for making distinctions among employees in terms of performance.

Traditionally, pay structures in the public sector were based on longevity or time-in-service, focusing on career management and advancement instead of employee performance. High-performing civil servants were not rewarded, and underperforming employees were not given incentives to improve. Since the 1980s, the push for greater accountability has resulted in an increased focus on performance management by

rewarding superior performers with pay increases. PRP structures have proven to be an attractive alternative to traditional public sector practices. Appealing to the logic of market-like mechanisms, proponents of compensation reform suggest that PRP systems have the potential to increase employee performance and therefore organizational productivity. Several PRP schemes have been implemented by linking individual, team, and/or organizational performance measurement to financial incentives. Financial incentives can be distributed as increases to base pay (i.e. merit pay), one-time bonuses, or a combination of the two (Kellough and Nigro, 2002: 146).

PRP is primarily grounded in expectancy theory and reinforcement theory. Expectancy theory proposes that employees will exert additional effort in exchange for monetary reward. Thus, expectancy theory is tied to an individual's conviction that increased effort will result in valued results (Rainey, 2009; Van Eerde and Thierry, 1996). Reinforcement theory posits that the incidence of a desired behavior (e.g. job performance) is subject to its consequences (e.g. pay) (Perry et al., 2009). Reinforcement theory, therefore, supports the idea that PRP schemes are a behavior-reinforcing mechanism for improved performance (Perry et al., 2006). Based on these theories, the expectation is that better performance will lead to the reward of increased pay and, moreover, that attaining this reward will be a positive reinforcing factor, compelling the employee to continue the improved performance to continue receiving the incentive.

Perceptions of fairness in the workplace may also affect employees' attitudes and behaviors. Predicated upon equity theory, research suggests employees place a premium on equity and fairness within the workplace (Adams, 1965). In terms of compensation, employees base their fairness perceptions on social comparisons between themselves and others. Equity theory suggests that employees are more concerned with obtaining equitable outcomes than with maximizing outcomes, thus, judging fairness in terms of work outcomes relative to their co-workers (Munyon et al., 2016). For example, perceptions of inequity or unfairness accrue when an employee believes that the rewards received for their work compare unfavorably to the rewards their peers receive for their contributions to the workplace (Schermerhorn et al., 2004). Such social comparisons may persuade individuals to adjust their outputs to moderate pay inequity. Further, equalizing monetary rewards irrespective of performance differences among employees may be problematic because individuals value fairness more than equality, to the point that they have a strong preference for fair inequality over unfair equality (Starmans et al., 2017).

Perceptions of fairness, support, and transparency in PRP schemes are also closely tied to psychological contract theory (Argyris, 1960; Rousseau, 2001; Wenzel et al., 2019). The notion of a psychological contract stems from the work of Rousseau (1989), whereby the employee and employer establish an informal schema identified by mutual beliefs, perceptions, and expectations of employment. This informal dyadic relationship is a distinct concept from the more formal written contract that generalizes written duties and obligations of the job to be performed. It is a mutually agreed upon and binding understanding between the employee and employer (Rousseau, 2001). With respect to PRP, the psychological contract is the expectation of future payments in exchange for the work to be done (Robinson and Rousseau, 1994). Fair, participatory, and transparent design reduces the perception of friction between the two parties of the

contract while fostering the intrinsic motivation of employees (Wenzel et al., 2019). Thus, fairness is instrumental to the effectiveness of PRP schemes, in that employees should have a clear perception that their contribution is directly related to their receipt of said reward. Unfortunately, ensuring fairness in public sector PRP schemes is often problematic owing to the difficulties in objectively defining and linking performance (Perry et al., 2009). Research has consistently demonstrated that employees often perceive PRP schemes as lacking a clear link between on-the-job performance and the amount of performance pay awarded (Kellough and Nigro, 2002; Perry et al., 2009).

