

Pay Flexibility and Government Performance

A Multicountry Study



THE WORLD BANK

June 2014

Contents

Acknowledgments	iii
Executive Summary	v
Introduction	1
The Theoretical Framework	5
Performance-related pay	5
Differentiation	6
Hypotheses	7
Context matters	8
Assessing the Evidence: Review of Literature on Performance-Related Pay	13
The Case Studies: Pay Flexibility Schemes	20
Performance-related pay	21
Differentiation	26
Assessing the Evidence from the Case Studies: Performance-Related Pay	30
Direct pay flexibility levers	30
Indirect pay flexibility levers	36
Assessing the Evidence from the Case Studies: Pay Differentiation	40
Direct pay flexibility levers	40
Indirect pay flexibility levers	42
Summary of Empirical Analysis	44
Fiscal impact	44
The degree of delegated authority	44
Conclusions and Policy Recommendations	47
How did the hypotheses hold up?	47
Implications for pay policy in developing countries	49
Notes	53
References	55

Boxes

1	Arguments for Long-Term Career-Based Incentives Rather Than Short-Term Performance-Based Incentives for Public Sector Staff	7
2	The “Effort Bargain”	8
3	Active Management and Managerial Effort	9

Figures

1	The Link between Pay Flexibility and Performance	9
2	Aggregate Findings on PRP	14
3	Findings by Research Quality (High Internal and External Validity Studies Only)	14
4	Findings for Craft and Coping Jobs	15
5	Numbers of Studies and Their Findings	18
6	PRP for Selected Careers in Brazil’s Federal Government	22
7	Minas Gerais PRP Bonus Size Increases over Time Based on Consistency of Performance	23
8	PRP as a Percentage of Total Pay in the Philippines	24
9	Pay Differentiation in Indonesia	27
10	Pay Differentiation in Brazil	28
11	Pay Differentiation in the Russian Federation	29
12	Perceptions of Effort in the BR Agencies in Indonesia	31
13	Results of the Performance Evaluation of Permanent Staff in Minas Gerais	32
14	Perceptions of Attendance in BR and Non-BR Agencies in Indonesia	33
15	Perceptions of PRP and Individual Performance Appraisals, Effort, and Motivation in Coping Jobs in the Philippines	34
16	Overall View among Civil Servants of the PRP Scheme in Coping Jobs in the Philippines	35
17	Perceptions of the Effect of PRP on Goal Setting, Monitoring of Targets, Teamwork, and Performance Appraisals in the Philippines	39
18	Perceptions of Effort, Morale, and Quality of Recruits between BR and Non-BR Agencies in Indonesia	41
19	Perceptions of Pay Inequity across Ministries and Agencies and within BR Ministries in Indonesia	42
20	Staff Perceptions of Management and Willingness to Accept Restructuring in Indonesia	43
21	The Link between Pay Flexibility and Performance	47

Tables

1	Hypotheses on the Impact of Pay Flexibility on Civil Service Performance	10
2	Five Key Design Elements of PRP Schemes	10
3	James Q. Wilson’s Classification of Job Types	11
4	The Studies Reviewed	13
5	Key Design Features of PRP in Case Study Countries	25
6	A Sample of Performance Indicators for the Group-Based Bonus Scheme in the Philippines	35
7	Summary of Evidence about the Impact of Pay Flexibility on Civil Service Performance	45

Acknowledgments

This report was prepared by a multicountry team of World Bank staff and staff from other international and national institutions. The lead authors are Zahid Hasnain and Nick Manning (World Bank), drawing on case studies from Chile by Mariano Lafuente (Inter-American Development Bank); Thailand and Malaysia by Willy McCourt (World Bank); Brazil by Barbara Nunberg (Columbia University) and Renato Brizzi (consultant); the Russian Federation by Maya Gusarova (World Bank) and Nikolay Klishch (National Research University Higher School of Economics); the Republic of Korea by Pan Suk Kim (Yonsei University); Indonesia by Zahid Hasnain; and the Philippines by Zahid Hasnain and Sheheryar Banuri (World Bank). It also reflects insights from surveys of government officials in Indonesia by Phil Keefer (World Bank) and Sheheryar Banuri and the

Philippines by Zahid Hasnain and Sheheryar Banuri, and a literature review on performance-related pay by Zahid Hasnain, Nick Manning, and Jan Pierskalla (Ohio State University).

The team is grateful to government officials from Korea, Indonesia, the Philippines, Malaysia, Thailand, Chile, Brazil, and Russia for their cooperation in sharing their views and providing data for the study.

The team is also grateful to a number of World Bank colleagues who helped facilitate data collection and commented on and provided guidance for this work. They include Erwin Ariadharma, Soren Davidsen, Pablo Fajnzylber, Frederico Gil-Sander, Kai Kaiser, Alma Mariano, Yasuhiko Matsuda, Shabih Mohib, Anne Sevilla, Geoffrey Shephard, Staffan Synnerstrom, Maria Tambunan, Rogier van den Brink, and Tony Verheijen.

Executive Summary

This study examines whether financial incentives through pay flexibility can improve the performance of staff in government bureaucracies. Its main messages for pay policy are:

- Pay flexibility can improve performance.
- Pay flexibility works most strikingly in changing managerial behavior.
- Improving public sector performance does not need to wait for systematic pay rationalization or pay simplification throughout government.
- Pay flexibility can work with rather than instead of long-term career incentives.
- The strategy and implementation of pay flexibility reforms must take into account the extent of fragmentation and complexity of the existing public sector pay structure in the country.

About the study

Pay flexibility is defined in the study as:

- *Performance-related pay (PRP)*: Enabling pay to differ for workers doing the same job by linking a portion of pay to the achievement of performance targets.
- *Differentiation*: Differences in pay between apparently similar workers across agencies, career groups, and geographical locations that reflect the need to compete for specific skills in the labor market.

And performance is measured by:

- Improvements in the quality of staff in a government ministry or agency through recruiting and retaining more highly skilled and motivated personnel.
- Improvements in individual line employees' "effort." Because measuring effort across

diverse jobs is difficult, it is often necessary to use proxies that reflect the "engagement" of staff (their commitment to the organization's mission and observed willingness to use flexibility in working methods to help deliver a product or service) or "organizational citizenship" (their willingness to provide extra effort to achieve goals).

The study aims to provide policy advice to the range of governments that are considering these reforms. In Organisation for Economic Co-operation and Development (OECD) countries, pay flexibility is a departure from the archetypical single pay scale, which has been found to be too rigid in an increasingly dynamic labor market and unable to meet performance objectives under fiscal constraints. By contrast, in emerging market and middle-income countries pay flexibility was either part of a move to liberalize a rigid, centralized pay model or introduced on top of an already complex pay regime in an attempt to formalize a haphazard structure that had evolved with little regard to the larger fiscal or incentive impact.

Drawing on the large theoretical literature on PRP, and on the much more limited one on differentiation, the study identifies two ways that pay flexibility can potentially improve performance:

- A direct effect of the financial incentive on staff behavior and quality, which we call the *direct pay flexibility lever*.
- An indirect effect through more effective management that in turn improves staff behavior, which we call the *indirect pay flexibility lever*.

Pay flexibility can potentially improve performance directly through financial incentives and indirectly through improved management

The primary question in this study is whether pay flexibility contributes to these changed methods of working and relationships to the job in ways that have a positive impact on outputs and outcomes

We used the literature to distill specific hypotheses on the possible impacts of PRP and of differentiation:

- *PRP hypotheses:*
 - Hypothesis 1. PRP as a direct pay flexibility lever:
 - Hypothesis 1a. PRP can have a direct incentive effect by improving individual engagement and organizational citizenship and by inducing staff to exert more effort to achieve the outputs and outcomes linked to the incentive.
 - Hypothesis 1b. PRP can have a direct incentive effect on staff quality, resulting in the recruitment and retention of more qualified staff who are likely to do well under the PRP scheme (sorting).
 - Hypothesis 2. PRP can act as an indirect lever by providing incentives for more effective management, resulting in improvements in the performance dialogue with staff—organizational goal setting, and teamwork toward achieving organizational goals and linking individual performance appraisals to those organizational goals.
- *Differentiation-based hypotheses:*
 - Hypothesis 3. Differentiation can act as a direct lever through the recruitment and retention of better-quality staff for priority activities (sorting).
 - Hypothesis 4. Differentiation can act indirectly by providing incentives for greater effort by managers, pushing them to improve the performance dialogue with staff and to increase efforts by line staff.

The evidence used to examine these hypotheses is for the most part self-reported perceptions of changed behaviors and falls short of the ideal of actual measures of outputs and outcomes. Wherever data on outputs (such as revenue collection, fines for nonfilers of taxes, and teacher attendance) and even outcomes (such as student test scores) were available, they were used in the study. However, this report focuses on the core administration, for which comprehensive measures of

performance are rarely available, and information on inputs, behaviors, and processes can usually be indirectly imputed only through staff perceptions. The primary question in this study is whether pay flexibility contributes to these changed methods of working and relationships to the job in ways that have a positive impact on outputs and outcomes.

Each of these hypotheses is hotly debated in the literature, with several arguments against both PRP and differentiation. The main points are that PRP can cause many perverse consequences when performance cannot be accurately measured, that PRP can crowd out intrinsic motivation, and that the pay inequity that both PRP and differentiation create hurts staff morale, effort, and teamwork. The study explores whether and in what contexts these hypotheses hold. Two key contextual variables are the design features of the PRP plan and the nature of the public sector job in which PRP is implemented. The relevant characteristics of the job (based on James Q. Wilson's typology) are whether the job outputs are easily observable and measurable to managers and external agents (*craft jobs*, such as teaching, health care, and revenue administration) or not (*cop-ing jobs*, such as finance, planning, and other core policy functions). The five PRP design elements are whether the incentive is individual or group based, the time horizon of the incentive, the nature of the performance evaluation, the size of the incentive, and the probability of receiving the incentive.

The study evaluates existing evidence through a comprehensive literature review of PRP. It then examines pay flexibility for the core public administration in emerging market countries through case studies of PRP and differentiation in Brazil, Chile, Indonesia, Korea, Malaysia, the Philippines, Russia, and Thailand. Data in the case studies were collected through structured interviews with experts and, in Indonesia and the Philippines, large representative surveys of government officials.

Findings from the literature review on PRP

The literature review, unlike other reviews of PRP, disaggregates the evidence by the quality

of the empirical study, the different types of public sector jobs based on the craft and coping classification, and the country context. The main findings: 93 of the 153 studies show some form of positive effect of PRP, and 65 of the 110 studies of craft and coping jobs find positive effects. Limiting the analysis to high-quality studies of craft and coping jobs, 37 of 53 report positive results. (The evidence is overwhelmingly for craft jobs, since there are only three high-quality studies of coping jobs, and there are no high-quality studies of coping jobs in developing countries.) Most of the literature has explored the effect of PRP on staff effort, with only a few studies examining the impact on staff quality through sorting, and even fewer exploring the impact of indirect pay flexibility levers.

Overall, the literature review finds support for Hypothesis 1a for jobs such as teaching, health care, revenue administration, and job placement, which have more measurable outputs and outcomes. There is more limited support for Hypothesis 1b of PRP improving sorting. Context is important, with more positive findings in developing countries, especially in teaching. Several studies identified problems of unintended consequences, or “gaming” of the incentive program. There is not enough evidence for Hypothesis 1 and Hypothesis 2—the direct incentive effect and the indirect incentive effect through improved management—in the core civil service in non-OECD countries.

Findings from the case studies: PRP

Brazil, Chile, Korea, Malaysia, the Philippines, Thailand, and, to a limited extent, Indonesia have implemented PRP. These PRP plans display a high degree of diversity in the five main design features, and all these countries have plans that apply to both craft and coping jobs.

The case studies reinforce support for Hypothesis 1 for craft jobs. In Brazil, especially in the state of Minas Gerais, the PRP scheme led to higher productivity in the police and the revenue authority, as measured by weapons seizures, number of police operations, and revenue collection. Notably, the performance targets for the revenue authority evolved as

frontline managers became aware of the perverse incentives that a myopic focus on only revenue collection could generate and later included taxpayer facilitation to ensure sustainability of revenue collection over the long run. There was also some evidence of a positive sorting effect in Minas Gerais. In the World Bank survey of government officials in Indonesia, the revenue agency scored the highest on questions gauging effort and staff engagement. In Chile, PRP was viewed quite favorably in the revenue authority and the civil registry. In Malaysia the only positive reports of PRP were from the revenue authority.

By contrast, there was little support for Hypothesis 1 for coping jobs, and in some cases evidence of negative effects. By definition, these are jobs for which outputs cannot be easily measured, so the performance evaluations that form the basis of the financial incentive are either based on subjective evaluations by supervisors and review panels or on quantitative input or process measures. Subjective appraisals put the onus on managers to credibly distinguish among staff, difficult in most bureaucracies. The tendency in Brazil, Chile, and Thailand was for the vast majority of staff to be given a best or next best performance rating. As a result, the performance bonus was given out with close to a probability of 1 and therefore became a de facto salary supplement that could not have a direct incentive effect. Where quantitative measures were used, they tended to be heavily process-oriented, increasing the risks of gaming behavior.

However, evidence from Brazil and Indonesia revealed that PRP had an impact on the extremes of the performance distribution by reducing staff absenteeism and helping discipline blatantly incompetent staff. Outlier staff members were fairly easy to identify, so managers could credibly sanction them.

Korea, Malaysia, and the Philippines tried to counter the tendency toward uniformly high performance ratings and an equal distribution of the performance bonus by mandating a forced distribution of performance ratings. This risky policy can harm staff morale, and its efficacy depends on the level

The literature review and the case studies find that PRP can increase productivity in teaching, health care, and revenue administration

In Indonesia differentiation helped gain buy-in from staff for restructuring and pushed management to improve recruitment procedures and performance appraisals

of trust in an organization and the legitimacy of performance appraisals among staff. In Korea the plan has been generally well received. For example, when given the choice between individually differentiated incentives and equal pay for all within a group, 32 of 44 government agencies opted for the former and none for the latter (with the others choosing some form of mixed option). By contrast, in the Philippines, which has a combination of individual- and group-based incentives from a mandated ranking of both working units and individuals within those units, the survey revealed little direct effect on staff effort. Instead, the survey found a divergence of views, depending on the performance ranking of the respondent, the effectiveness of the individual performance appraisal process, the transparency of individual performance ratings, and the impact of the incentive on staff morale.

The case studies found support for Hypothesis 2: PRP did improve performance through the indirect route of better management, specifically through setting better goals and encouraging greater teamwork. In the Philippines the World Bank survey showed that staff believed that the performance bonus scheme had motivated management to focus more on organizational target setting, monitoring the accomplishment of those targets, engaging staff in this process, and fostering greater teamwork and collaboration among staff. Overall, staff were positive about the PRP scheme, despite the lack of credibility of the individual performance appraisal process, suggesting that the direct and indirect pay flexibility levers were having opposite effects on individual productivity and organizational citizenship. In Minas Gerais, staff believed that PRP was integral to the performance agenda in the state and helped clarify expectations and individual goals and targets, broadly complemented results-based management reforms, and encouraged greater delegation of human resource management authority from the central personnel office to line ministries and agencies. Similar positive management effects were noted in Korea and, to a more limited extent, in Malaysia.

Findings from the case studies: Differentiation

The case studies found strong support for Hypothesis 3. Evidence from Brazil, Chile, Indonesia, and, to a lesser extent, Russia—all of which have significant differentiation—reveals that the higher paid agencies or groups of staff were better able to attract and retain high-quality staff. This increase in staff quality did not appear to come at the expense of demotivation through greater pay inequities for other groups. As the Indonesia survey showed, pay inequity has been the norm in many of these countries, and the added inequity from differentiation may be only marginally more demotivating for staff in the less privileged agencies. Differentiation needs to be limited, however, to a few priority groups of staff. Without these constraints, as the recent experience of Brazil suggests, differentiation can create disruptive competition among agencies over salary increases and discourage inter-agency cooperation.

There was also some support for Hypothesis 4. Data from Indonesia suggest that differentiation can spur other organizational reforms, where it helped gain buy-in from staff for restructuring and pushed management to improve recruitment procedures and performance appraisals.

Conclusions and policy recommendations

Overall, six messages can be drawn more generally for pay policy, particularly for the more challenging low-income country contexts in Africa, South Asia, and elsewhere. First, pay flexibility can improve performance, an important finding given the general skepticism in the public administration literature on this topic. These reforms are not a silver bullet, and they involve tradeoffs and risks. Poorly designed pay flexibility plans can cause considerable harm by encouraging perverse behavior and unintended consequences, and some task reallocation and gaming behavior should be expected. With proper monitoring, these tradeoffs can be managed. At a minimum, the study suggests that pay flexibility plans be implemented in craft jobs with a monitoring

regime in place to detect and respond to gaming behavior.

Second, pay flexibility works most strikingly in changing managerial behavior, which has several implications. The PRP scheme designed for coping jobs should encourage these management changes. A large group-based bonus PRP scheme is preferable, despite the difficulty in establishing unit-level performance targets, because it bypasses the problem of distinguishing between individuals' performance and puts the spotlight on management improvements as the key enabler linking PRP to better performance. Pay flexibility does not need to be introduced across the board—it should be introduced asymmetrically where there is some basic minimum level of managerial competence and should complement performance budgeting and other management reforms. In terms of sequencing, PRP should come after initiation of results-based management so that it has some managerial resources to build on, such as performance indicators and progress review mechanisms, and provides incentives to frontline managers to better use these resources.

Third, improving public sector performance does not need to wait for systematic pay rationalization or pay simplification throughout government. “Purposeful” pay complexity through pay flexibility can improve performance at the margin even when layered on top of a chaotic pay regime and poor human resource management. This is an encouraging finding given the technical and political challenges of comprehensive pay rationalization and the poor track record of such reforms.

Fourth, pay flexibility can work with rather than instead of long-term career incentives. The PRP scheme in Minas Gerais is a good example of complementarity between short- and long-term incentives, since the size of the incentive increased with sustained good performance.

Fifth, the strategy and implementation of pay flexibility reforms must take into account the extent of fragmentation and complexity

of the country's existing public sector pay structure. Although pay simplification is not a necessary prior step before introducing pay flexibility, the extent of “messiness” of the pay regime has implications for the pay flexibility strategy. In simpler systems, ambitious, across-the-board pay flexibility reforms pose less risk if there is explicit recognition of possible perverse behavior and unintended consequences and if experimentation and learning-as-you-go are built into the plan. Ideally, even in these systems pay flexibility would be initially restricted to a few organizations to limit the administrative burden of the validation and monitoring systems necessary for effective pay flexibility. In complex systems there is the risk that pay flexibility could degenerate into yet another element of the messy pay regime with few productivity gains and a further weakening of central fiscal control and management coherence. This risk can be mitigated by limiting pay flexibility to a select few high-priority organizational “islands,” which are chosen either because they are the highest priority or because they are managed relatively well. These are the staff whose productivity improvements are considered to be the most important for government performance.

And sixth, many questions remain, and much more research is needed. The evidence this study has drawn on primarily concerns self-reported perceptions of processes and behavior change and not improvements in actual outputs and outcomes, and so causal claims cannot be made. The impact of pay flexibility reforms varies considerably based on contextual factors that go beyond the two factors—the type of public sector job and the design features of the pay flexibility scheme—that this study has looked at. How pay flexibility interacts with existing formal and informal rules, culture, and the institutional arrangements and capacity within government to collect and validate data on performance and to coordinate pay flexibility across the public sector are all key issues that require more investigation.

Pay flexibility works most strikingly in changing managerial behavior, which has implications for the design of pay flexibility schemes

Introduction

An effective and efficient government is a central objective for countries at all levels of economic development. Government bureaucracies are very different from their private sector counterparts because they are bound by a legal framework and subject to influences from external actors that create a set of incentives in which a performance orientation, in the sense of managers and staff working purposefully to achieve an outcome, may be a minor objective. In some low- and middle-income countries the public sector is seen more as an avenue for political patronage and employment than as the executor of government policies aimed at growth, poverty reduction, and delivery of essential services. Yet there are few alternatives to government's extensive role in these areas and public sector reform remains a key development challenge.

This study examines whether financial incentives through pay reform, specifically pay flexibility, can improve the performance of public bureaucracies. It asks this question with the objective of providing policy advice to the range of governments, particularly in emerging market and low- and middle-income countries, that have introduced, or aim to introduce, these reforms. It is therefore a deliberately narrow, but focused and empirically detailed, analysis of one factor in the set of variables in human resource management—recruitment, promotion, training, and management practices—that impact the incentives of individuals in organizations. Within public bureaucracies, the study focuses largely on the “civil service” or “core public administration,” which is normally defined as the civilian

public administration excluding key service delivery staff such as teachers, medical personnel, and police. These limitations are in part to keep the analytics manageable and in part to home in on one of the more dynamic areas of human resource management in both OECD and developing countries over the past two decades.