While touted as the magic elixir to performance problems in government, PRP has consistently failed to deliver on this promise (Bae, 2021; Keraudren, 1994; Perry, 1986; Perry et al., 2006; Perry et al., 2009). PRP research has suggested that the failure to connect the incentive scheme with performance may be owing to an inability to distinguish poor performance from other levels of performance, inadequate financial support, and insufficient evidence to confirm improved performance led to the demise of federal PRP systems (Battaglio, 2014; Choi and Whitford, 2017; Perry et al., 2009). Some public management scholars also point to implementation breakdowns as the main cause of those failures (e.g. Egger-Peitler et al., 2007; Marsden and Richardson, 1994). For instance, PRP provisions may fall short of expectations because the performance appraisal is perceived as unfair (Gabris and Ihrke, 2000) or because performance ratings are simply disregarded when distributing monetary rewards. It is telling that the survey item from the 2018 Federal Employee View Point Survey with the lowest level of agreement is "Pay raises depend on how employees perform their jobs" (United States Office of Personnel Management, 2018: 5). Yet, PRP systems – supported by politicians and managers alike – remain a prevalent remedy to performance and productivity in the public sector. Despite the problematic history of PRP systems, a strong desire to rid the bureaucracy of inefficiencies continues to be a compelling enough reason to pursue alternative pay systems. Alternative pay systems in the public sector continue to rely upon theoretical mechanisms, such as psychological contract theory and equity theory, as mediators for promoting individual performance and organizational effectiveness.

Expanding on the merits of PRP, supporters also suggest that along with productivity, increases in pay can also contribute to positive employee attitudes and/or behaviors. The logic follows that in a work climate that supports and rewards productivity, employees will be more likely to express feelings of satisfaction, accomplishment, belonging, job security perceptions, and esteem (Brown and Sessions, 2003; Godard, 2001; Green and Heywood, 2008). However, recent work (Choi and Whitford, 2017) suggests that the psychological well-being of employees in agencies with such incentive schemes is less than sanguine. Indeed, by incentivizing a highly competitive workplace PRP may be undermining its intended purpose. Instead of promoting productivity, the competitive environment in the public sector might lead to unwanted competition, lack of job security, risk, and perceptions of unfairness (Choi and Whitford, 2017; Condrey and Battaglio, 2007; Godard, 2001). PRP schemes have also encountered increased compensation differences within public organizations owing to budgetary constraints (Perry et al., 2009). Such pay differentials can contribute to perceptions of unfairness, frustration, and lower morale for less productive workers (Green and Heywood, 2008). These negative

perceptions are more acute given problems measuring and evaluating performance accurately and fairly (Choi and Whitford, 2017).

In the context of our experiments here, the psychological contract is implied by the monetary reward that participants are expecting in each round based on the evaluation of their performance in the tasks assigned. We then manipulate the enforcement of the contract with the presence of the supervisor in the lab, who will either reward their performance according to the default PRP scheme described later or deviate from that scheme. This manipulation might be perceived as a breach to the psychological contract, where one of the parties perceives the other to have failed to fulfill promised obligations (Robinson and Rousseau, 1994). According to Rousseau (1989), such breaches are not inconsequential in that obligations have been forfeited and norms of conduct have been violated. Thus, if a subject perceives the supervisor is deviating from the PRP scheme, they see the psychological contract as breached. As a result, subjects, and by association employees, may perceive this manipulation of meeting/disappointing expectations about rewards as upholding/violating that implicit psychological contract.

Based on these broad insights, we formulated and tested the following hypotheses (Figure 1, online[2]):

Hypothesis 1: When a PRP scheme is formally adopted, performance contingent rewards in accordance with the PRP scheme will increase fairness perceptions relative to equal rewards in violation of the PRP scheme.

Hypothesis 2: When a PRP scheme is formally adopted, performance contingent rewards in accordance with the PRP scheme will increase performance relative to equal rewards in violation of the PRP scheme.

Hypothesis 3: Fairness perceptions will mediate the impact in accordance with violation of the PRP scheme onto performance.

Research design and methods

Parallel encouragement design

The use of experimental designs is a powerful tool for social scientists to establish causal claims empirically (Imai et al., 2013). However, as Imai et al. (2013) note, an important critique of experiments is that they "merely provide a black box view of causality and fail to identify causal mechanisms" (p. 5). In other words, although best suited to identifying average causal effects, randomized controlled trials are unable to describe the mechanisms through which such effects play out. Imai et al. (2013) identify a potential solution in the parallel design, whereby two randomized experiments are conducted in parallel and "each subject is randomly assigned to one of two experiments; in one experiment only the treatment variable is randomized whereas in the other both the treatment and the mediator are randomized" (Imai et al. 2013: 6). A parallel encouragement design is a variation of the base design that can be used in cases where a perfect manipulation of the mediator is not feasible, as is usually the case with psychological factors. "Under the parallel

encouragement design, experimental subjects who are assigned to the second experiment are randomly encouraged to take (rather than assigned to) certain values of the mediator after the treatment has been randomized" (Imai et al., 2013: 6). In other words, in the parallel encouragement design, experimenters first randomly split subjects into two arms. Then, for subjects in Arm 1, the treatment is randomly assigned but no manipulation of the mediator is conducted. For subjects in Arm 2, experimenters randomize the treatment and the indirect manipulation to encourage subjects to take either a high or a low level of the mediator. The elements of the design are as follows:

> *Treatment* – The treatment in our experiment consisted in disattending a PRP scheme that – knowingly to participants – could be arbitrarily disregarded or applied by their supervisor. More precisely, subjects in the treated condition T received a fixed reward, that was unrelated to their performance in a series of tasks, thus disregarding the PRP scheme. Subjects in the control condition C, instead, were rewarded according to the preset PRP scheme.

> *Encouraged mediator* – Our design is meant to test whether equalizing rewards in violation of a PRP scheme may negatively impact performance through the mediator of perception of being unfairly paid. As this mediator could not be perfectly manipulated, we indirectly manipulated fairness perceptions as follows. Subjects that were encouraged to take a high level of the mediator (MH) read, "the supervisor fairly distributes the incentives. If you did not receive the incentive you deserved, this was due to an honest mistake." Subjects encouraged to take the low level of the mediator (ML) were prompted, "the supervisor unfairly distributes the incentives. If you did receive the incentive you deserved, this happened totally by chance."

> *Outcome* – The outcome in our experiment was change in performance across a series of tasks, as described in the following section.

Participants and procedures

Following up on recent calls for increased laboratory experimentation in public management (Walker et al., 2017), we ran our study at the Bocconi Experimental Laboratory for the Social Sciences (BELSS), using the z-Tree software (Fischbacher, 2007) with 192 Bocconi students. Throughout the experiment, participants received all the instructions exclusively through their computer screens. Subjects earned a show-up fee of €6 plus a monetary reward for carrying out four tasks. On average, participants earned €14. Based on our parallel encouragement design, 64 participants were assigned to Arm 1 and 128 to Arm 2. Figure 2 (online) provides a concise overview of the procedure for the experiments.

Arm 1 of the parallel design

Participants in the first arm of the parallel design were randomly assigned to either the treatment condition T, in which the supervisor disregarded the PRP scheme and rewarded

all the subjects the same irrespective of their performance, or the control condition C, in which the supervisor rewarded every subject based on their performance according to the PRP scheme. Participants knew that the supervisor in the lab, who was a confederate, would either reward their performance according to the default PRP scheme described below or arbitrarily deviate from that scheme.

At the beginning of the experiment, participants were invited to take a seat in the lab, where they could only see a computer screen through which they received instructions about the PRP scheme and the four experimental tasks. The instructions are described in Appendix 1 (online), which also includes examples of screens presented to subjects when performing their tasks. The first task (*GPA Task*) consisted in calculating the grade point average (GPA) for 10 students based on information reported on a paper document and typing the GPA into a computer. The second task (*Grades Task*) entailed typing 10 students' grades from a paper document, which contained a list of 100 grades of as many students, into a computer. In the third task (*Eligibility Task*), subjects were presented with paper documentation about 10 families who had applied for poverty subsidies; participants were tasked with deciding whether or not applicants were eligible or not eligible for the subsidy, according to a set of criteria illustrated in the documentation, and then to type their decision into a computer. In the last task (*Tax Task*), subjects had to decide whether or not the tax return forms filed by 10 taxpayers were correct, according to a set of rules described in the experimental documentation, and then type their decision into a computer. Participants had a maximum of 90 s for the *GPA Task*, 60 s for the *Grades Task*, 60 s for the *Eligibility Task*, and 180 s for the *Tax Task*. After a practice round in which participants had a chance to familiarize themselves with the four experimental tasks without being measured nor rewarded, each student performed two rounds that counted toward their performance. Before each round, subjects were presented with the PRP scheme for that round. Across the two rounds, each participant encountered two different PRP schemes, one more generous than the other. The more generous PRP scheme entailed an €8 reward for subjects performing in the top quartile, a €4 reward for participants in the second and third quartiles, and no monetary reward for those in the bottom quartile. The less generous PRP scheme was similar, with the only difference being that all the rewards were half of those in the more generous scheme – that is, €4 for those in the top quartile, €2 for those in the two middle quartiles, and zero for the bottom quartile. In half of the sessions, subjects randomly encountered the more generous scheme in the first round and the less generous in the second round. The order was reversed in the other random half of sessions, with the less generous PRP scheme in round one and the more generous scheme in round two. At the end of each round, participants were presented with three pieces of information: the anonymized performance ranking of participants in their session, the monetary reward they should receive based on the PRP scheme, and their actual reward. Before leaving the laboratory, subjects answered a set of questions described in the *Measures* section and received their monetary rewards.