This focus on the civil service is important for three reasons. First, the core administration is the conduit of political authority—public servants in central agencies can claim, explicitly or implicitly, a reform mandate from those politicians with whom they work day to day. Second, civil service pay is often the implicit standard for pay across the public sector. And third, the center sets the rules of the game for the wider public sector as well as for the private sector. For example, the public expenditure and financial accountability arrangements created by the center (budget, accounting, and audit) determine how public and private schools and hospitals providing services on behalf of the government are held accountable. The rules and procedures set by the center determine how publicly owned banks, harbors, and airports, bodies that are outside general government, are regulated and how their contingent liabilities are to be managed. They thus shape the procurement arrangements that determine how many services are obtained.

What is “performance” and how should it be measured?

The study takes the simple view that organizational performance can be assessed through

The center sets the rules of the game for the wider public sector as well as for the private sector

The study assesses performance primarily through behavioral changes among staff in organizations

the following behavioral changes among the staff of an organization:

- Improvements in the quality of employees in a government ministry or agency through the recruitment and retention of more skilled personnel.
- Improvements in employees' "effort." Because measuring effort across diverse jobs is difficult, it is often necessary to use proxies concerning the "engagement" of staff (their commitment to the organization's mission and observed willingness to use flexibility in working methods to help deliver a product or an output) or "organizational citizenship" (their willingness to provide extra effort to achieve goals).¹

The evidence for these behavioral changes in the study is largely based on self-reported perceptions of staff and falls short of the ideal of actual measures of outputs and outcomes. Wherever data on outputs (for example, revenue collection, fines for nonfilers of taxes, and teacher attendance) and even outcomes (for example, student test scores) are available they are used. However, since the focus of this report is the core administration, such comprehensive measures of performance are rarely available, and information on inputs, behaviors, and processes can often only be indirectly imputed through staff perceptions. The primary question addressed in this study is whether pay flexibility contributes to these changed methods of working and relationships to the job in ways that are considered in the current literature to have a positive impact on outputs and outcomes.

What is pay flexibility?

Pay flexibility as an aspect of public sector pay policy is defined in this study as a departure from the traditional civil service pay model that emerged in the process of nation-state formation and the creation of modern bureaucracies in the nineteenth and early twentieth centuries (Ketelaar, Manning, and Turkisch 2007). As captured in (although not prescribed by) the writings of Max Weber, a modern bureaucracy acts, in principle, as a lasting, impartial, rule-abiding, and nonpartisan executor of laws and regulations, which

are devised by the political leadership (Weber 1978). The accompanying employment and salary systems, intended to allow bureaucrats to fulfill their role with minimal political interference, have often featured a work relationship governed by public not civil law; common and compressed salary scales based on grades and seniority not task; standardized and test-based recruitment; secure tenure; and generous benefits and regular across-the-board salary raises. While large private sector bureaucracies have historically developed similar characteristics, more recent assessments of pay arrangements in private organizations characterize them as having higher pay dispersion, higher average pay for skilled employees, less secure employment, more explicit performance incentives, and greater sensitivity of the individual wage to current supply and demand in the labor market (Eldridge and Palmer 2009; Perry, Engbers, and Jun 2009).

In Organisation for Economic Co-operation and Development (OECD) countries, moves toward pay flexibility in all economic sectors arose in the 1980s in the context of attempts to liberalize labor markets in response to increasing international economic competition. Pay flexibility was usually complemented with flexibility in other aspects of personnel management, such as numerical flexibility (the ability of employers to adjust the number of workers or hours worked to changes in demand); functional flexibility (the ability of employers to reorganize the competencies associated with jobs); and distancing (displacement of employment contracts by contracting out noncore tasks). In the public sector, traditional pay arrangements were, it was argued, unable to ensure that performance objectives were met within fiscal constraints, and many OECD countries began to move from the archetypically single pay scale toward more flexible pay arrangements that have "pay for performance" and "pay for skills," and not "pay for seniority," as their defining feature.

Pay flexibility takes place within constraints. The level and structure of pay needs to be set so that it maintains long-term fiscal sustainability and ensures that the wage bill does not crowd out other essential government

expenditures. Affordability constrains the extent to which compensation can be used as a lever to improve performance.

Although developing countries have not made substantial formal moves toward pay flexibility, in practice a somewhat ad hoc flexibility has emerged. In many countries the civil service pay structure is far from the Weberian model, encompassing a variety of allowances and salary supplements. Pay as a result varies considerably between individuals performing similar tasks based on individual circumstances. This ad hoc pay flexibility is in part a cumulative result of uncoordinated measures taken over time to improve the performance of, or to respond to lobbying by, particular groups of staff—for example, in salary increases through special allowances for service delivery staff or revenue officials. In these cases, pay flexibility is in part an outcome of political circumstance, such as bargaining between different employee unions and central authorities. Attempts to instill a mission orientation in these public administrations have not been very successful. One approach, often favored by the World Bank and other development partners, of trying to rationalize pay through the traditional single pay scale or a limited number of pay scales has a poor track record (Independent Evaluation Group 2008). Many developing countries, seeking to emulate OECD countries but also trying to improve the quality and motivation of staff, have been experimenting with linking pay to performance and providing higher pay to select groups of staff that are deemed particularly important.

This study defines pay flexibility as comprising two key design elements, which can be present in varying degrees. These elements are:

- *Performance-related pay (PRP)*:² Enabling pay to differ for workers doing the same job by linking a portion of their pay to the achievement of performance targets. How performance is measured, who measures it, and how it is linked to salary can all vary considerably. Performance can be based on qualitative assessments or quantitative measures of inputs, outputs, or outcomes, and assessed by direct supervisors, human

resource specialists, peer panels, or outside agencies. The financial incentive can be a combination of base pay and one-off bonuses or merit increases of base pay, and it can be awarded on an individual, small team, or larger departmental basis.

- *Differentiation*: Differences in pay between apparently similar workers across agencies, career groups, and locations that are primarily a function of the specific skills that the agency competes for in the labor market and of labor costs and the cost of living in the localities where agencies operate. Differentiation can affect all staff paid by the organization, with the result that there are agency-specific pay scales; particular occupational groups or cadres; specific individuals with scarce but vital skills; and specific locations entailing particular hardship.

Pay flexibility reforms are always framed in terms of improving performance, but they can be an element of two quite distinct strategies: to liberalize a rigid and centrally driven pay model, or to formalize an already complex haphazard or asymmetric structure that has evolved with little regard for the larger fiscal or incentive impact. The first strategy introduces asymmetry and complexity within the pay system in contrast to the apparent simplicity and order of standardized, centrally determined pay scales. The second emphasizes “purposeful complexity” to bring order to an unmanaged approach to pay policy. Both pay flexibility strategies aim to provide incentives for performance improvements for a target group of civil servants or to make pay more competitive for that group. The first is usually done through an ex post assessment of whether the target group of staff delivered on contracted outputs or objectives, the second through an ex ante assessment of the relevant labor market in which the target group competes.

Data sources for the study

The study draws on two sets of data. First is a comprehensive review of 153 studies on PRP that, unlike other reviews, disaggregates the available evidence by the quality of the empirical study; by differences in public sector contexts, particularly the different types of public

In developing countries, in practice a somewhat ad hoc flexibility has emerged

Pay reforms are not a silver bullet to enhance public sector performance

sector jobs; and by country context (developing country or OECD settings). This disaggregation allows for more nuanced conclusions on the impact of performance pay in the public sector.

Second is a set of case studies of OECD and emerging market countries (Brazil, Chile, Indonesia, Korea, Malaysia, the Philippines, Russia, and Thailand) that have undertaken either one or both of these pay flexibility reforms. Data in the case studies were collected through structured interviews with country experts and, in Indonesia and the Philippines, large perception surveys of government officials.³ These case studies, unlike the literature review, focus mainly on pay flexibility in the core civil service. The case studies exclude key service delivery staff such as teachers and doctors, since the literature review analyzes a number of high-quality studies of PRP for teaching and medical staff that provide strong evidence of causal relationships. There are few high-quality studies of PRP in the core civil service, and therefore the case studies focus on this area. There is also no significant literature on pay differentiation, which also is covered in the case studies.

Organization of the study

The next section outlines the theoretical framework, summarizing the main arguments for and against PRP. It also reviews the much smaller literature on pay differentiation. This discussion is then used to distill some hypotheses on the impact of pay flexibility on public sector performance through *direct levers*, which change the incentives of staff, and through *indirect levers*, which change the incentives of frontline managers, in both cases by acting on the incentives of staff to behave differently in their jobs and by changing incentives about joining, or remaining in, the public service. These hypotheses are then examined in the main empirical sections of the study, the review of the empirical literature on PRP, and the review of the findings on pay flexibility from the case studies. The final section summarizes the main findings and offers some policy directions.

This empirical analysis is more an assessment of the general plausibility of these hypotheses than a rigorous empirical test. The strength of this study is the breadth and richness of its contextual coverage, which comes at the expense of empirical depth and its ability to support causal statements. The evidence primarily concerns perceptions of processes and behavior changes and not improvements in actual outputs and outcomes. Much of the evidence from the case studies is anecdotal, based on conversations with government officials and different “experts.” Even the large perception surveys of government officials in Indonesia and Philippines, which are in many ways state of the art in core public administration research, provide little evidence to establish causality. The study is therefore more empirically tenuous than the recent impact evaluation literature on performance incentives in teaching and health care, but it is more rigorous than much of the work to date on pay reform in the core public administration.

It is important to emphasize also that this study is not premised on the notion that pay reforms are a silver bullet to enhance public sector performance. It is improbable that such silver bullets exist. Any reform is about tradeoffs, and whether these tradeoffs are worthwhile is highly context specific. Improving the performance of large bureaucracies is extremely difficult, and the key is to capitalize on small openings to achieve marginal improvements in productivity (World Bank 2012). Can pay flexibility provide such an opening, and are the potential tradeoffs, which are discussed in the next section, worth it in that particular country? The study also reviews the empirical evidence from research that is limited to pay flexibility, assuming that all else is constant; this limitation is both to keep the task manageable and to examine a variable that has a particular relationship with other aspects of pay reform. It cannot reach a conclusion that when pay flexibility “works” it is more important than other variables. These are important issues that merit a different study.

The Theoretical Framework

Performance-related pay

The case for and against PRP has strong theoretical underpinnings in neoclassical economics, organization theory, and behavioral economics. In the microeconomic principal-agent model, the problem that needs to be addressed is that the principal—the manager, the school principal, the hospital administrator—has to induce effort from his or her staff but cannot easily monitor their work. Under these conditions a fixed pay contract gives the employer little leverage to influence employee effort after hiring decisions have been made, a problem that becomes worse if employees are hard to fire. PRP is a means of addressing this problem of moral hazard by tying observable outputs, which are presumably correlated with unobservable effort, to pay.

The principal-agent model suggests that PRP can potentially ameliorate another problem, that of adverse selection. The agent has access to private and valuable information at the time of contract signing and low- and high-skill applicants are hard to distinguish based on public information. Hiring agencies need to offer contracts that induce high-quality applicants to apply and deter low-quality applicants from misrepresenting their qualifications. PRP can alleviate this sorting problem, since higher-quality personnel who expect to perform better under this pay system will be more likely to apply than low-quality applicants.

The model also offers strong arguments against PRP. Such incentive schemes require the ability to measure some relevant results to which pay is linked, and their design should tightly link the agent's actions to these

observable results. Critics have pointed out that these conditions are rare in the public sector. Results are hard to define, let alone measure, particularly in policy or administrative, as opposed to service delivery, organizations. Civil servants often work in large teams under the supervision of multiple managers, complicating the attribution of performance and responsibility of evaluation. A necessary condition for PRP may also be the presence of high levels of trust and transparency between employees and management to avoid arbitrary implementation and worker dissatisfaction (Kellough and Lu 1993). Implementing PRP in the absence of objective measures of results, lack of attribution, and lack of trust can breed resentment among staff and demotivate them, thereby resulting in lower effort.

A related problem is one of perverse incentives or unintended consequences, which come in several variants, the most extensively studied of which is the multitasking problem (Holmstrom and Milgrom 1991). When multiple tasks are performed, giving incentives for only a subset of those that are observable and contractible will not necessarily improve overall outcomes. Instead, employees may shift effort from noncontracted to contracted tasks, which under some circumstances can lead to worse outcomes. For example, the task of teaching can involve both instruction based on sound curricula and coaching on test-taking strategies, and poorly designed incentive schemes can encourage teachers to reallocate effort to the latter and away from the former (known as “teaching to the test”) to the detriment of human capital accumulation. Similarly,

PRP is a topic of considerable debate in the academic literature

PRP can provide frontline managers with incentives to increase awareness among staff of organizational goals by defining explicit performance standards

financial incentives for the provision of school meals to children to reduce malnutrition can result in reduced teaching and worse student learning outcomes.

Closely linked to the multitasking problem is “gaming,” or manipulation of the incentive system, which comes in two forms: manipulation of the data used to measure the performance incentive output, or manipulation of the output itself. Typical examples of the former are the setting of easy-to-reach performance targets and other forms of manipulating data to demonstrate that the targets have been achieved. The latter includes “cream skimming” or “cherry picking”—the deliberate selection of beneficiaries to improve program effects (Heckman, Heinrich, and Smith 1997)—or other forms of manipulation such as the provision of high-calorie food to students during test days (Figlio and Winicki 2005) and ratchet effects under which managers reduce their output increases to a modest increment so that expectations and future targets will be set at a low level.

Finally, an argument coming from organizational theory and behavioral economics is that PRP can reduce the intrinsic motivation that people have in their jobs (Frey and Osterloh 1999). These schemes cause workers to change their perception about organizational goals and values—for example, that the organization’s goals are not about public service but more about private profitability—leading to an overall reduction of effort. This crowding-out of intrinsic motivation can be especially salient if performance pay is introduced using antagonistic framing and can stifle creativity and collaboration.

These problems have prompted many to argue that the approach to performance in the public sector should be based on long-term career-based incentives rather than PRP (box 1).

The pros and cons of PRP have focused largely on the direct effects of financial rewards on individual incentives. Organizational theorists have also pointed to possible indirect effects of PRP even in light of the difficulties associated with measuring outputs amid a multiplicity of tasks. Scholars as diverse as James

Q. Wilson (1989) and Jean Tirole (1994) have recognized that the multiple demands placed on a public organization, and the multiple interests or principals it needs to serve, make it difficult to define a “goal” for the organization that can orient the staff. The introduction of PRP can start the process of goal or mission orientation by triggering improvements in management practices. PRP can provide frontline managers with incentives to increase awareness among staff of organizational goals by defining explicit performance standards, encourage frontline managers to focus more on working with their staff toward achieving these organizational goals and tracking them regularly, and increase the link between individual and organizational goals in individual performance assessments, thereby inculcating a focus on results within the organization (Marsden 2004, 2009; OECD 2005). In sum, PRP works indirectly by institutionalizing regular discussions of performance between management and staff, thereby altering the “effort bargain,” as elaborated in boxes 2 and 3.

Differentiation

The theoretical literature on pay differentiation is much more limited. The main argument for it is that differentiation enables an agency to set pay at a level that is appropriate for a given task in a specific labor market. This is deemed necessary given the considerable pay differentiation in the private sector in which interindustry wage differentials exist even after controlling for differences in human capital of employees (Dickens and Katz 1987; Krueger and Summers 1988; Groshen 1991). Differentiation can also reflect other factors, such as differences in union power and wage bargaining arrangements, sectoral monopoly or oligopoly profits, industry-specific technology shocks and innovations, and managerial approaches. Advocates of pay differentiation in the public sector stress functional similarities and competition with the private sector to motivate mirroring reforms. Overall, pay differentiation across agencies is meant to better reflect the heterogeneity of public services and signal commitment to particular organizational goals to employees and outside observers (Bender and Elliott 2003).

Box

Arguments for Long-Term Career-Based Incentives Rather Than Short-Term Performance-Based Incentives for Public Sector Staff

1

Many argue that in complex public sector environments, with complex and occasionally contradictory objectives and multiple principals, incentives for performance should rely on information that is hard to game over the longer term (Burgess and Metcalfe 1999). There are three sets of long-term career-based incentives: opportunities for long-term enhancement of rewards and status, competitive promotions, and deferred compensation.

The central idea behind long-term career benefits is that workers exert effort in order to influence actual or potential employers' beliefs about their talent, and that while performance information can be gamed in the short term, over time real performance becomes evident. Even when employees are paid a fixed wage, they are thus motivated by the effect their effort will have on future wages (Holmstrom 1982). However, long-term career-based incentives require some cofactors: employees cannot signal talent and effort to employers that fail to look afresh at effort on a regular basis or that are more interested in nonmerit based signals, or if the breadth and complexity of employee tasks are such that it cannot be clear where and whether they are succeeding (Dewatripont, Jewitt, and Tirole 1999a, 1999b).

Competitive promotions have been extensively reviewed within the framework of tournament theory with promotions seen as prizes allocated to the workers who rank higher than all others over a given period. Some evidence suggests that successive rounds of competition for jobs can reveal otherwise hidden performance traits

(Burgess and Metcalfe 1999). Competitive promotions are considered more reliable measures of ability than short-term performance assessments since “[s]enior civil servants are likely to be as motivated by [promotions] as they are financial rewards [since] the incentives to game to achieve reputational rewards are somewhat lower than the incentives to game in relation to financial rewards. This is for the simple reason that if the reward is reputation, a reputation for gaming amongst professional peers undermines the reward itself” (Ketelaar, Manning, and Turkisch 2007, 16). However, early promotions risk distorting the employer's perceptions of fast-rising staff, tending to promote them more automatically. There is also some evidence of diminished cooperation between staff who are in the same pool of candidates for promotion (Lazear 1989).¹

The final long-term career incentive widely used in the public sector is deferred compensation, where upward-sloping wage profiles can be structured to reflect experience and expertise. The argument is that deferred compensation provides incentives to workers to exert effort early in their careers in order to be promoted or not to be fired and hence lose a pay-off later in their tenure (Lazear 1981). However, if there is no serious risk of losing the long-term compensation gains, the rewards are simply provided in exchange for length of tenure or seniority. However, if rewards for seniority are provided together with a credible threat of nonadvancement or dismissal for poor performance, seniority is simply an easy-to-measure proxy for experience and serves to attract risk-averse but talented workers.

Note

1. See also the emphasis given to employment security and recruitment in the seven human resource management practices identified by Pfeffer (1998a, 1998b) as key to organizational effectiveness. These have been validated more widely, though an empirical review of the impact of these practices found no direct relationship between employment insecurity and organizational performance. The review did note that insecurity seemed to hinder development of other useful human resource management practices with a stronger link to performance (Ahmad and Schroeder 2003).

In developing countries the argument for differentiation, also called asymmetric pay reform or agency-level pay reform, is primarily pragmatic. The experience of comprehensive whole-of-government pay reforms in these countries has been disappointing, primarily because they lacked the political support of elites and were resisted by key stakeholders such as employee unions (Lindauer and Nunberg 1994). An agency-based approach may therefore be a more feasible alternative for incremental productivity improvements, with successful implementation creating a demonstration effect across government (Nunberg and Taliervo 2012).

The main argument against pay differentiation is that the departure from centralized and uniform salary scales reduces transparency and equity in civil service remuneration, thereby breeding resentment and demotivating staff. Uniform salary schedules were introduced in

Western bureaucracies in the nineteenth and early twentieth centuries precisely to combat problems of patronage and corruption (Odden and Kelley 2002). Pay differentiation across agencies can also reduce cooperation and complicate coordination across sectors or regions (Rexed and others 2007).

Hypotheses

Performance improvements are characterized in this study as increased effort by line staff, proxied as staff engagement or organizational citizenship, and better staff quality. The review of the literature suggests a number of hypotheses on how pay flexibility can cause these improvements (also illustrated in figure 1):

- *PRP hypotheses:*
 - Hypothesis 1. PRP can act as a direct lever by:
 - Hypothesis 1a. Having a direct incentive effect through increased

Box The “Effort Bargain”

2

“[E]very employment contract ... consists of two elements: (1) an agreement on the wage rate ... i.e. a wage-rate bargain; and (2) an agreement on the work to be done, i.e. an effort bargain” (Behrend 1957, 505). Early work on productivity assumed that the relationship between these two elements was fixed in some mechanical way—without a change in the wage-rate bargain, there would be no change in the effort bargain. Some flexibility could be found to adapt that employment contract, but only “within certain limits” (Coase 1937, 391).

Marsden, in examining why U.K. efforts to introduce PRP in the core civil service during 1980–2000 had failed to improve psychological motivation for staff, argued that the effort bargain could be changed in nuanced ways. “Less visible [than pay adjustments], but just as important for management, is its ability to revise job boundaries, and redefine the nature and standards of performance that it requires from employees. These standards, which may include qualitative aspects of performance, are usually the subject of a tacit understanding between staff and management” (Marsden 2004, 352).

In effect, frontline managers were using the informality that is always present in complex bureaucratic tasks, where “[i]t is common for jobs to deviate considerably from their formal job descriptions” and their inside knowledge of what workers were actually doing, since “[t]he features of a given job are

... accessible to higher management only through the eyes of firstline managers” (Marsden 2004, 352), to move some of the goalposts concerning work procedures and performance expectations. Marsden argued that while research showed that PRP for jobs within the core civil service had little impact on measures of individual motivation, the accompanying regular institutionalized performance discussions improved productivity in the agencies as a whole. Empirical work has highlighted the significant motivational effect of regular communication between frontline managers and their staff (Ahmed and others 2010).