Arm 2 of the parallel design

Arm 2 was the same as Arm 1 with the only difference that in Arm 2 both the treatment and the encouragement of the mediator were randomized. As a result, subjects were

randomly split into four experimental conditions: T×MH (i.e. the supervisor disregarded the PRP scheme and subjects were prompted to take high levels of perceived fairness); T×ML (i.e. the supervisor disregarded the PRP scheme and subjects were prompted to take low levels of perceived fairness); C×MH (i.e. the supervisor applied the PRP scheme and subjects were prompted to take high levels of perceived fairness); or C×ML (i.e. the supervisor applied the PRP scheme and subjects were prompted to take low levels of perceived fairness).

Measures

Performance. We measured task performance as the number of correct answers minus the number of incorrect answers. A correct answer counted as 1 point, a missed answer counted as 0 points, and a wrong answer counted as −1 points. Therefore, each subject's performance score can vary between −40 and +40. The 40 available points are equally distributed among the tasks. Our dependent variable was the difference in performance between round one, when subjects receive the incentive for the first time, and round two, when they performed the sequence of tasks for the last time.

Perceived fairness. We measured perceived fairness of the supervisor on a Likert-type item (1 = disagree strongly, 5 = agree strongly): "The supervisor rewarded me fairly". The item was presented to subjects at the end of the experiment, after the two sequences of tasks had been performed and the supervisor had already distributed the incentive.

Controls. At the end of the experiment, we asked participants questions about gender, age, and number of lab experiments already attended.

Results

Participants in our lab experiments were 20.5 years old and they had previously attended 4.6 experiments, on average. Forty-eight percent of them were female. The mean performance in the pre-treatment practice sequence of task was 10.3 points. On these variables, we did not detect any statistical difference among experimental groups assigned to different treatment and mediator conditions, with one exception: participants assigned to the PRP scheme were 6 months older, on average, than those assigned to the condition in which incentives were equally distributed.

Treatment effects

Performance improvement from round 1, when subjects receive the incentive for the first time, to round 2, when they perform the four tasks a second time, is equal to 1.97 points on average (SD = 4.93). Performance improvement is 0.75 (*p-value* = 0.146; N = 192) points higher when the PRP scheme is applied (Mean = 2.34, SD = 4.93) compared with when all participants receive an equal reward, irrespective of their performance (Mean = 1.59, SD = 4.93).

The relatively high *p-value* seems to be a consequence of the low power of our study linked to heterogeneous effects of the treatment across both subjects and tasks. The effect of implementing the PRP scheme on performance is indeed clearer for subjects who were ranked in the second quartile at the end of the first sequence of tasks (ATE = 2.09, *p-value* = 0.094; N = 55). Looking at the other quartiles, differences in performance improvement are not statistically significant. In addition, the average treatment effect is higher in the *Grades Task*, which is the most routine task of the session. When changing the dependent variable to performance improvement in this single task (Mean = 0.98, SD = 1.81), the effect of implementing the PRP scheme becomes 0.44 (*p-value* = 0.094; N = 192). For all the other tasks, differences in performance improvement are not statistically significant. This is consistent with findings from a systematic review showing that financial incentives turn less effective in improving performance in lab tasks as these become cognitively more complex (Bonner et al., 2000).

Causal mediation analysis

Perceived fairness of the supervisor was 3.54 (SD = 1.32) on average. Whether the supervisor applied the PRP scheme or not had a strong impact on perceived fairness (ATE = 1.10, *p-value* = 0.000; N = 192). In other words, perceived fairness of the supervisor was not independent of the treatment. When the incentive was equally distributed among participants (Mean = 2.99, SD = 1.40, N = 96), perceived fairness was almost one standard deviation lower with respect to when the incentive was distributed according to the PRP scheme (Mean = 4.09, SD = 0.98, N = 96) (Figure 5, online).

In order to identify the causal mediation effects of perceived fairness of the supervisor, standard approaches to causal mediation analysis (Baron and Kenny, 1986) would need to assume that the observed values of perceived fairness are conditionally independent of potential outcomes given the actual treatment status and observed pretreatment variables. This assumption is called the *sequential ignorability of the mediator* (Imai et al., 2013). In our case, this requires that the observed perceived fairness values are randomly chosen once the treatment is assigned. Figure 5 (online) suggests that this is unlikely to be the case. Therefore, the adoption of a parallel encouragement design seems to be particularly valuable here, for it allows for the testing of the causal mediation effect without relying on the sequential ignorability assumption, making it hold by design. Half of the subjects in our parallel experiment, in which the treatment was nonetheless randomized, were randomly encouraged to take low values of perceived fairness of the supervisor, while the other half of the subjects were randomly encouraged to take high values of the same variable.

Our manipulation (i.e. our encouragement) was effective in changing the perceived fairness of the supervisor, as perceived fairness in the low fairness condition (Mean = 3.28, SD = 1.49, N = 64) was significantly lower with respect to perceived fairness in the high fairness condition (Mean = 3.84, SD = 1.17, N = 64) (*p-value* = 0.018) (Figure 5, online).

We performed our design-based causal mediation analysis using the *R mediation package* developed by Tingley and colleagues (2014). Table 1 shows the outcome of

the analysis. The first model's dependent variable is overall performance improvement in the sequence of tasks. As far as this first model is concerned, there seems to be neither direct significant effect of implementing the PRP scheme nor indirect significant effect mediated by perceived fairness of the supervisor. The second model's dependent variable is performance improvement in the second task – that is, the most routine task. Here, we observed a direct significant effect of implementing the PRP scheme (0.6 points, $p =$ 0.04). However, perceived fairness of the supervisor does not mediate the effect of implementing the PRP scheme on performance. This is true for both the entire sample of analysis and for "compliers" – that is, those who responded to our encouragement towards different levels of perceived fairness. In other words, our estimates of the mediating effects are not informative as the sharp bounds cross zero.

Discussion and conclusion

In a laboratory experiment, we tested the effect of a PRP scheme on the task performance in four public administration domains mediated by individual perceptions of supervisor fairness. Our manipulation was predicated upon the PRP scheme introduced within the Italian public sector in 2009, which was never implemented owing to the financial crisis.

Participants in our experiments knew that a PRP scheme had been suggested to the supervisor in the lab but that the supervisor could deviate from it and distribute the incentives as he/she pleased. Performance improvement from the first to the second sequence of tasks was higher when subjects were paid according to the rule, rather than when the incentives were equally distributed among them. The difference in performance improvement between conditions was significant only in the second task, which is the most clerical oriented and least interesting of the four. Therefore, the hypothesis on the effect of

Table 1. Direct and indirect (through perceived fairness) effects of applying the merit pay scheme: design-based causal mediation analysis.

	Model 1 Overall performance improvement across all tasks (N = 192)	Model 2 Performance improvement on the most routine task (i.e. task 2) (N = 192)
Direct effect of PRP scheme		
Population	1.1 (0.14)	0.6 (0.04)*
Indirect effect (through perceived fairness) of PRP scheme		
Population (performance-unrelated pay)	0.0 (0.99)	−0.2 (0.65)
Compliers (performance-unrelated pay)	0.0 (0.99)	0.2 (0.28)
Population (performance-related pay)	−0.1 (0.67)	0.2 (0.25)
Compliers (performance-related pay)	−0.1 (0.41)	0.0 (0.99)

*Significant at the 5% level.

PRP scheme implementation on performance improvement is partially confirmed. Our results suggest that the effectiveness of merit pay schemes may be conditional not only on the correct implementation of the scheme, but also on the complexity of job tasks performed by public employees.

The importance of understanding causal mechanisms is critical to advancing experimental research in the social sciences, specifically public management (Imai et al., 2013; Walker et al., 2017). We contribute to public management research by disentangling the "black box" of experimental design to better appreciate causal processes. Our use of a parallel encouragement design in the experiments here provides an alternative that greatly enhances the power of laboratory experiments in public management research. As opposed to the single-experiment design, our parallel experiments not only uphold the randomization of the treatment, but also offer a means for manipulating a mediator – in this case, perceived fairness of the supervisor. More generally, our parallel encouragement design provides an alternative for informing social science (specifically, public management) theory regarding experimental conditions and whether treatments (and by assumption policies) are effective or ineffective. Such research has the potential to inform how to precisely design, develop, and implement policies effectively.