A more recent line of research concerning staff engagement has reached a similar conclusion. There is a close link between high levels of employee engagement and positive discretionary behavior or effort. The empirical literature tends to come from the private sector and is largely from developed countries, but there is an emerging literature concerning middle-income countries (Kular and others 2008). Engagement improves affective commitment (“I like this place”) and normative commitment (“I should stay”), which are both positively behavior changing.¹ Purcell and others (2003a) found that such commitment was affected by how large-scale objectives were explained and interpreted day to day: “[T]he vital ingredient in linking people management to business performance ... is primarily the task of line managers” (Purcell and others 2003b, 72).

Note

1. Rafferty and others (2005) review the literature and find empirical associations between commitment and increased job satisfaction (Vandenberg and Lance 1992); increased job performance (Mathieu and Zajac 1990); higher sales (Barber, Hayday, and Bevan 1999); lower employee turnover (Cohen 1991); less intention to leave (Cohen 1993; Balfour and Wechsler 1996); and lower absenteeism (Cohen 1993; Barber, Hayday, and Bevan 1999).

individual engagement and organizational citizenship and by inducing staff to exert more effort toward achieving outputs and outcomes linked to the incentive.

- Hypothesis 1b. Having a direct incentive effect on staff quality, resulting in the recruitment of higher quality staff who are likely to do well under the scheme (sorting).
- Hypothesis 2. PRP can act as an indirect lever by providing incentives for greater effort by managers on the performance dialogue, resulting in better organizational goal setting, teamwork toward achieving organizational goals, and linkage of individual performance appraisals to those organizational goals. This improved performance dialogue in turn results in greater effort by line staff.
- *Differentiation-based hypotheses:*
 - Hypothesis 3. Differentiation can act as a direct lever by resulting in the

recruitment and retention of better-quality staff for priority activities (sorting).

- Hypothesis 4. Differentiation can act indirectly by providing incentives for greater effort by managers, putting the spotlight on management to improve the performance dialogue with staff. This improved performance dialogue in turn results in greater effort by line staff.

These potential direct incentive effects on productivity, and indirect effects through changed management practices, are hotly debated. Table 1 summarizes the arguments for and against them.

Context matters

The main objective of the empirical analysis is to examine whether and under what contexts these hypotheses hold. This study considers three contextual variables in particular: the *nature of the PRP scheme*; the *nature of the job* for which the scheme is being implemented; and the *degree of delegated authority* that agencies have to manage

Box Active Management and Managerial Effort

3

PRP encourages frontline managers to institutionalize performance discussions with staff, thereby renegotiating the “effort bargain.” The logical next questions are: What flexibilities can managers use to propose to staff in this renegotiation? And why should frontline managers themselves be motivated to undertake such a renegotiation, which is, after all, an additional effort for them?

In considering the first question, it is important to underscore that these are likely to be informal, “between the rules” changes to working practices. The emphasis on managerialism within the public sector in the managerial and academic literatures since the 1970s has highlighted that the real tradeoff is between increased accountability for managers and reduced structural restrictions on their use of inputs (money, resources, and people) (Knies and Leisink 2013). But loosening those formal structural constraints is difficult: “Overall public sector management reforms are inducing more entrepreneur-like behavior patterns and attitudes on the part of public managers ... (but) entrepreneurialism can only be introduced at the expense of the more traditional input-oriented type of governing with accountable scrutiny” (Koch 1996, 39–40). More likely, it is the behavior of managers, rather than their formal power, that leads to successful implementation at the unit level (Norrgrén and Schaller 1999). This view is confirmed by the research in the United Kingdom by Purcell and others (2003a), who note that high levels of organizational performance are not achieved simply by having well-conceived human resources policies and practices in place: what makes the difference is how these policies and practices are implemented. How frontline managers show leadership, deal with employees, and exercise control are major issues—and it in these areas

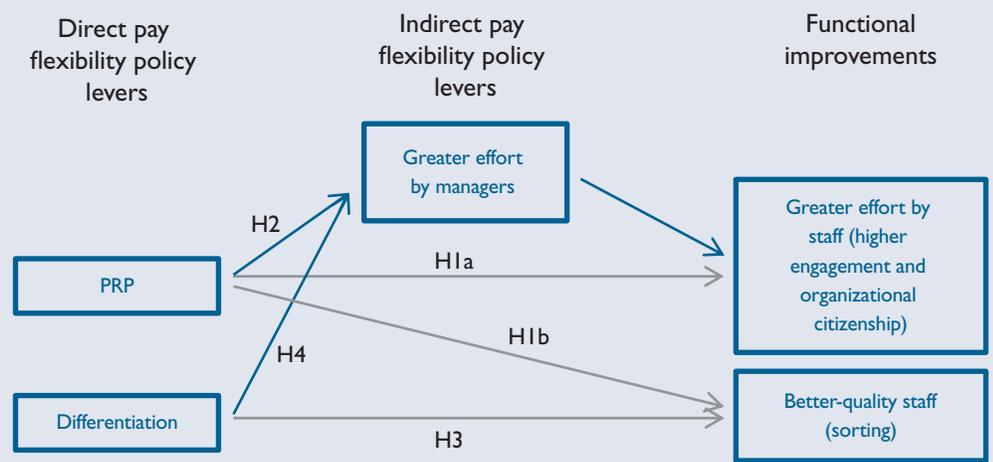
where they can exercise the greatest amount of discretion, “bringing human resource policies to life” (Hutchinson and Purcell 2003).

In sum, line managers play a crucial role in providing positive feedback and recognition and in finding opportunities for staff to develop and the scope to exercise responsibility (Armstrong 2010). They thus play a major part in influencing employee attitudes toward the organization and their jobs (Purcell and Hutchinson 2007). Where this shows in practical terms is the frontline manager’s role in modest modifications to a staff member’s work allocation, responding as well as possible to the skills of the staff member and demonstrating a link between the tasks and organizational goals and how they fit with the contribution of others (Purcell and Hutchinson 2007).

On the second question, there are three likely reasons why frontline managers might want to expend additional effort to find small, “between the rules” informal ways to better motivate their staff. First, it seems reasonable to assume that putting money on the table for staff performance attracts senior management attention and so making frontline managers sensitive to their unit’s performance regardless of whether they are themselves on PRP. Second, managers will have a heightened awareness of the organization’s objectives since they are now spending much of their managerial time translating them into work targets for their staff. Third, performance discussions with their staff are, inevitably, performance feedback sessions for themselves, likely bringing options for these informal “between the rules” issues to their attention.

Figure The Link between Pay Flexibility and Performance

1



their personnel on their own discretion rather than by centrally managed rules. Other factors in the case studies, not treated as systematically but nonetheless important, include the nature of the political context and potential synergies and complementarities with other reforms.

PRP schemes can be distinguished based on five design elements: whether the bonus is based on individual or group performance, the time horizon of the financial incentive, the nature of the performance evaluation, the size of the bonus, and the probability of receiving

Table 1 Hypotheses on the Impact of Pay Flexibility on Civil Service Performance	
Arguments for ←	→ Arguments against
PRP	
Hypothesis 1. Direct pay flexibility levers	
<i>Hypothesis 1a. Direct incentive effect on effort (PRP can directly affect individual engagement and organizational citizenship)</i>	
<ul style="list-style-type: none"> Induces staff to exert more effort toward achievement of the outputs and outcomes linked to the incentive 	<ul style="list-style-type: none"> No effect on effort: Difficult to measure outputs in the public sector given lack of objective performance measures Negative effect on effort: Given the difficulties in measuring outputs, the perceived unjustified pay inequity breeds resentment Negative effect through unintended consequences: "Gaming behavior" Behavioral economics: Crowding out of intrinsic motivation, which reduces effort
<i>Hypothesis 1b. Direct incentive effect on staff quality (PRP can have a direct effect in improving the recruitment and retention of better-quality staff)</i>	
<ul style="list-style-type: none"> Attracts higher-quality staff who are likely to do well under the scheme (sorting) 	
Hypothesis 2. Indirect pay flexibility levers (PRP can act indirectly by providing incentives for improved management)	
<ul style="list-style-type: none"> Positive effect through better management of the "effort bargain," in particular improvements in: <ul style="list-style-type: none"> Organizational goal setting Teamwork task allocation toward achieving organizational goals Linking individual performance appraisals to organizational goals 	<ul style="list-style-type: none"> Negative effect: Pay inequity results in harmful competition, hurting teamwork and reducing staff trust in management
Differentiation	
Hypothesis 3. Direct pay flexibility levers (Differentiation can have a direct effect in improving the recruitment and retention of better-quality staff)	
<ul style="list-style-type: none"> Direct incentive effect on staff quality: More targeted recruitment of high-quality staff for priority activities within governing fiscal constraints (sorting) 	<ul style="list-style-type: none"> Induces staff in the lower-paying segments of the government to exert less effort; reduces pay transparency; and increases pay inequity, thereby breeding resentment
Hypothesis 4. Indirect pay flexibility levers (Differentiation can act indirectly by providing incentives for improved management)	
<ul style="list-style-type: none"> Puts the spotlight on management to improve the performance dialogue 	<ul style="list-style-type: none"> Hurts interagency cooperation

the bonus (table 2). Rewarding team performance can have certain advantages, ranging from reduced evaluation costs to avoiding harmful competition between employees. However, basing rewards on team outputs can also lead to free-riding, where some team members reduce their effort and rely instead on the work of others. The awards can be one-off bonuses or merit increments to salary that are permanent and cumulative, or can be based on the past year's performance or on multiple years' performance. Performance evaluations can be based on quantitative performance targets or subjective assessments against individual results agreements. Small bonuses have little

incentive effect, while large bonuses can further encourage gaming and under extreme circumstances result in "choking under pressure" and therefore hurt performance (a phenomenon known as the Yerkes-Dodson Law). Finally, if the probability of receiving the performance bonus is either close to 0 or to 1, the incentive will have no impact (Bruns, Filmer, and Patrianos 2011).

On nature of the job, this study relies on the fundamental insight of James Q. Wilson (1989) that government bureaucracies vary along two dimensions: whether the tasks performed by the individuals in the organization, or the inputs of labor, are easily observable and

Table 2 Five Key Design Elements of PRP Schemes					
2	Individual- or group-based awards	Time horizon of the incentive	Nature of the performance evaluation	Size of the award	Probability of receiving the award
	Individual awards can breed harmful competition, but team awards can encourage free-riding	Short-term annual bonuses or longer-term merit increments. Single- or multi-year performance appraisals	Quantitative performance targets may be more objective but are rare for public sector jobs	Small awards have limited effects, but very large awards can further encourage gaming and perverse incentives	If the probability of receiving the performance bonus is either close to 0 or 1, the incentive will have no impact

Source: World Bank staff.

measurable to managers and external agents; and whether the outputs or outcomes of these jobs are easily measurable. Inputs may be hard to monitor because they may be highly technical and esoteric in nature (for example, a doctor performing a diagnosis) or because these actions cannot be readily observed since staff are in faraway locations or act out of view of managers (for example, forest rangers or rural teachers). Outputs may be hard to measure because indicators may be difficult to find, or, more fundamentally, to define what the goal of a particular agency is.

Table 3 provides a classification of job types, with the simplifying assumption that jobs with multiple dimensions are located within the cell that represents the most complex of those dimensions.⁴ The top left box describes “production jobs,” in which outputs are easily measurable; the production process consists of repeatable, mechanical tasks that are observable to an outside monitor; and controllability is likely to be high. Typical examples are manufacturing factory-floor jobs, the postal service (where letter sorters can be observed and the speed of mail delivery measured), and municipal services like garbage collection. If the production process is not directly observable but outputs remain measurable, such jobs are termed “craft jobs.” With recent advances in measuring learning outcomes, teaching can be classified as a job in which the exact process of production is hard to ascertain but desired outputs are to some degree quantifiable. Similarly, some of the outputs of health care, particularly in preventive

services like child immunization, are also more measurable. Other examples include tax collection, job placement services, and auditing.

In the bottom row are “procedural jobs” and “coping jobs.” Both are characterized by difficult-to-measure outputs, but again they differ in the observability of the production process to an outsider. Procedural jobs like the military have clearly defined and highly routinized and monitored inputs, which can be specified in standard operating procedures that proliferate and that regulate every detail of work. Administrative jobs in general policy units of the central government neither produce easily measurable outputs nor have transparent production processes. These coping jobs present the most challenging functional contexts for PRP and more generally for inducing a mission orientation in the organization.

For this study we use the distinction between jobs with observable and unobservable production processes as a rough proxy for distinguishing between jobs that tend to be found in the private or public sectors, respectively. Within the latter, we use the distinction between those public sector jobs where outputs are measurable and those where they are not as a more precise proxy for distinguishing between broad public service jobs and those in the core administration. We take those studies of PRP in coping jobs as likely the best measure of the impact of PRP in the primarily policy jobs in the core civil service.

The study also looks at organizational autonomy, or delegation, in the use of the factors of production—the extent to which

Coping jobs present the most challenging functional contexts for PRP

		Actions or internal production process of the job	
		Observable	Not observable
Outputs from the job	Relatively easily measurable	<p><i>Production jobs:</i> Simple repetitive stable tasks, specialized skills</p> <p>Examples: Manufacturing, sales, simple clerical tasks, postal service, garbage collection</p>	<p><i>Craft jobs:</i> Application of general sets of skills to unique tasks, but with stable, similar outcomes</p> <p>Examples: Auditing, revenue collection, teaching, medical practice, job placement work</p>
	Not easily measurable	<p><i>Procedural jobs:</i> Specialized skills; stable tasks that can be routinized through standard operating procedures, but unique outcomes</p> <p>Examples: Military</p>	<p><i>Coping jobs:</i> Application of generic skills to unique tasks whose outcomes cannot be evaluated in the absence of alternatives</p> <p>Examples: Policy formulation, administration; managerial jobs in large private sector organizations</p>

Source: Adapted from Wilson (1989).

agencies have the authority to manage their personnel on their own discretion rather than by centrally managed rules. This autonomy can include the ability to hire and fire staff, establish agency hiring standards and salaries, and make outsourcing decisions while retaining central control only over the agency's total wage bill.⁵

Delegation can potentially impact the efficacy of both PRP and differentiation. For PRP to work, managers in an agency need some degree of freedom to define the agency and individual goals and to evaluate staff performance, the assumption being that these managers have more information about outputs and inputs in these jobs and are therefore

better able to make these decisions. Delegated authority allows agencies to complement new organizational goals with adjustments in the workforce, hiring and firing staff members, reorganizing teams, and providing financial incentives (Rexed and others 2007). Civil service agencies that have little control over personnel planning or salary structures have few tools to reorganize the production of services, since their main input is human capital. Joint authority over service delivery, input allocation, and overall costs increases the responsibility and accountability of senior management as the overseeing authority and thus allows better assessment of performance and service quality.

Assessing the Evidence: Review of Literature on Performance- Related Pay

The first piece of evidence on the impact of pay flexibility is based on a comprehensive review of the existing literature on PRP that disaggregates the studies by type of job, using the classification discussed in the previous section, and study quality. In total, 153 empirical studies of PRP were considered in this review (see Hasnain, Manning, and Pierskalla [2012]) for the full list), of which 110 were of craft and coping jobs and 17 were of coping jobs specifically (table 4). The research to date on the subject has largely focused on advanced countries—in the review 127 studies are in OECD countries, and only 26 are in developing countries. The empirical literature also employs a range of methodologies, from earlier observational and often qualitative studies to the more recent field randomized control trials (RCTs) and laboratory experiments explicitly aimed at teasing out causality. These experimental studies have

been exclusively on craft jobs and production jobs, with none to date on coping jobs.

These reviewed studies were grouped into three categories to capture the effect of PRP: *positive* if their findings provide evidence for the effectiveness of incentive schemes;⁶ *neutral* if the study is largely descriptive or finds contradictory evidence; and *failed* if the evidence indicates no effect or a negative effect for PRP. Figure 2 shows the overall frequency of results. Most of studies (93 of the 153) present evidence for the effectiveness of performance pay schemes, with experimental studies showing more positive findings than observational ones.

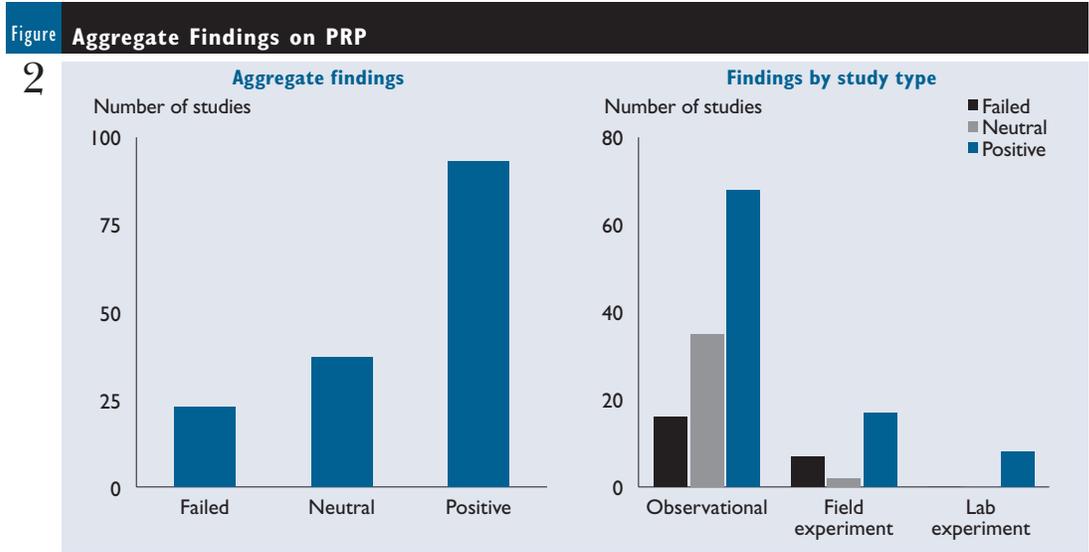
In drawing conclusions, however, it is important to distinguish the findings by study research quality. Study quality was ranked in two different ways. First, each study was assessed for its internal validity, or the strength

Table 4 The Studies Reviewed

4

Country and methodology	Types of jobs					Total
	Production	Procedural	Coping	Craft	Unclassified	
<i>OECD study</i>	27	0	16	71	13	127
Observational	14	0	16	58	13	101
Field RCT	7	0	0	13	0	20
Lab experiment	6	0	0	0	0	6
<i>Developing country study</i>	1	0	1	22	2	26
Observational	0	0	1	15	2	18
Field RCT	0	0	0	6	0	6
Lab experiment	1	0	17	1	0	2
<i>Total</i>	28	0	17	93	15	153

Source: Hasnain, Manning, and Pierskalla 2012.

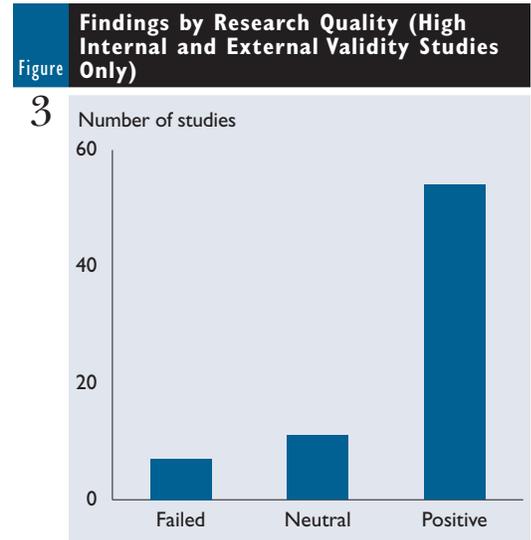


Source: Hasnain, Manning, and Pierskalla 2012.

of the causal arguments being made, using a five-point ranking (from weak to strong):

1. No empirical study or faulty research design.⁷
2. Descriptive; small sample size.⁸
3. Secondary data analysis or descriptive data analysis; small sample size; some statistical analysis.⁹
4. Quasi-experimental design; reasonable sample size; conclusions based on statistical analysis.¹⁰
5. Laboratory experiments; RCTs; large sample size; strong statistical analysis; strong conclusions.

Second, studies were also evaluated on the dimension of external validity, or the extent to which causal connections drawn in the specific context of the study would remain valid if replicated in other contexts. For example, lab experiments and RCTs offer strong evidence of causality (high internal validity) but in a specific context—they tell us the average impact of a particular intervention in a particular location with a particular sample at a particular point in time. They are often accused of being low on external validity because the study subjects (usually college students in the case of laboratory experiments) are not representative of the general population, or in this case the population of interest (civil servants), and the requirements of the experiment imply conditions that may not approximate real world settings.