Turning to PRP specifically, our experiments support the notion that equalizing monetary rewards irrespective of performance differences among employees may be problematic, consistent with the idea that individuals do not necessarily value equality if this comes at the cost of unfairness (Starmans et al., 2017). What matters is that you adhere to the PRP scheme. By ignoring the PRP scheme and giving out equal performance rewards, managers undermine the scheme and the overall effectiveness of the plan. Additionally, given our focus on specific tasks in the parallel encouragement design, adherence to the PRP scheme would appear to be particularly important for routine activities rather than less routine tasks. A more cautious interpretation of our results may be that not following through with PRP provisions is detrimental for employee morale because it generates perceptions of unfairness, which seems more likely to translate into lower performance for routine tasks. This finding provides support to previous claims that PRP may play out differently depending on the complexity of the task at hand and be better suited for tasks that are less interesting and less ambiguous (Perry et al., 2009; Weibel et al., 2010). In this respect, our experimental design proves to effectively mirror – at least partially – the variation in task complexity that one can observe across real public administration activities. Indeed, although all four experimental tasks were quite repetitive, the second task was less interesting and ambiguous than the others, thus allowing us to estimate how task complexity moderates the performance impact of contingent pay.

Further, while most research supporting the appreciation of fairness in distribution relies on perceptions and attitudes from survey research, we offer experimental evidence. That is, in return for their "hard work" there is an expectation of being rewarded, thus reinforcing the power of psychological contract theory in buttressing the effectiveness of PRP. The default PRP scheme that was shown to subjects may be interpreted as an implicit psychological contract. Our manipulation of meeting/disappointing expectations about rewards models upholding/violating that implicit psychological contract. This

would speak to incomplete contract theory, which is often used in this area of research (Grossman and Hart, 1986; Hart and Moore, 1988). In interpreting our findings it should be considered the specific violation of the implicit psychological contract included in our experimental design, that is an equal distribution of the available resources. Evidence exists that public employees are particularly concerned with favoritism as a specific threat to pay for performance in their agencies (e.g. Kellough and Nigro, 2002), rather than giving equal rewards to all employees. This means that our analysis might underestimate the effects of violating the implicit contract, given that some participants received less than expected but others were ultimately given more than what they should have received. That said, we emphasize that our interest in this specific violation is determined by its relevance in the Italian public administration context (Corte dei Conti, 2003; Ministero per la Pubblica Amministrazione, 2008).

The evidence provided in this article should be interpreted in light of some limitations that pave the way for future research directions. The main limitation of our study lies in construct validity. As we simultaneously manipulated both the enforcement/violation of an announced rule and the distribution of the incentives, we do not know whether the small effects were caused by the former or by the latter. In addition, as is common with laboratory experiments, our design may be prone to external validity concerns (e.g. Walker et al., 2017). For instance, students voluntarily participating in our lab experiment may not be representative of public sector workers. However, there is wide evidence suggesting that students' experimental responses are largely equivalent to those of other individuals (Anderson and Edwards, 2015). As another concern, especially with respect to the mediation analysis, the study might be under-powered, as experimental groups of the parallel-encouragement design were made of 32 students only. The low power of the study might increase the risk of committing a type II error – that is, failing to reject the null hypothesis when this is false. As to fairness perceptions, we opted to measure them after the last round to avoid priming subjects about our research questions. Moreover, measuring the same construct midway through the experimental procedure and at the end would have been risky owing to the testing threat and the anchoring bias. Finally, the experimental tasks that students in our sample completed may not be representative of all the real-world activities that public employees carry out, some of which may be less measurable than in our lab setting. However, concerns about the lack of publicness in our experimental tasks may be partially mitigated by extant evidence suggesting that numerous tasks are equal across sectors and not sector-specific (Christensen and Wright, 2011; Rainey and Bozeman, 2000). Nonetheless, replication studies may test our findings across different treatments, samples, and contexts.

Declaration of conflicting interests

Funding

ORCID iDs

Paola Cantarelli [iD] https://orcid.org/0000-0002-9914-2548
Paul Battaglio [iD] https://orcid.org/0000-0003-0826-912X

Supplemental material

Supplemental appendices for this article are available online at http://journals.sagepub.com/home/ras

Notes

1. Legislative Decree 150/2009, which introduced the merit pay scheme, was followed by another decree a few months later, according to which the merit pay scheme was frozen. This was mainly owing to the European debt crisis and the urgent need for resources of the government.
2. All the figures and Appendix 1 are published online; see https://journals.sagepub.com/home/ras.

References

Adams JS (1965) Inequity in social exchange. In: Berkowitz L (ed.) *Advances in Experimental and Social Psychology*. Orlando, FL: Academic Press, pp.267–299.