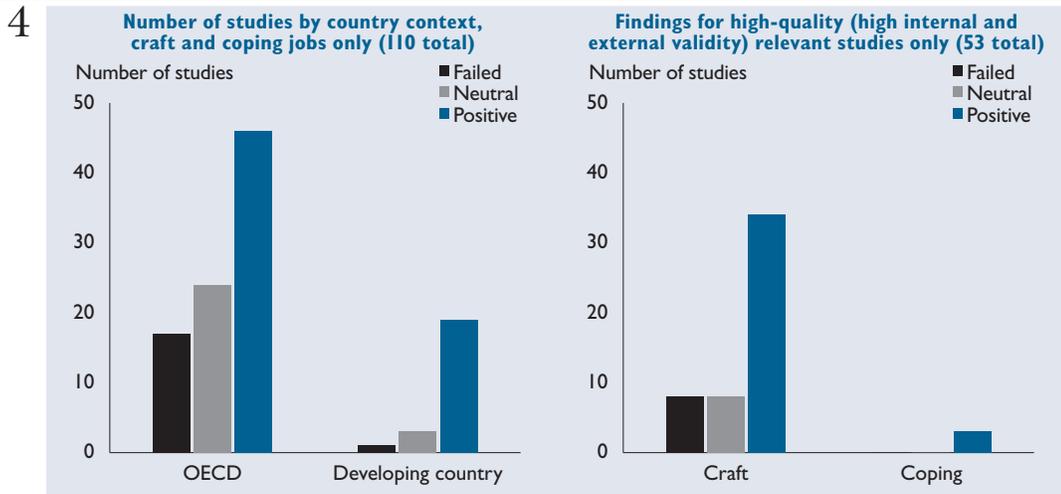


Note: Seventy-two studies.
Source: Hasnain, Manning, and Pierskalla 2012.

Figure 3 applies these two quality filters to these 153 studies, resulting in 72 high-quality studies (ranked 4 or 5 on the internal validity scale, and “high” on the external validity scale);¹¹ 54 of these high-quality studies show positive results.

The effect of PRP in craft and coping jobs is particularly interesting, since these jobs most closely resemble public sector organizations. Figure 4 presents the evidence for all the reviewed studies for craft and coping jobs (110 studies) and for the high-quality studies of craft and coping jobs (53 studies). The overall evidence is generally quite positive, though

Figure Findings for Craft and Coping Jobs



Source: Hasnain, Manning, and Pierskalla 2012.

the evidence is almost exclusively for craft jobs, and mostly in OECD settings.

There is an extensive and growing literature on performance pay for teachers. In the United States most observational studies have primarily examined the impact of performance incentives on student test scores, though a few studies, such as Clotfelter, Diaz and others (2004) and Clotfelter, Glennie, and others (2008), show, using detailed data from North Carolina's schools, that accountability and performance pay systems can help retain quality teachers. The evidence is mixed with regard to student test scores. Cross-sectional studies from the American National Educational Longitudinal Survey show positive results (Figlio and Kenny 2007; Winters and others 2009). A number of studies identify problems of gaming, such as outright cheating (Jacob and Levitt 2003; Jacob 2005) or, more subtly, the adjustment of the caloric content of school lunches to improve cognitive ability on test days (Jacob and Levitt 2003; Figlio and Winicki 2005; Jacob 2005).

Outside the United States an analysis by Atkinson and others (2004) finds clear positive effects of performance pay for British schools. A set of observational studies uses data from an Israeli policy experiment with tournament-based teacher competition for bonuses and a regression discontinuity and difference-in-difference design to approximate random

treatment assignment (Lavy 2008, 2009). These studies by Lavy also show significant gains in student achievement. They also identify changes in teaching methods, enhanced after-school teaching, and increased teacher responsiveness as the key mechanisms for these improvements.

A number of field experiments have evaluated the impact of performance pay for teachers on reducing absenteeism and improving learning outcomes. The findings are generally mixed but are, interestingly, more positive for developing countries. Duflo, Hanna, and Ryany (2010) show that monitoring teacher attendance through tamper-proof cameras linked to financial incentives in rural India led to a strong reduction in teacher absenteeism and increased students' test scores. By contrast, Kremer and Chen (2001) found that in Kenya subjective monitoring arrangements by an individual in the institutional hierarchy (such as a school's headmaster) may not work because the monitor might shirk, try to avoid confrontation, or collude with the workers. These studies suggest that impersonal, external monitoring through technology coupled with a clear, credible, and automatic threat of punishment and promise of reward was the key design feature for program success.

A field experiment in 50 Kenyan schools linking teacher salaries to student test scores failed to find lasting effects (Glewwe, Ilias,

There is a large and growing empirical literature on PRP in education that surprisingly finds more positive results in developing countries

and Kremer 2010): teacher attendance did not improve, and teachers did not adjust their teaching methods or conduct more preparation sessions. Students in treated schools did perform better during the program duration, but these gains did not carry beyond the study period. However, a large-scale field experiment in a representative sample of 300 government-run rural primary schools in India found that bonus pay linked to the mean improvement of student test scores in an independent learning assessment led to a statistically significant and substantively meaningful improvement of student outcomes (Glewwe, Ilias, and Kremer 2010).

RCTs in the United States have been quite negative on the effect of teacher incentives on student outcomes. A field experiment conducted in New York City public schools also failed to find statistically significant effects of team incentives for teachers on student outcomes (Fryer 2011). A related study that also assessed the effects of the New York City group incentive program on classroom activities and teacher turnover and qualification, apart from test scores and teacher effort, similarly found no effects (Goodman and Turner 2010). A three-year experimental evaluation of the Project on Incentives in Teaching in metropolitan Nashville schools also found no significant effects of bonus incentives on student test scores (Springer and others 2010).

In the health sector a number of RCTs have been implemented to determine the role of performance pay on health worker productivity, patient treatment, and outcomes. The majority of studies assess these questions in the context of OECD health care systems. Kouides and others (1998) implemented an RCT offering financial incentives to a randomly selected set of primary care physicians based on influenza immunization rates of the elderly as part of a Medicare demonstration project. Doctors in the treatment group performed more immunizations. However, Hillman and others (1998, 1999) used two RCT designs to provide incentives for cancer screenings for women of age 50 and above and pediatric immunizations, respectively. In both studies the authors document no significant difference between

treatment and control groups. Similarly, a RCT implemented by Grady, Lemkau, and Caddell (1997) found no clear effects of financial incentives on mammography referrals by primary care physicians.

By contrast, a set of studies, also focusing on pediatric immunizations, found that performance incentives increased immunization rates by several percentage points compared with the control group (Fairbrother, Hanson, and others 1999; Fairbrother, Siegel, and others 2001). An RCT at the clinic level found that financial incentives improved treatment of smoking cessation outcomes (Roski and others 2003). Work on performance pay for cognitive services interventions by pharmacists also found positive effects (Christensen and others 2000).

To our knowledge, the only two RCTs on performance pay in health care in a low-income country are a study by Basinga and others (2010) in Rwanda and a study by Singh (2010) in India. Basinga and others used an RCT design to evaluate performance pay in Rwandan primary health care centers. The authors took advantage of a sequenced roll-out of the scheme across Rwandan health care facilities, collecting data on child preventive care and prenatal delivery. To isolate the performance-pay effect from a general increase in resources, comparison facilities received an equivalent increase in their budgets. The study, using information from 166 facilities and 2,158 households, found large effects on all central outcome measures, but with particularly striking effects for services with the highest payoffs and smallest necessary staff effort.

Singh (2010) treated three groups of mothers and staff providing child care and nutritional advice to them in Chandigarh, India. In one group the workers received performance pay; in a second group the workers had no performance pay but the women that they worked with were separately given factual information about nutrition; and the third group received both treatments. The study found that children's weights improved only in the third group compared with the control group.

Note that nearly all the studies in the health care sector focus on fairly narrow types of

performance pay and specific, single-outcome measures in preventive care rather than on overall multidimensional patient treatments and outcomes.

Revenue authorities and job placement agencies also provide examples of craft jobs where, though work methods are hard to observe, the outputs—number of audits conducted, tax fines collected, job-seekers finding employment within a specified time period—are more easily measurable. Kahn, De Silva, and Ziliak (2001) examined a 1998 incentive scheme in Brazil and found that it resulted in a 75 percent increase in fines per inspection. Burgess and others (2010) used an RCT to examine the impact of a pilot team-based incentive scheme based on revenue collection and audit targets introduced in 2002 in a U.K. indirect tax assessment and collection agency. The tax yield increased for the treatment team relative to the control group, with the increases due to more time spent on audits, which resulted in recovery of greater tax revenue. By contrast, Bertelli (2006) found that, in the U.S. Internal Revenue Service, the incentive scheme “crowded in” intrinsic motivation at the lowest pay levels and crowded it out at the highest levels. Similarly, a set of studies of performance incentives for agencies with responsibility for training and recruitment found considerable evidence of gaming among agency staff in the choice of termination date of the training for the participants (Asch 1990; Heckman, Heinrich, and Smith 1997; Courty and Marschke 2004).

Only a few studies have attempted to evaluate the sorting effect of PRP. An experiment with 115 Australian students that tried to distinguish the potential incentive and sorting effect of performance pay found evidence for both hypotheses (Cadsby, Song, and Tapon 2007). In addition, it found that low-productivity subjects were less likely to sort into pay-for-performance jobs and that subjects with higher levels of risk aversion avoided PRP, suggesting important unintended side effects. In a field experiment in the private sector (Bandiera, Barankay, and Rasul 2006), some managers were treated with the introduction of a performance-pay system, and productivity

of lower-tier workers was used as an outcome measure. The study found evidence of both incentive and sorting effects—that is, managers support their high-productivity workers and fire the least qualified employees.

To date there have been few high-quality studies of coping jobs. Of the three studies reviewed, two were of performance pay for managerial positions in the private sector, and only one was for core administrative jobs in the public sector. All these studies showed positive effects of the performance incentive. For example, Hochberg and Lindsey (2010) found a positive impact of stock options for company rank-and-file on firm profits. Dowling and Richardson (1997) used an employee perception survey to examine PRP for managers in the U.K. National Health Service and found a modest positive effect of the incentive on manager motivation and effort. No high-quality studies were found of this type of job within developing countries.

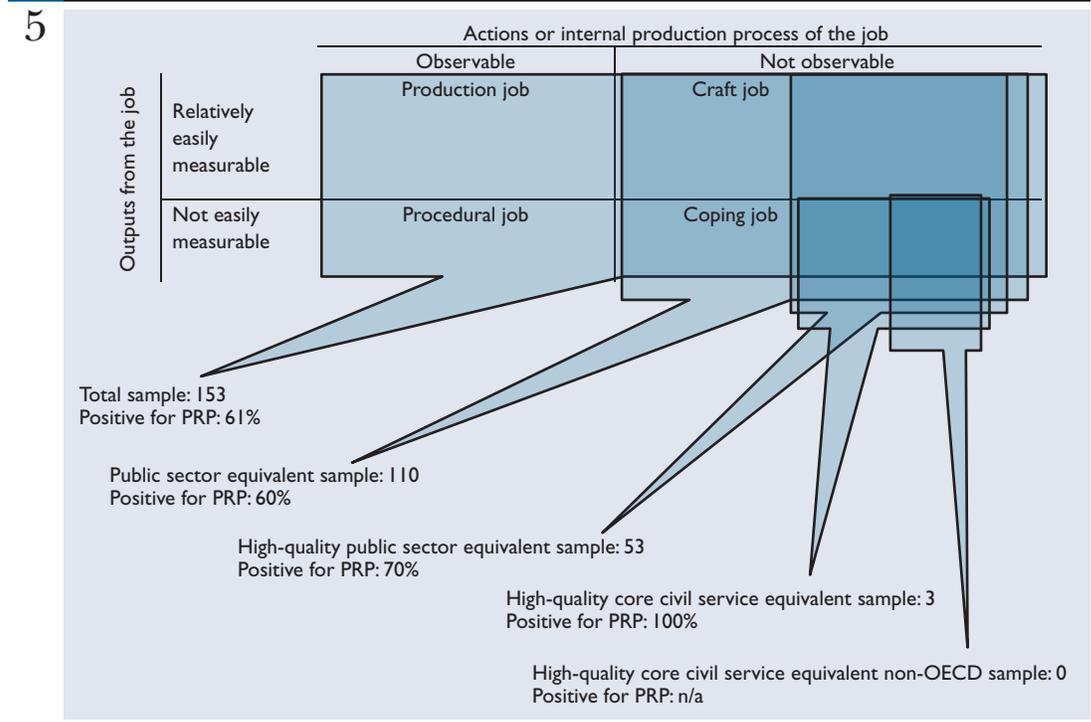
To summarize, when we winnow the pool of studies to identify the subsets that are most relevant to the tasks facing senior administrators within the core civil service in non-OECD settings, the number of studies from which we might draw policy lessons becomes quite small (figure 5). Therefore, no conclusions can be drawn from the existing literature on the effects of PRP in these organizational contexts.

The review does enable us to conclude that the incentive theory prediction—that PRP has a role to play in craft jobs where outputs are readily observable, such as teaching, health care, and revenue administration jobs—holds true, even though the day-to-day actions of staff are unobservable. This apparently confounds, at least in the short term, the concern in behavioral economics about crowding out intrinsic incentives. This conclusion is supported by observational and experimental studies in developing countries, where the evidence is generally more positive than in OECD settings.

At the same time, several observational studies identify problems with unintended consequences, generically subsumed under gaming the incentive scheme, which can subvert the original intentions of the reforms. With the

There are too few high-quality academic studies of coping jobs to draw any firm conclusions

Figure 5 Numbers of Studies and Their Findings



current evidence though, it remains unclear whether incidents of gaming have a net negative effect in the presence of increased productivity. Furthermore, while explicit incentive schemes certainly increase the opportunity for gaming, standard civil service arrangements have their own unintended incentive effects, with employees engaging in behavior that increases the chances of easy work assignments or promotions. It is simply unknown whether existing forms of gaming are worse than similar behavior under performance pay. In addition, there might exist important cultural differences in the prevalence of gaming performance standards in the public sector between developed and developing countries. While to our knowledge no explicit research on this question exists, work on the prevalence of corruption, behavioral norms, and the effectiveness of anticorruption efforts suggests that gaming in highly corrupt bureaucracies might be more problematic.

For RCTs, the evidence again speaks in favor of the potential utility of performance pay for craft jobs. Comparing various laboratory experiments, the results suggest that explicit

performance incentives can work, but the studies employ easily measurable performance indicators and use fairly unrepresentative subject pools. Both concerns should caution policy makers against accepting the results independently of other research. However, similar results have been found across a varied set of experimental settings, test locations, and subject pools, and the overall findings do resonate with the observational literature, improving overall credibility and external validity.

The strongest form of evidence comes from field experimental studies for craft jobs that neatly address concerns of internal and external validity. Here, evidence is somewhat more mixed. Several studies of teacher incentive programs found no or transient effects of bonus pay systems in U.S. schools, but in the developing world evidence has been more positive. The discrepancy between teacher incentives in the developed and developing world could stem from the relative magnitude of incentives compared with normal salary, or from higher marginal effects in the education production function in developing countries. Many factors affect the education process, all of which

are likely lacking in many developing country schools. Improving one input aspect, such as teacher presence and effort, could have conceivably larger marginal effects than the same input improvement in a developed country school.

For coping jobs in developing countries, the evidence has to draw more on the case studies, which will not meet the rigorous standards of an RCT but can nonetheless provide insights on the tradeoffs involved.

In sum, the evidence in relation to the hypotheses set out earlier is primarily around Hypothesis 1: that PRP has a direct impact on staff effort and on the type of staff recruited or retained. The literature shows no relevant support for Hypothesis 1 for coping jobs but reasonable evidence concerning craft jobs (both within and outside OECD settings) even though gaming is a persistent problem. Some sorting effect is noted, primarily in OECD country settings.

The Case Studies: Pay Flexibility Schemes

Fiscal sustainability was a trigger for the introduction of PRP in Brazil and Korea, echoing moves in the OECD countries a decade earlier

Brazil, Chile, Indonesia, Korea, Malaysia, the Philippines, Russia, and Thailand represent an interesting diversity in their pay flexibility schemes. It's worth noting that performance improvement was usually only one reason for introduction of these plans, with fiscal and political reasons also important. Fiscal constraints meant that the wage bill had to be brought under control, and a different mechanism for salary increases had to be found or to be more explicitly linked to productivity improvements. Fiscal sustainability was in particular a trigger for the introduction of PRP in Brazil and Korea, echoing moves in the OECD countries a decade earlier. Korea introduced PRP in the aftermath of the Asian financial crisis in 1997, and Brazil similarly moved in this direction in the wake of the fiscal crises affecting Brazilian states in the early 2000s as part of a package of budget and management reforms to increase the efficiency of the public sector.

Political crisis or a change in regime was another motivation. Politically, PRP has often been viewed as a means to enforce responsiveness, particularly among senior bureaucrats, to the political leadership, or as a politically more acceptable way of increasing public sector wages. In Thailand the main driver for reform was former prime minister Thaksin Shinawatra, who had a degree of executive power that was unprecedented in Thai politics and who, given his business background, sought to use PRP to infuse private sector responsiveness into the civil service. In the Philippines, PRP was the initiative of the newly elected reform-minded president Benigno Aquino, who,

limited to one six-year term in which to deliver on his promises, needed to quickly infuse some dynamism into a notoriously sluggish bureaucracy.

Beyond these fiscal and political motivations, it is not apparent from the case studies that these pay systems had achieved a minimum threshold of functionality before introduction of second-generation pay flexibility reforms. With the exception of Thailand and, surprisingly, the Philippines, these countries had highly fragmented pay regimes. In general, it was the norm for pay to vary between workers doing similar jobs based on their agency, occupational group, or geographic location, or because of idiosyncratic personal characteristics. While in some cases, these differences reflected official policies as specified in laws or regulations, more often they were uncoordinated outcomes of ad hoc policy changes or political power, and they often provided incentives for unproductive behavior.

As an example, in Indonesia base pay accounts for less than 20 percent of total compensation for mid- to senior-level staff, and there are hundreds of different types of allowances and supplementary payments for different categories of staff. Among these are honoraria, which are additional payments to staff for taking on tasks and responsibilities in addition to their regular duties, such as attending workshops or participating in meetings that are called by other agencies, or working on a team on a special project. These honoraria are paid in cash and are not included in the personnel budget, and they can amount to a third of total cash compensation. It is not

hard to see that honoraria can induce unproductive behavior, since they are paid in a highly nontransparent manner and encourage staff to maximize the number of project teams they belong to and the number of meetings and workshops they participate in.

In almost all cases, pay flexibility measures added another layer to an already complex pay regime, but one that was, at least rhetorically, motivated by the desire to enhance individual and agency productivity. The question then is whether the tradeoffs in terms of increased inequity and possible lack of transparency and lower intrinsic motivation were offset by the direct and indirect productivity-enhancing effects of these changes.

Performance-related pay

In the case countries, Brazil, Chile, Korea, Malaysia, the Philippines, Thailand, and, to a limited extent, Indonesia have implemented PRP. These PRP schemes display a high degree of diversity in the five main design features (summarized in table 5 below). The schemes in Brazil, Chile, the Philippines, and, to some extent, Korea have both individual- and group-based incentives, either through two separate schemes or through a formula that combines assessments of group and individual performance. Chile and Korea also have distinctive PRP schemes for senior civil servants. The award can be paid either as an annual bonus, as in Chile and the Philippines, or a merit increment, as in Korea and Thailand. In Brazil the bonus increases based on the number of good successive performance evaluations staff have. The performance assessment mechanism usually uses measures of organizational and individual performance, but with notable variations in the use of quantitative standardized indicators versus individually specific results agreements and short- versus long-term measures. The size of the awards varies considerably, from a low of 3 percent of basic pay in Thailand to a high of more than 100 percent of basic pay in Chile (for the Senior Executive Service). Finally, some of the countries—Korea, Malaysia, and the Philippines—have tried to counter the usual tendency of equal distribution of PRP (which renders it a salary

supplement) by mandating a distribution of performance rankings of organizations or staff.

PRP is ubiquitous in the Chilean public administration, with schemes that apply to most central administration institutions (across-the-board schemes) and others that are specific to particular institutions, such as the revenue authority. The across-the-board schemes are an annual institutional-level bonus based on the achievement of ministerial/agency targets and a team-level bonus defined internally in each agency. Until 2011, the agency-level targets consisted of achievement of outputs and improvements in processes such as basic human resource practices, planning and management control, procurement, and audit. Each agency defined its action plan in agreement with the Budget Office and self-evaluated its accomplishments of the previous year, with a group of experts validating the report. This procedure changed in 2011, with the parent ministry having a much stronger role in discussing the agency's targets and a stronger focus on results rather than on outputs and processes. For the working unit scheme, the concerned line minister or head of agency divides the institution by groups and defines annual goals and performance indicators for each (collective performance agreements), and the internal audit unit evaluates the level of accomplishment annually. The size of the award ranges from 4 percent to approximately 8 percent of base pay for both the agency-level and working unit-level schemes, implying that staff in a high-performing unit in a high-performing agency (usually over 95 percent of public sector staff) could receive an annual bonus of as much as 16 percent of base pay.

A government-wide individual PRP scheme was introduced in 1998 but abandoned in 2003, as supervisors faced with the challenge of forcing a normal curve—when traditionally all employees had received the top qualification in the performance appraisal system—rotated the bonus between groups of employees so that each received an equal share over the medium term. Today, Chile has only an individual performance bonus scheme for members of the

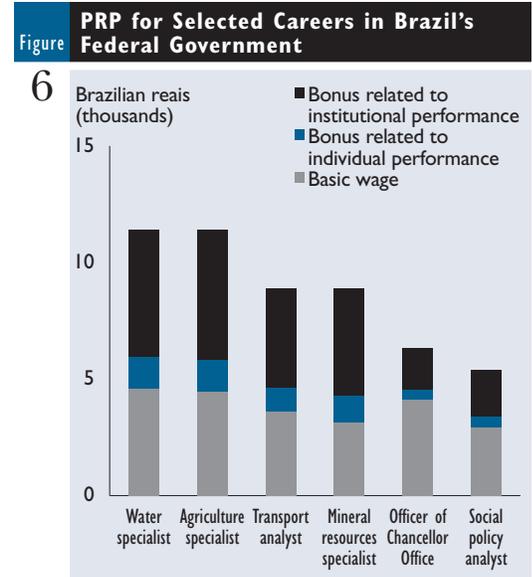
PRP is ubiquitous in the Chilean public administration, with schemes that apply to most central administration institutions (across-the-board schemes) and others that are specific to particular institutions

Brazil, both at the federal level and in the state of Minas Gerais, has a performance bonus scheme that consists of individual- and group-based awards

Senior Executive Service (SES), which is based on evaluations by the immediate supervisor and uses targets established in individual results agreements. The size of the maximum bonus is determined by the Ministry of Finance in consultation with the Civil Service Directorate and the concerned agency based on market conditions and an assessment of the level of pay needed to attract and retain suitable candidates. The incentives can be high, with the bonus equal to as much as 100 percent of base pay if all targets are met in the case of the head of a hospital in a remote location, but are usually a maximum of 30 percent of base pay for other SES positions. Almost all SES members receive the complete bonus, and the PRP is therefore a de facto salary supplement rather than a performance incentive.

Brazil, both at the federal level and in the state of Minas Gerais, has a performance bonus scheme that consists of individual- and group-based awards (this study covered only Minas Gerais at the state level). Remuneration practices are largely organized around career categories that group professional positions on a common salary spine. In the federal government a range of careers have prescribed breakdowns for performance allowances derived from either individual or institutional criteria (figure 6). The performance assessment allocates points based on the achievement of individual, working unit, and institutional goals, with the individual component having a weight of 20 percent and the group component a weight of 80 percent. Each point has a monetary value, the details of which vary by career group, and is at the discretion of the concerned agency. The institutional component of the PRP is usually linked to goals set in the multiyear plan for each government entity. Interviews with government officials stressed that there is no explicit system of results agreements across government units or even a precise set of indicators set out in these broad umbrella plans. Nonetheless, some ministries have established institutional goals that are set out in their own specific regulations, which are published annually online.

The Brazilian state of Minas Gerais has a similar formula for PRP but with some



Source: Ministry of Planning.

interesting variations and considerably more vigor in its application. The group-based bonus, called the Productivity Premium, began in 2003 and is based on the percentage of institutional goals achieved, as identified in the results agreements that agencies sign with the governor and that are cascaded down to working units. There is a time-in-service requirement, so that bonus size is also conditional on the number of days that staff work. The individual bonus, adopted in 2007, applies to all public servants recruited after 2003 and is calculated on the basis of both institutional goals reached (derived from the results agreements for the whole agency) and the individual performance appraisal of the specific public servant, with 30 percent weight given to the former and 70 percent to the latter. The value obtained by measurement of these two categories is then multiplied by a factor that takes into account the basic salary of the public servant and the percentage associated with the number of satisfactory evaluation cycles in the course of the civil servant's career. This multiyear satisfactory evaluation requirement was designed to provide long-term incentives for performance, with the public servant seeing a substantial increase in bonus size over time. For example, a civil servant with 5 satisfactory performance evaluations will receive a 10 percent bonus, while one with 15 will receive a

30 percent bonus, with even higher bonuses possible in the future (figure 7).

The Philippines until recently had very little pay flexibility, with public sector remuneration approximating the single pay spine across government departments and agencies and allowances making up less than a quarter of total monetary compensation. In 2012, at the initiative of President Aquino, the government introduced an individual- and group-based performance bonus that applies to all central government departments and agencies. The objective of this scheme, as stated in the presidential order that promulgated it, is to help deliver on the Social Contract of the Aquino administration. The Social Contract lays out a broad reform agenda to promote inclusive growth in the Philippines, with combating corruption and “professional, motivated, and energized bureaucracies” as key elements for achieving the government priorities of higher revenue generation and improved outcomes in education, health, and other services. The PRP scheme is a deliberately ambitious, government-wide reform initiative that is designed to “shake the bureaucracy out of its lethargy” to achieve some sustainable impact in the remaining three years of the Aquino administration.

The scheme, as currently designed, is aimed at achieving three objectives, not all of which

are explicitly stated in the policy guidelines: improve agency performance as measured by achievement of departmental performance targets and key presidential priority programs; improve individual performance; and improve agency compliance with existing governance-related laws and regulations, such as transparency in procurement, financial management, and disclosure of information. The PRP scheme provides an integrated group and individual award.

Departments and agencies can qualify for the performance bonus if they meet 90 percent of their agreed performance targets and additional good governance criteria. Within departments, comparable working-level units (policy bureaus, implementing units, and services) will be force-ranked into three categories, and staff within these units will also be force-ranked into three categories so that 10 percent of units and 10 percent of staff within those units are classified as best performers, 25 percent of units and staff are in the second category, and 65 percent are in the bottom category. While the government’s ambition is to have the working unit performance targets cascaded to the individual performance evaluations, the individual bonuses are currently based on the traditional, trait-based performance evaluation tool that has long been in use in the civil service. Individual bonuses, which are flat sums and not a percentage of salary, can vary from 35,000 pesos to 5,000 pesos, depending on where staff are in the three-by-three matrix of working unit and individual performance, implying that as a percentage of pay the highest incentives can vary between 10 to 18 percent of total compensation for clerical and junior technical staff to less than 5 percent for senior management (figure 8).

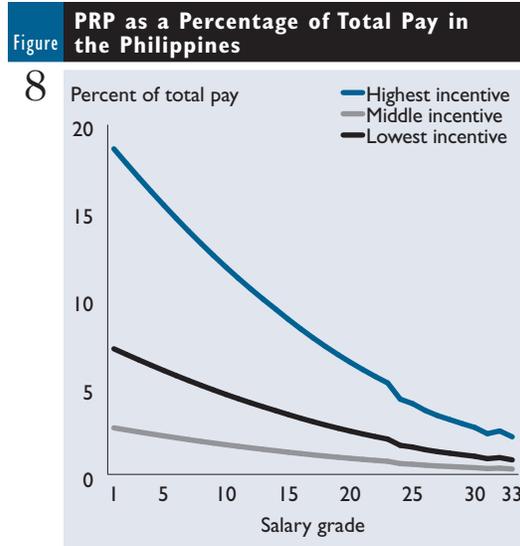
Almost all of Malaysia’s recent pay reforms have concentrated on PRP. PRP was introduced in 1992 as a merit increment scheme based on a forced distribution of assessments: only 3 percent of employees could qualify for a single PRP increment in addition to the “normal” increment, and only a further 2 percent could qualify for a double increment. (At the other end of the distribution, 5 percent of staff would not receive any increase, and would become subject to

The Philippines until recently had very little pay flexibility, with public sector remuneration approximating the single pay spine across government departments and agencies



Source: Government of Minas Gerais.

In Korea, PRP has been one of several reforms aimed at strengthening performance management, particularly in the wake of the Asian financial crisis



Source: Department of Budget and Management.

disciplinary action.) This scheme was modified substantially a decade later. A new “competence-based assessment” was introduced in response to complaints of manager favoritism by the civil servants’ union, CUEPACS. In practice, this took the form of an assessment of employees’ performance at a training workshop. Although the competence assessment constituted only 30 percent of the total assessment, in practice it counted for more, since the competence assessment was viewed as a better discriminator of performance than the managerial assessment. This competency-based assessment also eventually faced resistance from CUEPACS, this time about the length of time that its members were spending on assessment workshops, and the indignity of having to go “back to school.” The system is being changed again, and the government has set up a pay commission to review overall pay arrangements.

Malaysia has performance evaluation at the agency as well as individual level. There are National Key Result Areas; ministerial and department heads’ key performance indicators; “star ratings” of department performance conducted by the Malaysian Administrative Modernization and Management Planning Unit; reports by the auditor-general (which are reported in the press); and ISO 9000 quality ratings and client’s charters. Government departments are required to report their performance on all these measures.

In Korea, PRP has been one of several reforms, such as performance budgeting, aimed at strengthening performance management, particularly in the wake of the Asian financial crisis. Historically the pay structure had been based on seniority, but in 1999 the government introduced PRP to improve recruitment and create a more performance-oriented culture in the civil service. These PRP schemes have spread widely in the public sector, with 71 percent of public bodies using PRP. Two types of PRP schemes are currently in use: an annual merit increment program for senior officials (members of the Senior Civil Service [SCS]) and a performance bonus program for middle- and lower-level officials.

For the SCS, PRP is based on individual performance, organization-level performance based on the concerned staff’s position as a manager (in terms of citizen satisfaction with service delivery, citizen satisfaction on major public policies, and similar criteria), and job-related abilities (core competencies, customer-orientation, and so forth). The performance incentive is high, with a maximum of 15 percent of annual base pay increase to the top performers (as a merit increment, this is a permanent increase in base salary), and a 10 percent and 5 percent increase for the next two performance categories, respectively. The nature of the performance agreement depends on the type and task of the ministry. For example, policy-oriented departments have more qualitative targets, while the service delivery-related departments have more quantitative targets.

For junior staff, the minister in charge has discretion in deciding what proportion of PRP should be based on group or individual performance. The most common is a fully individual scheme, which provides for significant variations in the size of the annual bonus across individual performance categories. This scheme is currently in operation in 30 of 44 government agencies and is based on a forced distribution, with the top performers (the top 20 percent) getting an annual bonus equal to 172.5 percent of basic monthly pay (or approximately 14 percent of annual pay), the

Table 5 Key Design Features of PRP in Case Study Countries

5

Country	Individual- and group-based awards		Time horizon of the award	Nature of the performance evaluation		Size of the award	Probability of receiving the award
	Individual based	Group based		Individual based	Group based		
Chile	<ul style="list-style-type: none"> • SES only • Specific schemes for certain agencies, such as the revenue authority • General across-the-board scheme was introduced but later abandoned 	<ul style="list-style-type: none"> • Achievement of working unit goals (collective performance agreements) • Achievement of institutional goals (Management Enhancement Program); applicable to almost all staff 	<ul style="list-style-type: none"> • Annual bonus only, paid in four installments during the year 	<ul style="list-style-type: none"> • Annual performance appraisal of SES members. Evaluation led by immediate supervisor and based on targets established in individual results agreement (three years) 	<ul style="list-style-type: none"> • Based on achievement of agency and working unit targets. Internal audit unit evaluates achievement of working unit targets 	<ul style="list-style-type: none"> • From 30% to 100% of base pay for SES, based on the job • From 4% to 16% of base pay for the group awards (institutional and working unit) for all staff 	<ul style="list-style-type: none"> • Almost 100% of staff receive the individual bonus (for SES) and the institutional and working group bonuses
Brazil (federal government)	<ul style="list-style-type: none"> • All staff. Based on productivity at work, commitment, knowledge, and compliance with rules • 20% weight in overall bonus 	<ul style="list-style-type: none"> • Achievement of working unit and institutional goals • 80% weight in overall bonus 	<ul style="list-style-type: none"> • Annual bonus only 	<ul style="list-style-type: none"> • Details at the discretion of the ministry but based on a combination of self-assessment, supervisor assessment, and team member assessment 	<ul style="list-style-type: none"> • Based on goals established at the beginning of each evaluation cycle 	<ul style="list-style-type: none"> • Specific legislation for each career identifies the monetary value of each point 	
Brazil (Minas Gerais)	<ul style="list-style-type: none"> • Individual performance allowance (ADE) • The ADE depends on both individual performance assessment (70%) and achievement of institutional goals (30%) 	<ul style="list-style-type: none"> • Productivity premium. Formula based on working days without absenteeism and achievement of institutional results agreements 	<ul style="list-style-type: none"> • Annual bonus, but the size of the bonus increases over time with successive satisfactory performance evaluations 	<ul style="list-style-type: none"> • Based on a combination of self-assessment, supervisor assessment, and team member assessment 	<ul style="list-style-type: none"> • Each institution defines its action plan in agreement with the Secretary of Planning. Institution self-evaluates accomplishment through a committee with at least one member of the Secretary of Planning 	<ul style="list-style-type: none"> • Productivity premium: Maximum size of award is one month's salary • ADE: Varies with number of successive performance evaluations, ranging from 5% to 70% of base pay 	<ul style="list-style-type: none"> • 96% of staff receive a satisfactory rating; 74% receive the highest rating; and 22% the next highest rating
Korea, Rep.	<ul style="list-style-type: none"> • Two types of schemes, one for members of the Senior Civil Service (SCS) and one for all other staff 	<ul style="list-style-type: none"> • For SCS, agency performance is one element of the staff performance evaluation • For all other staff, the minister in charge has discretion in deciding the mixture of group and individual award 	<ul style="list-style-type: none"> • Merit increment for senior staff; annual bonus for junior staff 	<ul style="list-style-type: none"> • For SCS: appraisal based on agreement between agency head and heads of working units and includes individual and organizational performance- and job-related abilities 		<ul style="list-style-type: none"> • For SCS: Merit increment as high as 15% of annual base pay • For other staff: Bonus as high as 14% of annual base pay 	<ul style="list-style-type: none"> • Forced distribution for fully individual schemes: 20% of staff get the highest bonus • 30 government agencies have fully individual bonus schemes; none has a full group bonus and equal individual bonus scheme; and 14 have schemes combining individual- and group-based bonuses
Malaysia	<ul style="list-style-type: none"> • All staff from 1992 to 2012. PRP suspended at time of writing, with possibility of reintroduction for senior staff only 	<ul style="list-style-type: none"> • All staff have received one month's pay bonus in recent years (Inland Revenue: 5 months, based on performance) 	<ul style="list-style-type: none"> • Merit increment 	<ul style="list-style-type: none"> • Performance assessment scheme was later modified to a competency assessment scheme, and then later suspended 			<ul style="list-style-type: none"> • Forced distribution with only 2% of staff receiving a double merit increment
Thailand	<ul style="list-style-type: none"> • All staff 	<ul style="list-style-type: none"> • None 	<ul style="list-style-type: none"> • Merit increment 			<ul style="list-style-type: none"> • Most employees receive a pay increase of less than 3% 	<ul style="list-style-type: none"> • Most employees receive something
Philippines	<ul style="list-style-type: none"> • All staff 	<ul style="list-style-type: none"> • Department- and unit-level performance 	<ul style="list-style-type: none"> • Annual bonus 	<ul style="list-style-type: none"> • Based on traditional, "trait-based" individual performance assessment 	<ul style="list-style-type: none"> • Departments/agencies need to meet 90% of targets and 100% of good governance conditions. Working units are then evaluated against their targets 	<ul style="list-style-type: none"> • Varies from 1% to 18% of total pay 	<ul style="list-style-type: none"> • Forced distribution: 1% of staff receive the highest bonus
Indonesia	<ul style="list-style-type: none"> • Bureaucracy Reform (BR) allowance conditional on staff attendance 	<ul style="list-style-type: none"> • Revenue authority: group award based on tax collections 	<ul style="list-style-type: none"> • Monthly allowances for individual BR pay; three bonus payments for the revenue authority scheme 	<ul style="list-style-type: none"> • Fingerprint machines are used to measure attendance 	<ul style="list-style-type: none"> • Tax collections of the tax office 	<ul style="list-style-type: none"> • A month's BR allowance is cut if staff fail to meet minimum attendance requirements 	<ul style="list-style-type: none"> • All staff in the agency get the BR allowance

Source: Urbanization study team.

In Chile key government agencies departed from the former single pay scale in the late 1970s

next category receiving 125 percent, the third category receiving less than 85 percent, and the bottom category (the lowest 10 percent) getting no bonus. In departments where individual performance is hard to evaluate objectively, the minister can decide to distribute the individual bonus equally, with the size of the bonus based purely on departmental performance. However, to date no government agency has opted for this variation.

In Thailand, PRP was a major reform of the Thaksin government and mandated by the Civil Service Act 2008. Under the scheme, in each half-year cycle employees can receive a salary increase of up to 6 percent. In theory, an employee could get 12 percent in a year, following the manager's assessment of progress against the key performance indicators that employees are supposed to produce at the start of every year. Because the PRP budget is capped at 3 percent for each cycle (6 percent for the year), few employees in practice receive 6 percent in a cycle, let alone 12 percent over a year. In the first year of operation, most employees received a pay increase of just below 3 percent.

The 2008 procedure was preceded by an elaborate job regrading and simplification exercise, so that managers had a framework of grades and job descriptions to use as the basis for their PRP assessments. The public service's 225 job classes were drastically simplified to four broad job categories: executive, management, knowledge, and general.

Indonesia does not have across-the-board PRP in central ministries and agencies. However, some agencies, particularly those that receive an additional allowance, have introduced fingerprint machines to record attendance and hours at the office, and these additional allowances are cut if staff are absent without permission or do not work the required number of hours. A PRP scheme in the revenue agency gives a group bonus for achievement of property tax collection targets by relevant units. The bonus is given three times a year and can be between one and four times monthly basic pay, depending on the amount of tax revenues collected by the tax office.

Differentiation

In Brazil, Chile, Indonesia, and Russia total pay for a civil servant in a similar job varies across public entities based on the government agency, location, and the specifics of individual contracts.

In Chile the “controlling institutions”—among them the Superintendence of Pensions; Superintendence of Health; the central public procurement office, ChileCompra; the Unit for Financial Analysis; and the Defense of the Competition Agency—and the “finance sector institutions”—the Treasury, Budget Office, and revenue authority—have separate pay scales that are higher than those of other public entities. These agencies departed from the former single pay scale in the late 1970s, since they were viewed as key government agencies in charge of raising revenues, of regulating, and of exercising control over the public sector. In addition, most ministries and agencies receive a “critical functions” allowance of up to 100 percent of their remuneration from the Budget Office to distribute among few staff, with agency heads deciding who receives the benefit and how much selected staff receive. This allowance is normally used to increase the pay of key middle-management positions (for example, a Budget and Planning Division chief for a line ministry) and a small number of staff. Moreover, approximately 60 percent of public sector employees are contractual (*contrata*) staff with individual contracts determined by the agency head in which pay levels can be set at any point of the pay scale within limits determined by the Budget Office.

Indonesia has a complex structure for civil service pay with numerous allowances that result in significant differentiation of pay between staff doing similar jobs. For staff in central government ministries, cash compensation (excluding in-kind allowances for car, housing, and utilities) has three main elements: basic salary, allowances, and honoraria. Basic salary and allowances are considered a “fixed” or guaranteed allocation that staff receive regardless of the type or level of work that they are engaged in. Honoraria are, at least on paper, given to staff for taking on activities and responsibilities above and beyond their regular duties.

A major source of differentiation, and the focus of this study, is the significant variation in the levels and structure of pay between the few central ministries and agencies that are deemed to be strategically important or undergoing internal reforms. This reform exercise is called Bureaucracy Reform (BR). The BR agencies—as of 2012, 14 of 76 total central government ministries and agencies—include the Ministry of Finance, the audit institutions, the planning agency, and the personnel agency. They have additional, supplemental salary scales, though the BR pay scales have been largely determined in an ad hoc manner and differ across agencies, with seven separate BR pay scales in place in BR agencies.

Average total monetary compensation in BR agencies is between two and over four times that of non-BR agencies, depending on staff seniority (figure 9). Pay is also more decompressed in BR agencies, which means that pay increases quite significantly for staff over their career. Moreover, BR pay is based on a job evaluation–based parallel grading structure, so staff of similar seniority can be assigned to different grades, with the result that staff with the same seniority can earn very different BR allowances within the agencies. For example, there can be a sixfold difference in the BR allowance of the senior-most staff, leading to a significant variance in total pay.