Anderson DM and Edwards BC (2015) Unfulfilled promise: Laboratory experiments in public management research. *Public Management Review* 17(10): 1518–1542.

Argyris C (1960) *Understanding Organizational Behavior (Dorsey Series in Anthropology and Sociology)*. Homewood, IL: Dorsey Press.

Ariely D, Bracha A and Meier S (2009) Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review* 99(1): 544–555.

Bae KB (2021) The effect of pay for performance on work attitudes in the private, public, and non-profit sectors: A panel study from South Korea. *International Review of Administrative Sciences*. Epub ahead of print 26 July. DOI: 00208523211027329.

Baron RM and Kenny DA (1986) The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51(6): 1173.

Battaglio RP (2014) *Public Human Resource Management: Strategies and Practices in the 21st Century*. Thousand Oaks, CA: SAGE/CQ Press.

Belle N (2015) Performance-related pay and the crowding out of motivation in the public sector: A randomized field experiment. *Public Administration Review* 75(2): 230–241.

Belle N and Cantarelli P (2015) Monetary incentives, motivation, and job effort in the public sector: An experimental study with Italian government executives. *Review of Public Personnel Administration* 35(2): 99–123.

Bonner SE, Hastie R, Sprinkle GB, et al. (2000) A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research* 12(1): 19–64.

Brown S and Sessions JG (2003) Earnings, education, and fixed-term contracts. *Scottish Journal of Political Economy* 50(4): 492–506.

Choi S and Whitford AB (2017) Employee satisfaction in agencies with merit-based pay: Differential effects for three measures. *International Public Management Journal* 20(3): 442–466.

Christensen RK and Wright BE (2011) The effects of public service motivation on job choice decisions: Disentangling the contributions of person-organization fit and person-job fit. *Journal of Public Administration Research and Theory* 21(4): 723–743.

Condrey SE and Battaglio RP (2007) A return to spoils? Revisiting radical civil service reform in the United States. *Public Administration Review* 67(3): 424–436.

Corte dei Conti (2003) Sentenza 44/2003, Sezione 2. Roma.

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

Egger-Peitler I, Hammerschmid G and Meyer R (2007) Motivation, identification, and incentive preferences as issues for modernization and HR strategies in local government—First evidence from Austria. In: Annual conference of the EGPA, Madrid, Spain, 2007, pp. 1–22: Public personnel policies study group.

Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2): 171–178.

Forest V (2008) Performance-related pay and work motivation: Theoretical and empirical perspectives for the French civil service. *International Review of Administrative Sciences* 74(2): 325–339.

Fryer RGJr, Levitt SD, List J, et al. (2012) *Enhancing the Efficacy of Teacher Incentives Through Loss Aversion: A Field Experiment (No. w18237)*. Cambridge, MA: National Bureau of Economic Research.

Gabris GT and Ihrke DM (2000) Improving employee acceptance toward performance appraisal and merit pay systems: The role of leadership credibility. *Review of Public Personnel Administration* 20(1): 41–53.

Gneezy U and Rustichini A (2000) Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115(3): 791–810.

Godard J (2001) High performance and the transformation of work? The implications of alternative work practices for the experience and outcomes of work. *Industrial and Labor Relations Review* 54(4): 776–805.

Green C and Heywood JS (2008) Does performance pay increase job satisfaction? *Economica* 75(300): 710–728.

Grossman SJ and Hart OD (1986) The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy* 94(4): 691–719.

Gupta N and Shaw JD (2014) Employee compensation: The neglected area of HRM research. *Human Resource Management Review* 24(1): 1–4.

Hart O and Moore J (1988) Incomplete contracts and renegotiation. *Econometrica: Journal of the Econometric Society* 56(4): 755–785.

Imai K, Tingley D and Yamamoto T (2013) Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176(1): 5–51.

Kellough J and Nigro L (2002) Pay for performance in Georgia state government: Employee perspectives on GeorgiaGain after 5 years. *Review of Public Personnel Administration* 22(2): 146–166.

Keraudren P (1994) The introduction of performance related pay in the British civil service (1982–88): A cultural perspective. *International Review of Administrative Sciences* 60(1): 23–36.

Lah TJ and Perry JL (2008) The diffusion of the civil service reform act of 1978 in OECD countries A tale of two paths to reform. *Review of Public Personnel Administration* 28(3): 282–299.

Marsden D and Richardson R (1994) Performing for pay? The effects of 'merit pay' on motivation in a public service. *British Journal of Industrial Relations* 32(2): 243–261.