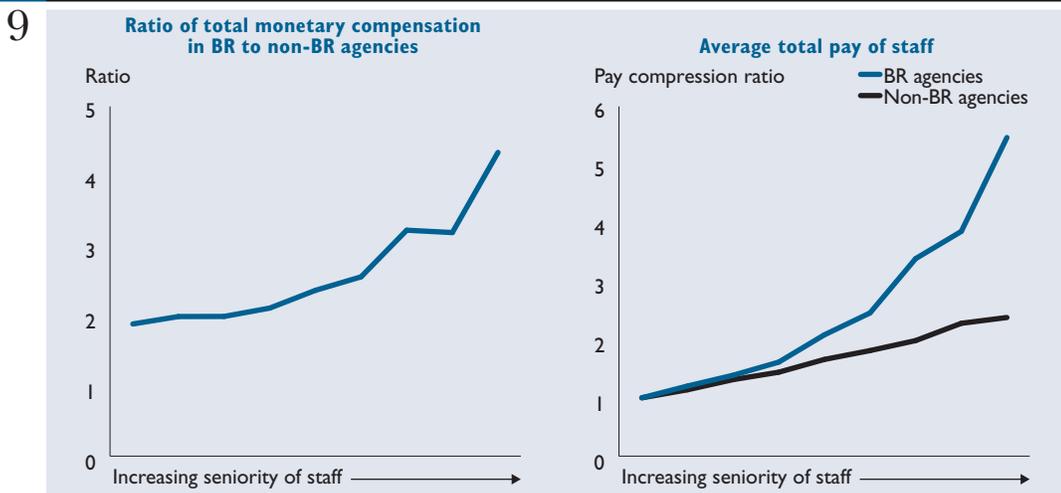
In Brazil considerable pay differentiation across careers exists at both the federal and state levels. In recent years the number of careers has grown significantly, with each ministry demanding its own set of identifiable careers to validate its status in the bureaucratic pantheon. As of 2011 the federal government had 161 careers and similar professional structures, each with its own salary structures, usually varying in the size and terms of allowances and administrative arrangements, a framework that is replicated at the state level. This differentiation is instituted through legislation specific to each position or career group, often through intensive bargaining with employee unions. One noteworthy source of differentiation is “subsidy” remuneration for 32 priority careers, a unified compensation package that guarantees salary increases for a specified period (and is thus not subject to periodic negotiations) and makes all compensation pensionable.

The difference in pay between the federal and state levels is also significant. States pay salaries significantly lower than their federal equivalents, since they are constrained by the Fiscal Responsibility Law of 2000, which places a ceiling on state government wage expenditures.

Figure 10 shows the variation in pay between careers in the federal government, variation for a particular career between the federal

As of 2011 Brazil’s federal government had 161 careers and similar professional structures, each with its own salary structures, usually varying in the size and terms of allowances and administrative arrangements

Figure 9 Pay Differentiation in Indonesia

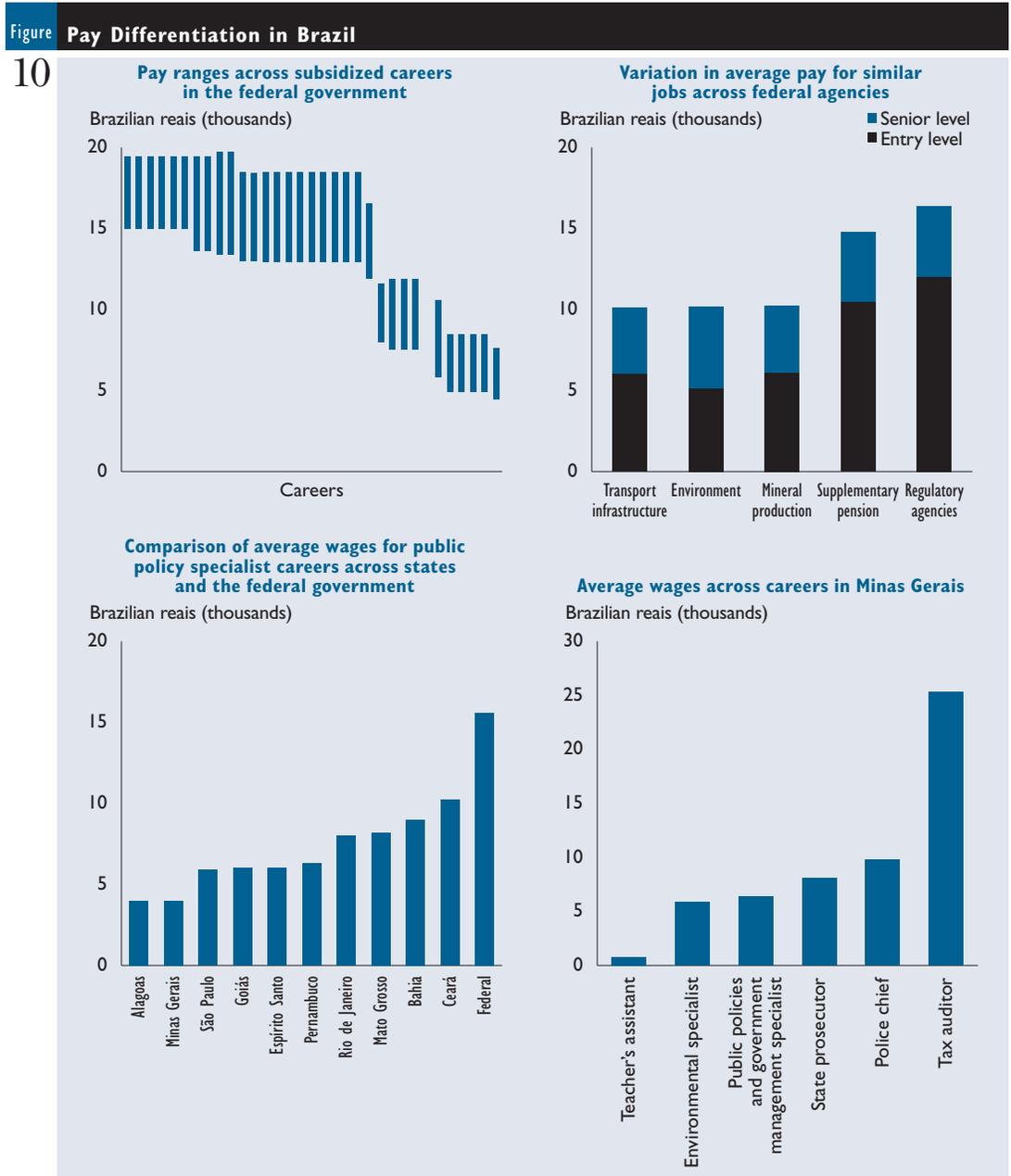


Note: The pay compression ratio is the salary of a particular staff as a multiple of the lowest paid staff. Source: World Bank staff calculations based on Ministry of Finance data.

government and the states, and variation in pay across careers in the state of Minas Gerais.

Pay flexibility in Russia consists largely of pay differentiation and significant use of PRP in the service delivery sectors. For the core civil service, there is significant variation in pay levels across federal ministries, between federal civil servants working in Moscow and those in the regions, across subnational administrations, and across staff in similar jobs

(figure 11). In federal ministries average pay can be three times higher in ministries responsible for policy development and regulatory oversight compared with agencies undertaking service delivery and supervision, and much of this variation is at the senior manager level. Average compensation in territorial authorities is less than half the average compensation in central authorities for similar jobs, with the gap larger in managerial jobs.¹²

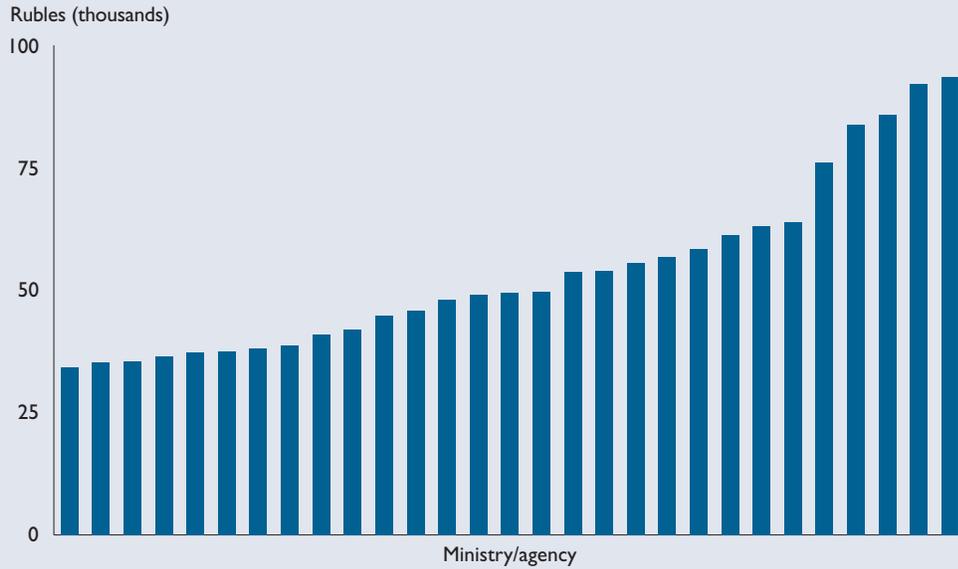


Sources: Ministry of Planning; Andrade 2011; Minas Gerais Civil Service Directorate.

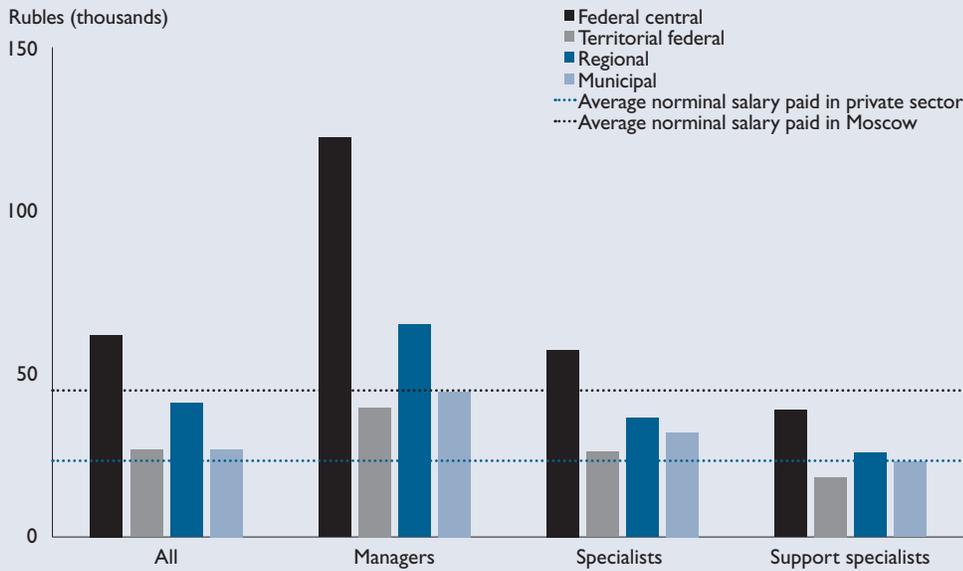
Figure 11 Pay Differentiation in the Russian Federation

11

Average monthly salary across federal ministries



Average salaries across levels of government



Source: Rosstat and HSE Institute for Civil Service Development data.

Assessing the Evidence from the Case Studies: Performance-Related Pay

In coping jobs the tendency has been for PRP to degenerate into a de facto salary supplement

Direct pay flexibility levers

The case studies reinforce the main finding from the literature review that PRP can have a direct incentive effect in craft jobs for which performance, either individual or group, can be measured with objective data. By contrast, in coping jobs, for which performance assessments are more subjective, the vast majority of staff tend to get high ratings, with the result that PRP degenerates into a de facto salary supplement. Some countries have tried to address this problem by requiring mandatory distribution of performance ratings, but that solution is not without implementation challenges and it depends heavily on the level of trust and general functionality of human resource management in the civil service. Our findings are that in low-trust environments mandatory distributions can have a negative impact on staff morale and performance.

Craft jobs

Brazil provides several instances of productivity improvements in the police and the revenue authority. In the state of Minas Gerais two of the performance targets for the police were for weapons seizures and police operations, both of which increased dramatically after the introduction of the performance incentive, the former by nearly 60 percent and the latter by almost 50 percent. De Assis (2012) found that staff attributed this increased activity to the introduction of agreed-upon institutional goals and the accompanying PRP financial incentives.

Revenue administration has been a priority area for results-based management and PRP

in Brazil, both at the federal and state levels. The federal PRP scheme for tax collectors was introduced in the late 1980s in the context of severe fiscal pressures and the need to raise tax revenues and was the subject of the empirical study by Kahn, De Silva, and Ziliak (2001) reviewed earlier. This individual- and group-based scheme awarded a bonus to tax inspectors based on individual evaluations by supervisors as well as the collection performance of the local tax agency based on fines collected and achievement of other targets (total tax collection, number of inspections, collection of overdue taxes). The strong rebound in revenues—significant increases in revenue collection and individual productivity, as measured by a 75 percent increase in fines per inspection—was attributed in part to PRP. Despite its apparent success, this federal scheme was not politically sustainable, and under pressure from the employee unions was later abandoned and replaced by a high, fixed-pay system.

In Minas Gerais the Secretariat of Finance introduced an individual- and group-based PRP scheme in 2005. The group award is in large part based on attainment of revenue targets, with staff attendance required to participate in the bonus. The individual award was based on the achievement of goals that are linked to a cascade of results agreements. These results agreements have grown in sophistication over time in recognition of the perverse incentives that a pure focus on revenue collection can engender. In this case, staff had come to focus on increasing short-term revenues even though building enduring

relationships of trust with potential taxpayers was a more effective way to broaden the tax base on a sustained basis. Individual evaluations now increasingly focus on operational priorities (for example, concentrating efforts on large taxpayers) and also include taxpayer facilitation and relationship-building variables such as frequency of site visits.

In Chile, PRP was viewed quite favorably in the revenue authority and the Civil Registry. The Civil Registry's PRP scheme is based on net customer satisfaction as measured by a survey conducted by an independent source, normally a university or consulting firm selected by competitive bid. The survey asks about wait times and citizens' experiences with online services. Within the agency's customer satisfaction unit there is a perception that the bonus has helped improve staff performance.

In Indonesia the revenue authority—the Directorate General of Taxation (DG Tax) of the Ministry of Finance—is the only government agency that has introduced an additional performance bonus (on top of linking BR pay to staff attendance, as in the BR agencies). This bonus is based on achieving property tax collection targets and goes to all staff of the tax office. In the World Bank survey of government officials, DG Tax ranked highest on the staff effort question (“whether others in the agency contribute more than is expected of them”) (figure 12). These differences among the four BR agencies surveyed—DG Tax, the Directorate General of Treasury (DG Treasury) in the Ministry of Finance, Bappenas (the national planning agency), and MenPAN (the national oversight agency on personnel matters)—suggest that factors other than the BR allowance may have an impact, and the differences between the two Ministry of Finance agencies suggest that factors specific to DG Tax, such as its performance bonus, may be important, a point that was also emphasized in the expert interviews.

In Malaysia the only positive accounts of PRP come from the revenue authority, where interviews with management revealed a perception that staff effort is linked to the corporate targets of increased revenue collections. This sanguine management view is qualified



Source: World Bank survey.

by the view of the authority's own staff, who were generally neutral about PRP's impact on staff effort. The revenue authority is also a rare example of an agency that enjoys substantial human resource management autonomy in Malaysia and where executives were motivated and had the authority to create a new work culture. PRP is thus part of a package of human resource management reforms that has succeeded in creating a performance culture in the authority.

The case studies of PRP suggest that in craft jobs a combination of group and individual bonuses can be effective and that worries about free-riding in group bonuses may be exaggerated. The size of the bonus in many cases was significant, for example, in the case of Brazilian tax officials. Another interesting finding is that managers, at least in some cases, were aware of the risks of gaming and had evolved performance assessments to minimize it, as exemplified by the revenue authority in Minas Gerais.

Coping jobs

The case studies revealed several challenges in implementing PRP for coping jobs that imply that the financial incentive had little direct positive effect, and in some cases possibly negative effects, on staff effort. By definition, these are

In Chile, PRP was viewed quite favorably in the revenue authority and the Civil Registry

In Brazil, PRP had some effect on the extremes of the performance distribution by helping in disciplining blatantly incompetent staff

jobs in which outputs cannot be easily measured and therefore the performance evaluations that form the basis of the financial incentive are either based on subjective evaluations by supervisors and review panels or on some quantitative input or process measures. Given these difficulties in performance assessment, the “steady state” in most bureaucracies is for the vast majority of staff to be given a best or next best performance rating in performance evaluations, with the result that the performance bonus is given with close to probability 1 and so cannot have a direct incentive effect. Why most bureaucracies converge to this norm is a question that cannot be easily answered here. One reason may be the likelihood that staff will be working with each other for a long period of time, in contrast to the private sector, where low turnover encourages a nonconflictual attitude.

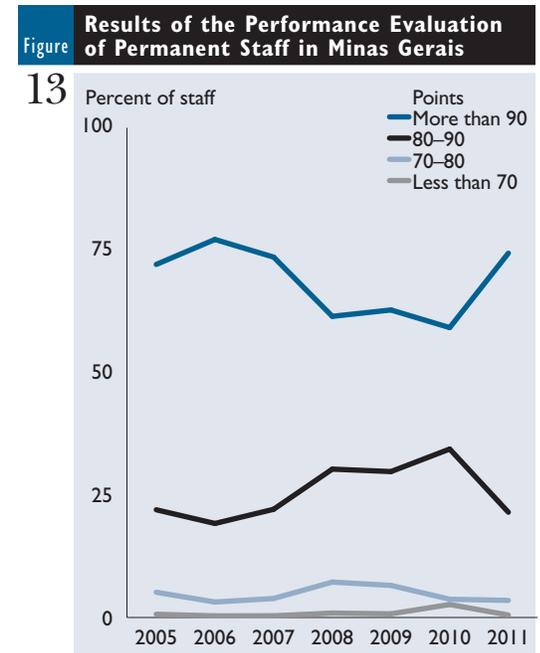
Chile, despite its reputation as having an efficient, performance-oriented public administration, abandoned a government-wide individual PRP scheme after five years because managers were rotating the award, with a third of the staff getting the highest bonus every year. Similarly, in the current PRP scheme for senior managers almost all SES members receive the complete bonus, and PRP has therefore become a salary supplement rather than a performance incentive. Interestingly, despite this high bonus, the turnover rate among the SES cadre has averaged 50 percent over the past decade, reflecting the difficulty of attracting and retaining senior talent in a highly dynamic labor market.

In Thailand most of the staff who were interviewed felt that PRP had only marginally improved work effort because of the small size of the performance bonus (the bonus budget is 3–6 percent of the total salary budget), the fact that most staff got the bonus, and the influence of nonmonetary motivators such as sense of duty. Some staff suggested that the effect would be bigger if the amount was higher. While staff have made the usual complaints about the weak links between pay and performance, overall they support the scheme. The government has recognized that the 6 percent ceiling on PRP increases is a constraint and is planning to increase it.

The state government of Minas Gerais in Brazil has also received considerable international attention for its performance orientation, but it has similarly found it difficult to use annual performance appraisals to meaningfully distinguish among staff. The overwhelming majority of civil servants receive satisfactory ratings (more than 70 points), though there is some consideration to differentiating between the top performers (more than 90 points) and those who score above the threshold but are not outstanding performers (between 80 and 90 points). This performance distribution has not fundamentally changed since 2005, when the individual performance bonus scheme was introduced, with the top two performance ratings given to almost 90 percent and 20 percent of staff, respectively (figure 13). Individual PRP in Minas Gerais is therefore also largely a salary supplement.

Interviewees in Brazil did note that PRP had some effect on the extremes of the performance distribution by identifying the outlier staff and helping in disciplining and even dismissing blatantly incompetent staff. PRP may also have reduced the number of leave-takers and discouraged extended absences from service.

Similarly, in Indonesia the World Bank survey of government officials found that the



Source: Minas Gerais Civil Service Directorate.

linking the supplemental pay in the BR agencies to staff attendance may have reduced the proportion of staff leaving work early. In the BR agencies 11 percent of staff reported that more than 20 percent of their coworkers in the agencies left work early or spent excessive time on personal matters, far fewer than in non-BR agencies, where 30 percent of the staff reported that their coworkers were similarly delinquent (figure 14). This difference is quite striking and cannot be fully explained by the proliferation of fingerprint attendance recording machines in ministries and agencies.¹³

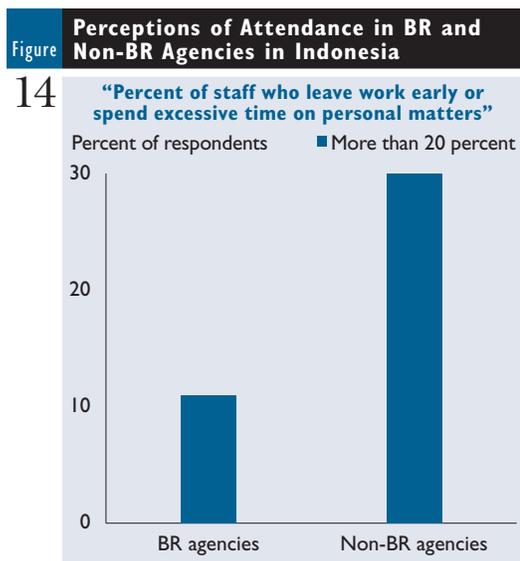
To some extent, the use of PRP to sanction only the outlier staff makes sense, since they can be easily identified and justified to most staff as deserving special treatment. By contrast, without objective performance measures, it is very difficult for supervisors to credibly discriminate between the achievement of the majority of the staff who are in the “fat end” of the normal distribution of performance, and any significant variations in pay for this group could easily be viewed as unfair and breed resentment among staff.