Ministero per la Pubblica Amministrazione (2008) Circolare n. 7/08 in merito all'applicazione dell'articolo 71 del D.L. n. 112/08. Roma.

Munyon TP, Jacobs JD, Carnes AM, et al. (2016) Pay-for-politics: Considering the variable compensation-organizational politics relationship. In: Vigoda-Gadot E and Drory A (eds) *Organizational Politics*. Cheltenham, UK: Edward Elgar Publishing, pp.115–147.

Perry JL (1986) Merit pay in the public sector: The case for a failure of theory. *Review of Public Personnel Administration* 7(1): 57–69.

Perry JL, Engbers TA and Jun SY (2009) Back to the future? Performance-related pay, empirical research, and the perils of persistence. *Public Administration Review* 69(1): 39–51.

Perry JL, Mesch D and Paarlberg L (2006) Motivating employees in a new governance era: The performance paradigm revisited. *Public Administration Review* 66(4): 505–514.

Pfeffer J (1998) Six dangerous myths about pay. *Harvard Business Review* 76(3): 108–120.

Rainey HG (2009) *Understanding and Managing Public Organizations*. San Francisco: Jossey-Bass.

Rainey HG and Bozeman B (2000) Comparing public and private organizations: Empirical research and the power of the a priori. *Journal of Public Administration Research and Theory* 10(2): 447–470.

Robinson SL and Rousseau DM (1994) Violating the psychological contract: Not the exception but the norm. *Journal of Organizational Behavior* 15(3): 245–259.

Rousseau DM (1989) Psychological and implied contracts in organizations. *Employee Responsibilities and Rights Journal* 2(2): 121–139.

Rousseau DM (2001) Schema, promise and mutuality: The building blocks of the psychological contract. *Journal of Occupational and Organizational Psychology* 74(4): 511–541.

Schermerhorn JR, Hunt JG and Osborn RN (2004) *Core Concepts of Organizational Behavior*. Hoboken, NJ: John Wiley & Sons Inc.

Starmans C, Sheskin M and Bloom P (2017) Why people prefer unequal societies. *Nature Human Behaviour* 1(4): 1–7.

Tingley D, Yamamoto T, Hirose K, et al. (2014) Mediation: R package for causal mediation analysis. *Journal of Statistical Software* 59(5): 1–38.

Tummers L (2019) *Behavioral Public Administration*. DOI: 10.1093/acrefore/9780190228637. 013.1443.

United States Office of Personnel Management (OPM) (2018) *Federal Employee Viewpoint Survey*. Washington, DC: OPM.

Van Eerde W and Thierry H (1996) Vroom's expectancy models and work-related criteria: A meta-analysis. *Journal of Applied Psychology* 81(5): 575–586.

Walker RM, James O and Brewer GA (2017) Replication, experiments and knowledge in public management research. *Public Management Review* 19(9): 1221–1234.

Weibel A, Rost K and Osterloh M (2010) Pay for performance in the public sector—benefits and (hidden) costs. *Journal of Public Administration Research and Theory* 20(2): 387–412.

Wenzel AK, Krause TA and Vogel D (2019) Making performance pay work: The impact of transparency, participation, and fairness on controlling perception and intrinsic motivation. *Review of Public Personnel Administration* 39(2): 232–255.

Paolo Belardinelli is a Fellow in Public Policy and Administration at the London School of Economics. His research focuses on behavioral public policy and administration.

Nicola Belle is an Associate Professor at the Sant'Anna School of Advanced Studies, where he works within the Management and Health Laboratory of the Institute of Management and serves as a vice scientific coordinator of the EMbeDS (Economics and Management in the era of Data Science) Department of Excellence. His research focuses on behavioral human resource management. Specifically, he studies how we

can use behavioral science to understand work motivation and improve decision-making in the context of mission-driven organizations.

Paola Cantarelli is an Assistant Professor at Scuola Superiore Sant'Anna (Management and Healthcare Laboratory, Institute of Management and EMbeDS, Pisa, Italy). Her research focuses on behavioral human resource management and work motivation in mission-driven organizations.

Paul Battaglio is Professor of Public and Non-profit Management at the Public and Nonprofit Management Program, School of Economic, Political, and Policy Sciences at the University of Texas at Dallas and Affiliate Professor in the Institute of Management and EMbeDS of the Scuola Superiore Sant'Anna. His research interests include public human resource management, organization theory, behavioral public administration, comparative public policy, and research methods.