Some countries, such as Korea, Malaysia, and the Philippines, have tried to counter this tendency of uniformly high performance ratings and equal distribution of the performance bonus by mandating forced distribution of performance ratings. This policy is risky, and its efficacy depends very much on the legitimacy

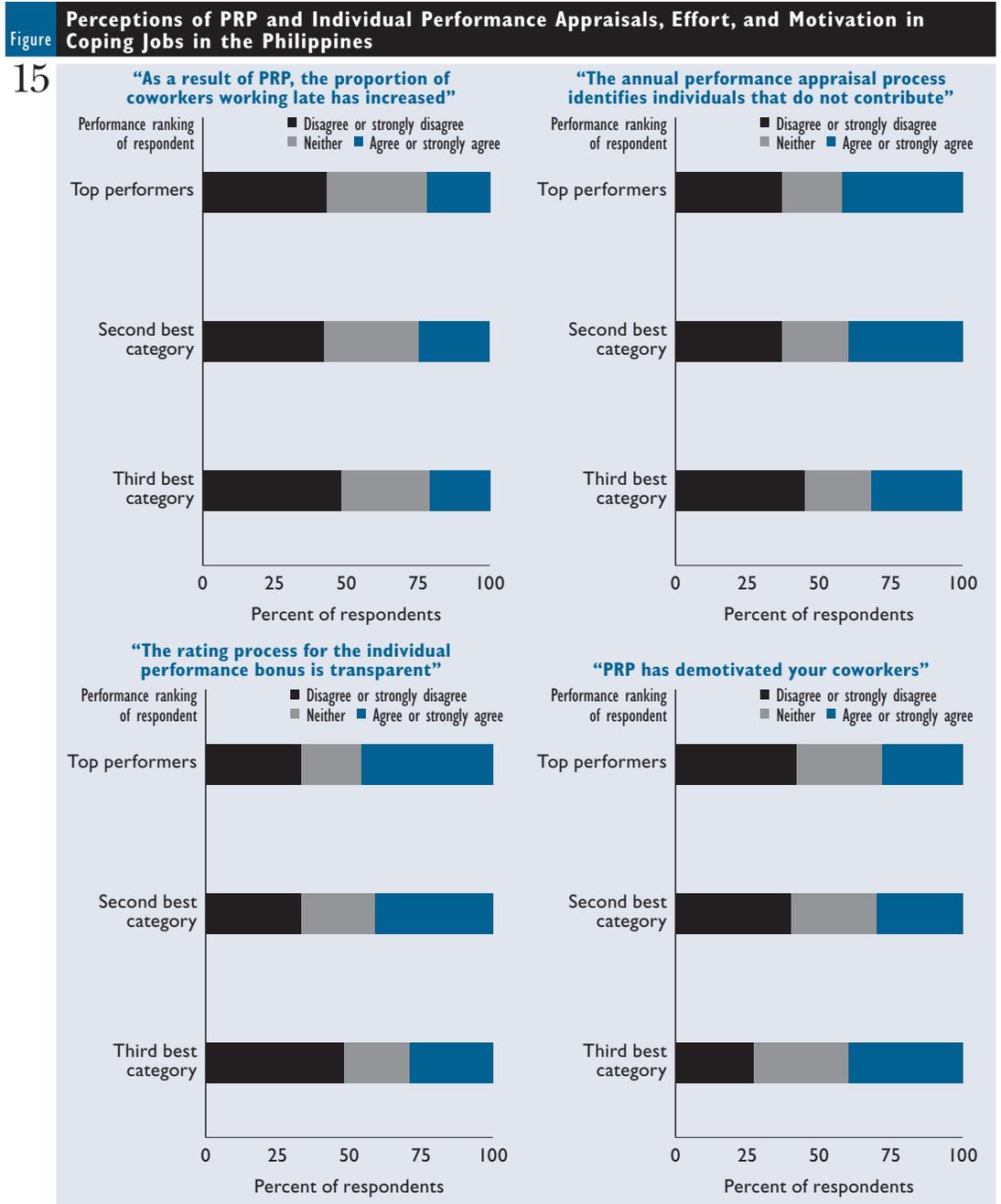
that performance appraisals enjoy among staff. In Malaysia the forced distribution introduced in 1992 was abandoned 10 years later under pressure from the civil service union following complaints of manager favoritism. It could in fact be argued that this policy did harm staff morale, teamwork, and the relationship between management and staff. Similarly, in the Philippines many of the staff interviewed complained about individual forced ranking as being highly subjective and unfair and potentially hurting morale. Even in Korea, generally regarded having as a high-performing civil service, studies have noted significant differences in perceptions of PRP between central government and local government staff, with PRP in the former being viewed as relatively well implemented and in the latter viewed generally negatively and not supported by staff (Han 2010; Lee 2010). In less accomplished civil services, mandating performance distributions can risk hurting individual and agency performance.

The World Bank survey of government officials in the Philippines revealed that staff in coping jobs clearly did not believe that the performance bonus had any positive effect on effort, as indicated by the unanimity in their disagreement with the statement that the proportion of coworkers working late had increased (with between 42 and 48 percent disagreeing or strongly disagreeing) (figure 15). Staff views diverged based on individual performance rankings, and therefore the size of the bonus, on the effectiveness of the individual performance appraisal process (between 37 and 35 percent disagreeing or strongly disagreeing that it identified individuals who did not contribute); the transparency of the individual performance rating (between 29 and 46 percent agreeing or strongly agreeing that it was transparent); and the impact of the incentive on staff morale (between 28 and 40 percent agreeing or strongly agreeing that it had demotivated their coworkers). The bottom-ranked performance category, which comprises the biggest group in the distribution, was quite clear that the rating process was not transparent and that the performance bonus had demotivated their coworkers.

Without objective performance measures, it is very difficult for supervisors to credibly discriminate between the achievement of the majority of the staff who are in the “fat end” of the normal distribution of performance



Source: World Bank survey.



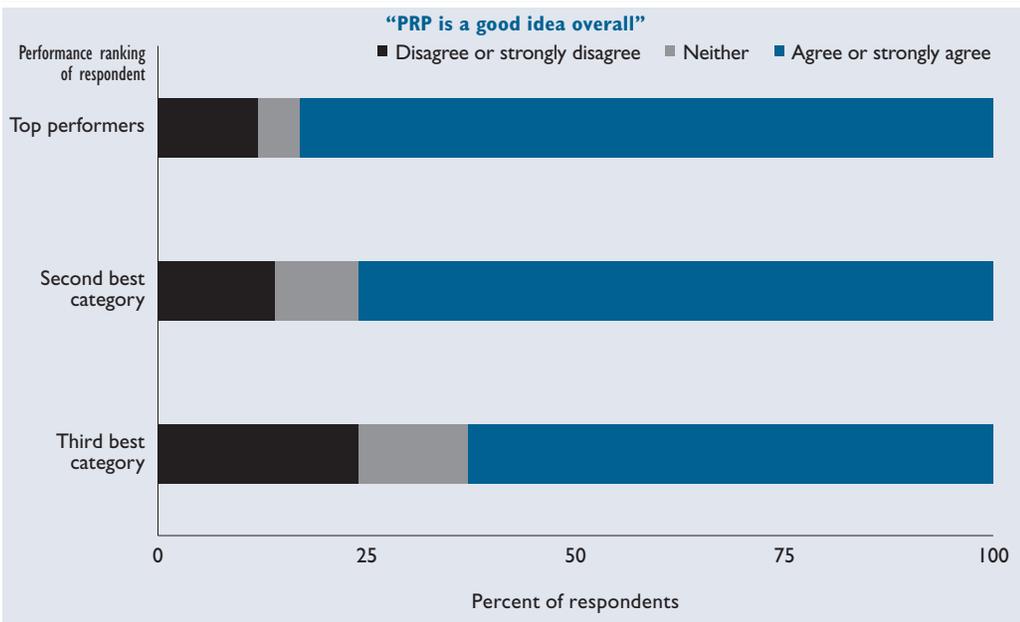
Source: World Bank survey.

These perceptions suggest that the direct incentive effects of PRP in the Philippines are weak and possibly negative. Yet the overall view of the scheme among staff is positive, with even a majority of the bottom-ranked staff believing that it is a good idea (figure 16), suggesting that other factors might be at play, as explored below.

For coping jobs, group-based PRP linked to working unit and institutional goals is not

without its own complications. On paper at least, group-level bonuses can be more objective and based on agency and working unit outputs and therefore potentially more acceptable and more likely to have a direct incentive effect. For policy units, as opposed to service delivery units, however, these performance indicators tend to be process oriented and therefore more vulnerable to manipulation and gaming.

Figure 16 Overall View among Civil Servants of the PRP Scheme in Coping Jobs in the Philippines



Source: World Bank survey.

In the Philippines, for example, the desire to use quantitative output indicators (which in the Philippines are called major final outputs) for all agencies has led to some sub-optimal choices. The performance indicators for the Department of Budget and Management include the number of advisories issued (counting the number of documents) and the quality and timeliness of these advisories, in contrast to the output indicators of service delivery departments such as roads and education (table 6). Similarly, another performance

indicator measures the ratio of authorized to received requests from departments for various budget and organizational change proposals. These indicators are vulnerable to gaming (for example, splitting one advisory into two in order to increase the output) and have unclear links with performance. By contrast, the service delivery departments have more measurable and independently verifiable output indicators.

Gaming was a major concern in Chile's group-based bonus schemes, since indicators

Table 6 A Sample of Performance Indicators for the Group-Based Bonus Scheme in the Philippines

Department	Performance indicators
<i>Policy and oversight departments</i>	
Department of Budget and Management	<ul style="list-style-type: none"> • Number of advisories and directives issued • Number of agencies whose spending capacity was evaluated • Release of funds for priority expenditures within five days after the receipt of approval from President
Housing and Land Use Regulatory Board	<ul style="list-style-type: none"> • Policies on housing, land use planning, and real estate development reviewed or conducted within the year • Proportion of local land use committee members trained • Percent of mediation conferences conducted within 60 days over total number of complaints filed
<i>Service delivery departments</i>	
Department of Public Works and Highways	<ul style="list-style-type: none"> • Number of road maintenance projects • Number of projects paving unpaved roads • Percent of projects completed in accord with plans and specifications
Department of Education	<ul style="list-style-type: none"> • Improvement in proportion of grade 6 pupils' test scores • Improvements in school graduation rates • Proportion of private schools with permit to operate or acquired recognition

Source: Data from the respective departments.

PRP can encourage managers to foster more teamwork among staff and to improve the dialogue between managers and staff

were mainly process oriented, self-imposed, and self-evaluated, resulting in the deliberate setting of easy-to-meet targets. The Budget Office, which was charged with the review process, could not feasibly control gaming behavior in more than 200 institutions. Moreover, staff unions put considerable pressure on agency authorities to ensure that targets were achievable and no portion of the variable pay was at risk. Practically all Chilean public sector institutions have consistently achieved their performance indicators, indicating that PRP has served as a more politically viable way to grant pay increases rather than as a direct incentive for performance improvement.

Similar problems of gaming were found in the group-level schemes in Brazil. Some institutional goals received inordinate attention, drawing energy away from other essential tasks. Other organizations gamed goals by establishing low bars that the agency could not fail to meet—for example, the Ministry of Environment, which attained 300 percent of some targets.

There is some evidence from the case studies that PRP may have had a direct incentive effect of sorting. Inspired by Chile, the government of Minas Gerais instituted a Public Entrepreneurs Program in 2007 that created 90 positions, open to recruitment from the private sector and within government, whose job description was to help implement achievement of government goals. These public entrepreneurs were charged with championing statewide reforms across the public sector and were paid substantially on a performance basis. Half the public entrepreneurs came from the private sector, and recruits were generally considered to be of high caliber. The methods of recruiting this cadre were also more stringent than for the rest of the public service and included an oversight committee and an evaluation committee. As of 2009 the program was operating in 13 different ministries, and 25 percent of the positions were managerial or other strategic positions.

In sum, the case studies show that PRP has a limited direct incentive effect in coping jobs because the probability of receiving the incentive is close to 1, either because managers

refuse to distinguish between staff in their performance appraisals or targets are manipulated so that most are achieved. The case studies do not suggest that PRP caused a decline in intrinsic motivation or that a lack of pay transparency caused problems, with the notable exception of when it was implemented through a forced distribution. In most cases the scheme defaulted to a salary supplement, which in some cases did have a positive effect on recruitment and retention (as in Minas Gerais) and in others did not (Chile's SES). Insistence on objective, measurable indicators for these jobs led to an abundance of input and process indicators, which resulted in some perverse consequences and gaming behavior, though it is unclear what effect gaming had on agency productivity. One can speculate that this plethora of indicators did cause some harm through performance measurement fatigue and skepticism of public employees and the creation of an army of employees in each agency to ensure that the gaming was correctly executed and no staff's pay was at risk.

Indirect pay flexibility levers

PRP can theoretically improve individual and agency productivity through another, more indirect channel by encouraging gradual changes in management that in turn could improve individual employee performance. The three relevant management practices are goal setting, teamwork, and individual performance assessments. PRP can give managers incentives to shift an organization's focus to results rather than inputs through the discipline of setting organizational targets and regular monitoring of progress. It can encourage managers to foster more teamwork among staff and to improve the dialogue between managers and staff. And it can improve the individual performance appraisal process by better linking individual results agreements to organizational goals. All of these practices help inculcate a "performance culture" or "mission orientation" within the agency.

The emphasis here is on gradual and evolutionary change. The usual path is for coping agencies in the first stages of PRP to have a surfeit of input and process indicators, which

leads to the problems mentioned. Over time, however, these indicators can be refined and can approach, even in coping jobs, a degree of output focus—through, for example, the more frequent use of perception surveys among service users (which are normally other government agencies).

In Chile the government that took office in 2010 was well aware of the challenges of implementing PRP and made efforts to modify the performance management system through better control from supervising ministries. This agenda included setting and evaluating performance targets for the agencies within their sector, giving more weight to actual results than to processes, and encouraging agencies to set higher targets by recognizing partial completion of targets as opposed to recognizing full achievement or no achievement.

In Korea the general perception is that PRP helped improve the quality of individual and organizational performance appraisals over time, complemented performance budgeting reforms, and has now achieved legitimacy and support within the central government. In the initial years the system was viewed as very lenient, particularly for senior officials, who received uniformly high ratings. Pressure from the National Assembly led to the introduction of forced distribution of ratings in 2009. Forced distribution appears to have been implemented reasonably well, at least for senior civil servants in the central government, particularly after organizational performance measures were linked to individual evaluations of senior civil servants. Government officials also perceived PRP as a tool to supplement their own salaries and to help managers manage more effectively. The success of this scheme can be gauged by the fact that the majority of government agencies (30 of 44) have opted for the individually differentiated forced ranking scheme and none has opted for the purely group-based bonus scheme with equal individual payments for all staff within the group.

PRP in the state of Minas Gerais in Brazil was part of a package of reforms introduced between 2003 and 2010 to institutionalize a results focus in the state administration. These reforms included administrative restructuring,

public expenditure reforms, and the performance contracting through cascading results agreements, with the whole process coordinated by the center through the newly created Secretariat for Human Resources and Management (SEPLAG) reporting to the governor. PRP was developed in the context of a results-based management framework, and, as discussed, consisted of both individual- and group-based bonuses.

In interviews, staff concluded that PRP was integral to the performance agenda in Minas Gerais. Staff reported that PRP had helped clarify expectations and individual goals and targets, and that these results were due to changes in the planning process, which led to a self-reported boost in morale and a performance-oriented culture and better results on the job.

Another notable aspect of the reforms in Minas Gerais is that the package of results-based management and PRP—the data do not permit distinguishing between the two on this point—encouraged greater delegation of human resource management authority to the implementing agencies. The most significant feature of this autonomy was the ability of agencies to change their organizational structure and hence staff positions and functions without prior approval from the central finance and personnel agencies, provided these changes were budget neutral. This autonomy was granted by SEPLAG on the condition of performance: agencies could continue to function autonomously if they performed satisfactorily and met at least 60 percent of their targets. If they fell below this threshold, autonomy would be withdrawn until their performance improved. This is a good example of how results-based management, of which PRP was an integral component, opened the door for other human resource management reforms, with SEPLAG realizing that this autonomy was necessary to make results-based management work.

In the Philippines there are signs that the PRP scheme is giving managers incentives to take results-based management more seriously. The Philippines has had an agency performance framework called the Organizational

In Minas Gerais
results-based
management and PRP
encouraged greater
delegation of human
resource management
authority to the
implementing agencies

In Philippines, staff surveyed believed that PRP had motivated management to increase its focus on target setting and monitoring and to engage staff in the process

Performance Information Framework (OPIF) in place since 2007, but agencies have to date viewed OPIF largely as a documentation exercise with almost no budgetary implications. Group-based PRP, which uses OPIF indicators, has stimulated efforts to improve OPIF performance indicators and to create a system for reviewing progress in achieving the targets, both within agencies and by the Department of Budget and Management.

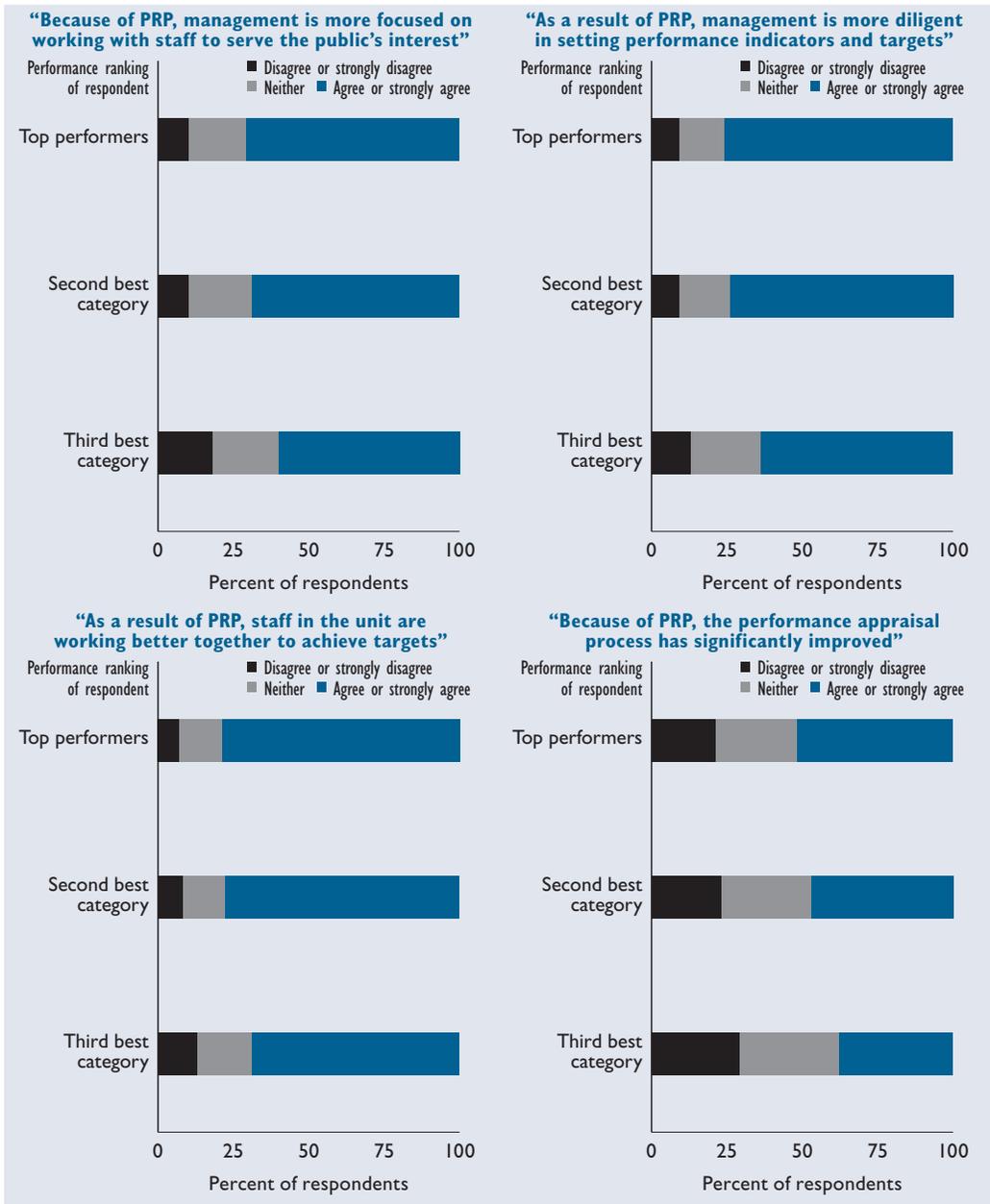
The World Bank survey of government officials in the Philippines asked questions to explore the hypothesis that PRP can improve management practices in the three areas of goal setting and monitoring, better teamwork, and improvements in the individual performance appraisal process. Staff perceptions reveal positive responses to each of these questions that cut across individual performance rankings (figure 17). Interviewees noted that the performance bonus scheme has motivated management to increase its focus on target setting and monitoring and to engage staff in the process. Staff across the performance spectrum strongly believed that management was more focused on working with staff to serve the public interest (between 60 and 78 percent agreeing or strongly agreeing) and more diligent in goal setting and in monitoring accomplishment against goals (between 64 and 76 percent agreeing or strongly agreeing). They were similarly clear in their views that teamwork in achieving departmental performance targets had improved as a result of PRP (between 69 and 79 percent agreeing or strongly agreeing), a surprising finding since forced rankings might be expected to create harmful competition among staff. Staff believed, though less strongly, that PRP had

triggered improvements in the performance appraisal process (between 38 and 52 percent agreeing or strongly agreeing). Taken together, the evidence suggests that PRP has inculcated management improvements that can over time help create a performance culture in the bureaucracy.

These positive findings from the survey may explain why the performance bonus scheme is viewed generally favorably in the Philippines despite the concerns about its impact on individual morale discussed earlier. It suggests that the group bonus and the individual bonus in the Philippines work in opposite directions: the individual bonus has a negative direct incentive effect on effort, while the group bonus has a positive indirect effect through better management. It is too early to tell what the overall effect on government performance will be.

In Malaysia, PRP has had a much more problematic history, but the overall assessment is that a culture of performance has gradually come into being over and above the outcomes of particular performance reforms and without minimizing the vestiges of patronage that survive. In this analysis the enduring significance of PRP and the other performance reforms is that they have reinforced the message that the government as an employer wants employees to raise their game. As in Minas Gerais, PRP in Malaysia has complemented results-based management reforms that have received a fair amount of international attention. However, the effect in Malaysia has been much weaker, possibly because the use of the forced distribution caused a severe reaction from the civil service unions and may have further weakened managers' already weak incentives to distinguish between the performances of staff.

Figure 17 Perceptions of the Effect of PRP on Goal Setting, Monitoring of Targets, Teamwork, and Performance Appraisals in the Philippines



Source: World Bank survey.

Assessing the Evidence from the Case Studies: Pay Differentiation

In China the Internal Revenue Service can better compete in the market for the best talent coming out of local universities thanks to its higher pay scale

Direct pay flexibility levers

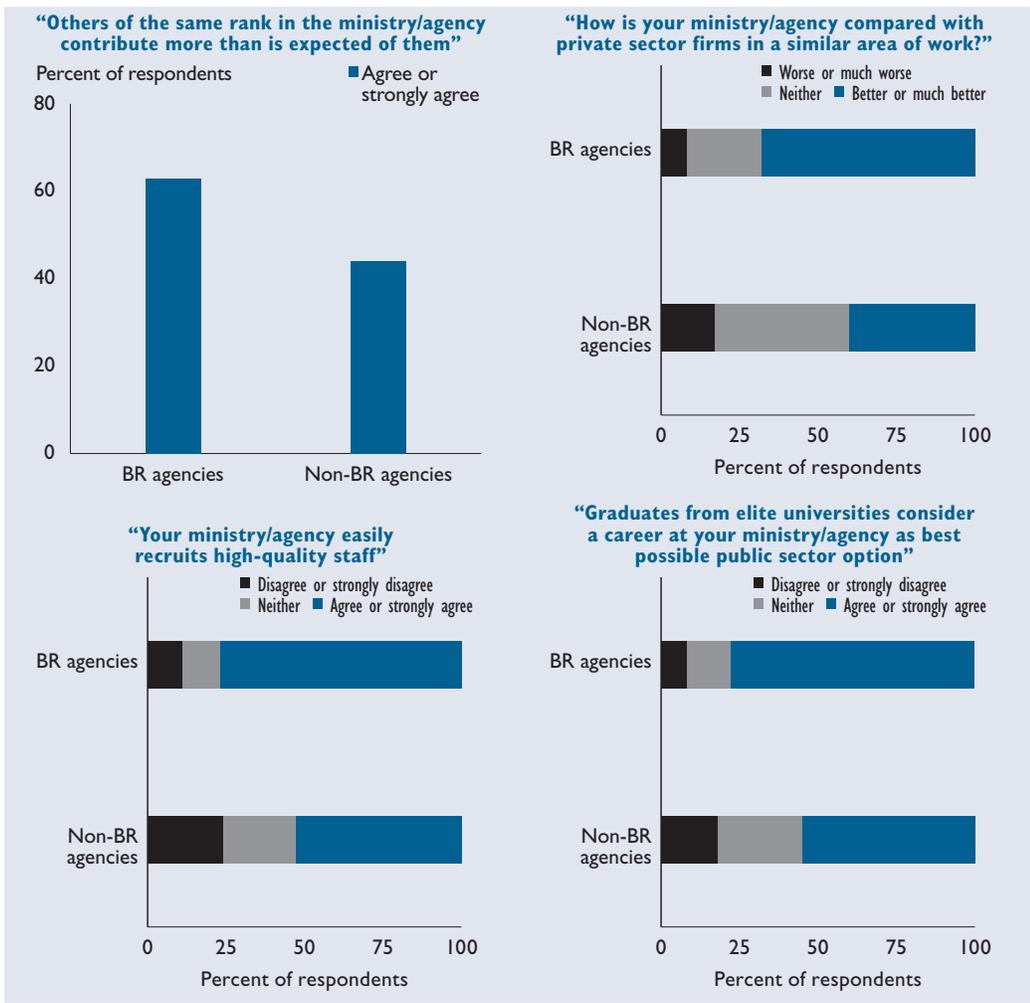
The overall evidence from the case studies is that pay differentiation can be a successful way, within the available fiscal space, to improve the quality of staff in some prioritized public entities. Differentiation causes resentment among staff of nonprivileged entities, and it is difficult to conclude whether these negative effects outweigh the positive effects in the high-pay agencies. Presumably, the initial prioritization, to the extent that it was driven by technocratic rather than political reasons, indicates that the government valued productivity improvements in these agencies as particularly important.

In Indonesia the World Bank survey asked some questions on staff perceptions of effort, engagement with the mission of the agency, and the quality of new recruits. The general view of respondents was that staff in BR agencies work harder than staff in non-BR agencies and had higher morale, and that BR agencies received higher-quality recruits (figure 18). Of the respondents from BR agencies, 63 percent agree or strongly agree that others contribute more than expected, significantly higher than all other agencies. In addition, 68 percent of staff in BR agencies said that their agency was a better or much better place to work than private sector firms in similar areas of work, compared with 40 percent in non-BR agencies. Only 11 percent of survey respondents in the BR agencies, compared with 27 percent of respondents in the non-BR agencies, either disagreed or strongly disagreed with the statement that their agency is able to recruit high-quality staff.

More anecdotal evidence from Brazil, Chile, and Russia points to similar conclusions. In Brazil the highest-paid careers, particularly subsidized careers that pay in a single package with no allowances and therefore full parity between pay and pension, were reported to have the highest application rates in merit exams. In Chile differentiation to create “high pay” and “low pay” offices was a conscious decision of successive governments in order to recruit and retain the best-quality staff for priority functions such as revenue collection, finance and treasury, and audit. The Internal Revenue Service, which has its own, higher pay scale, is a notable example. Interviewees noted that this public sector agency can better compete in the market for the best talent coming out of local universities. The critical functions allowance and the high degree of variation in pay based on individual contracts for the *contrata* staff have enabled the government to attract and retain highly skilled personnel in key posts across all institutions. In Russia the evidence also suggests a positive, if much more limited, effect on staff recruitment due to inefficiencies in competitive recruitment procedures and human resource management that watered down the pay incentive effect. The effect seems to be stronger on retention; staff turnover in the Ministry of Finance, in which the salary level and compression ratio are among the highest in the federal government, is very low.

When weighing the positive effects of differentiation on recruitment and retention against possible negative effects due to pay inequities, it is important to recognize that complexity

Figure 18 Perceptions of Effort, Morale, and Quality of Recruits between BR and Non-BR Agencies in Indonesia



Source: World Bank survey.

and lack of transparency in pay is the norm in the countries reviewed in this study. The plethora of allowances implies that salaries for similar jobs vary for a variety of individual-specific, agency-specific, and geographic reasons. Therefore, a government’s decision to prioritize certain functions and pay them more to attract talent adds another layer to an already highly inequitable system. This added inequity may be only marginally more demotivating to staff in the less privileged agencies, who have after all been living with this perceived unfairness for much of their working lives.

This point came out in the Indonesia survey. Greater individual and agency productivity in the BR agencies has indeed come at the expense of increased pay inequities between

agencies. The survey revealed very strong perceptions of unfairness within the agencies that do not receive the BR allowance, with less than 25 percent of respondents agreeing or strongly agreeing that pay in their agencies is fair compared with others doing the same job in other agencies (figure 19). Interestingly, a similar perception of unfairness exists in the non-Ministry of Finance BR agencies even though they receive higher pay than the non-BR agencies, reflecting their lower pay than the Ministry of Finance.

The survey respondents were also asked whether their pay was fair compared with others doing the same job in their agency. Interestingly, pay is considered to be less fair within the non-BR agencies than within the

The impact of differentiation on productivity is highly contextual—too much differentiation can create disruptive competition between agencies

BR agencies, despite the significant variation in BR allowances for staff in similar ranks within a BR agency (see figure 19). Surprisingly, higher-ranked staff in the BR agencies were more likely to say that their pay was fair compared with others doing similar jobs in their agency, despite the fact that pay inequities increase with seniority. These findings suggest that the existing pay system was already viewed as unfair prior to the introduction of BR pay, and that therefore the additional inequity introduced by BR may not be more demotivating in the lower-paid agencies.

Clearly, the impact of differentiation on productivity is highly contextual. Too much differentiation can create disruptive competition between agencies, as recent experience in Brazil suggests. Spurred on by perceived inequities, each career group bargains fiercely for increased allowance provisions to raise its compensation—a highly contentious process, as indicated by the numerous strikes, both at the national and state level. These resentments also find their way into the labor courts, where public employees sue, often successfully, to achieve remunerative equity with others in the public service based on the principle of “equal pay for equal work.”

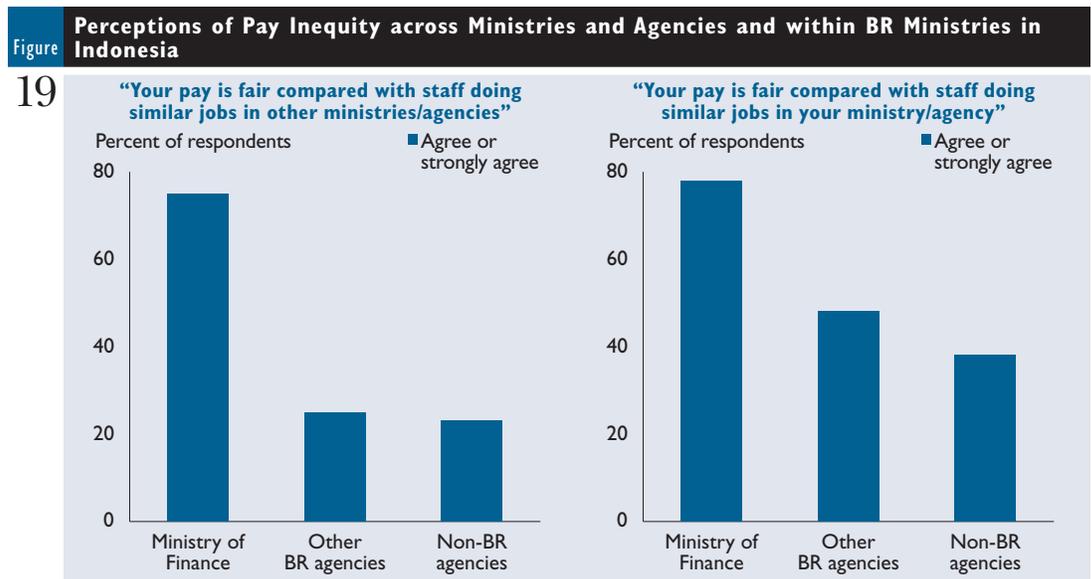
Taken to an extreme, the incentive effects of differentiation can also create a sorting problem. The possibility that differentiation can

enable particularly high public sector wages for certain privileged groups of civil servants may attract a less desirable type of worker. A comparison of public and private sector salaries in Brazil reveals that the public sector has become a wage leader for certain careers, raising concern that high public sector wages might attract recruits to the public service who are more motivated by money than by the desire for public service. Considerable media coverage in Brazil has recently focused on certain public employees being paid at levels far beyond what is justifiable.

Indirect pay flexibility levers

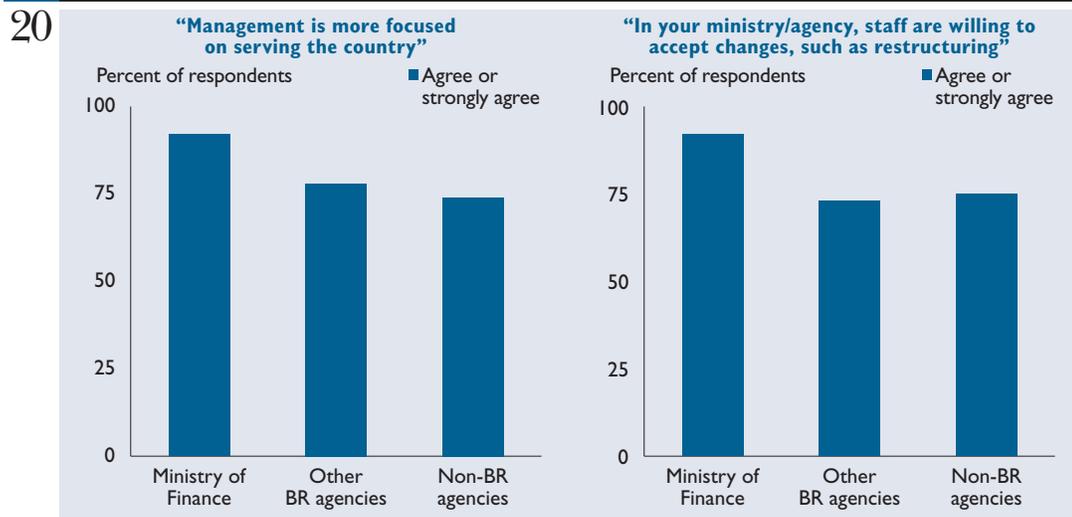
The hypothesis that pay differentiation induces managers to improve the performance dialogue with staff found some support in Indonesia but could not be examined effectively in the other cases. In Indonesia the BR process is meant to go beyond compensation reforms to emphasize managing for results. Experts who were interviewed noted the linkages between BR pay increases and organizational changes in the Ministry of Finance and the external audit agency (BPK). These pay increases gained the buy-in of staff on restructuring and also put pressure on the concerned minister to show results. In the words of one expert, BR pay helped “oil the wheels.”

These expert views are confirmed by responses from the staff survey (figure 20).



Source: World Bank survey.

Figure 20 Staff Perceptions of Management and Willingness to Accept Restructuring in Indonesia



Source: World Bank survey.

Staff in the BR agencies were more likely to agree with the statement that management has become more focused on serving the country's interests. Senior staff who were most exposed to management were also most consistent in their responses about management commitment. Successful reform also requires employees to embrace change, and

respondents from BR agencies were more likely to agree with the statement that employees in their agencies were willing to accept changes such as restructuring. On both of these questions, however, the differences were mostly from the Ministry of Finance, which has been undertaking BR for the longest period.

Summary of Empirical Analysis

In theory, PRP can reduce wage bill pressures because it provides the employer the tactical option of proposing that pay increases be provided only as enhanced performance bonuses

Table 7 summarizes the evidence from the empirical analysis. Following the approach of the literature review, the overall evidence is recorded as “positive” if there are clear perceptions of performance improvements among the staff interviewed; “negative” if the evidence indicates no effect or negative effect of pay flexibility; and “neutral” if there were opposite countervailing effects. The nature of the job has a significant bearing on the impact that PRP has on performance through direct incentive effects but not in how other aspects of pay flexibility work.

There is potential for the effective use of PRP in craft jobs through both individual- and group-based bonus schemes linked to outputs and outcomes. However, for coping jobs the evidence suggests that only group-based bonuses are effective even if they require a forced distribution of unit-level performance ratings with their accompanying rivalries. In either case, financial incentives must be significant (at least a month’s salary is a useful rule of thumb), and the measures should try to capture longer-term and sustained performance improvements to counter gaming.

Fiscal impact

The case studies provided conflicting evidence on the affordability of pay flexibility. In theory, PRP can reduce wage bill pressures because it provides the employer the tactical option of proposing that pay increases be provided only as enhanced performance bonuses—which are less costly since they are normally not pensionable (Marsden and French 1998). This was clearly a motivation for PRP in a number of

countries. Similarly, differentiation could in theory disrupt public sector-wide wage negotiations, thereby also limiting pressures for general wage increases.

The case studies showed that in Thailand, PRP added 3 percent to the wage bill with little indication of any commensurate improvement in performance, implying that it could have been a politically easier way to justify a wage increase. A similar motivation was in play in Brazil at the federal level. By contrast, in Chile, PRP was initially a response to pressure to increase public sector pay, which the government tried to manage by linking it to performance. The envelope for public sector pay has not decreased since PRP started, but pressure for across-the-board pay increases other than those linked to inflation has receded since its introduction. Similarly, differentiation may have limited the power of unions in Chile, but it fueled competition for pay increases between the unions representing the various careers in Brazil.

The degree of delegated authority

The case studies provided insufficient evidence to assess the role of organizational autonomy, or delegation, on pay flexibility. The only examples of autonomy are from Brazil and Chile. In Chile this delegation was in the form of managerial authority to set the contractual terms of most civil servants who are designated *contrata* staff and in the discretion to allocate the critical functions allowance. The head of an agency can decide how much to pay and in which category to hire *contrata* employees by arbitrarily defining their grade on the pay scale and thereafter modifying this

Table 7 Summary of Evidence about the Impact of Pay Flexibility on Civil Service Performance

PRP		
Hypothesis 1. Direct pay flexibility levers		
<i>Hypothesis 1a. PRP can directly affect individual effort</i>		
	<i>Craft job</i>	<i>Coping job</i>
<i>Literature review</i>	Reasonable evidence concerning teaching, health care, and other craft jobs (within and outside OECD settings) even though gaming is a persistent phenomenon	No relevant evidence in relation to coping jobs
<i>Case studies:</i>		
Chile	Positive for revenue authority	Negative since all staff got the bonus and therefore there was no incentive effect
Brazil (federal government)	Positive for revenue authority	Negative since all staff got the bonus
Brazil (Minas Gerais)	Positive for revenue authority and police	Weakly positive on effort: Weeded out the outliers
Malaysia	Weakly positive for revenue authority	Negative since forced distribution hurt teamwork and made for contentious management-staff relations
Thailand	Not examined in the case study	Negative since all staff got the bonus; the bonus was also very small
Philippines	Weakly positive for revenue authority	Negative: Staff are broadly supportive of the scheme but there was no reported impact on effort
Indonesia	Positive for revenue authority	Weakly positive on effort: Reduced absenteeism
Korea, Rep.	Not examined in the case study	Positive: Broad support in the central government for the individual bonus scheme. Most agencies chose individual ranking and none chose a group-based equal bonus
<i>Hypothesis 1b. PRP can have a direct effect in improving the recruitment and retention of better-quality staff</i>		
<i>Literature review</i>	Sorting effect noted in OECD settings	
<i>Case studies:</i>		
Brazil (Minas Gerais)	Some sorting effect	
Hypothesis 2. Indirect pay flexibility levers (PRP can act indirectly by providing incentives for improved management)		
<i>Case studies:</i>		
Chile	Weakly positive: Gradually improved results-based management; encouraged more delegation of human resource management authority	
Brazil (Federal)	Weakly positive	
Brazil (Minas Gerais)	Positive: Complemented performance budgeting reforms; encouraged delegation	
Malaysia	Neutral: Complemented results-based management but prompted a backlash from the unions	
Thailand	No evidence	
Philippines	Positive: Improved goal setting and monitoring; complemented and strengthened results-based management	
Indonesia	No evidence	
Korea, Rep.	Positive: Has complemented results-based management and performance budgeting	
Differentiation		
Hypothesis 3. Direct pay flexibility levers (Differentiation can have a direct effect in improving the recruitment and retention of better-quality staff)		
<i>Case studies:</i>		
Indonesia	Positive: Improved recruitment and effort in high-pay agencies; unclear if pay inequity reduced effort in low-pay agencies	
Chile	Positive: Improved recruitment and effort in high-pay agencies; unclear if pay inequity reduced effort in low-pay agencies	
Brazil	Neutral: While recruitment and retention improved in high-pay agencies, it also set up harmful competition between agencies	
Russian Federation	Weakly positive: Improved retention and led to relatively low vacancy rates in high-pay ministries	
Hypothesis 4. Indirect pay flexibility levers (Differentiation can act indirectly by providing incentives for improved management)		
<i>Case studies:</i>		
Indonesia	Positive: Raised stakes for management and opened up space for reforms in Ministry of Finance	
Chile	No evidence	
Brazil	No evidence	
Russian Federation	No evidence	

grade as needed. The only limitations are that the grade of the *contrata* employee cannot be higher than the highest permanent staff grade defined by law for each organization and that personnel expenditures remain within Budget Office limits. The critical functions allowance,

which can be up to 100 percent of gross pay, is given to contracted or permanent staff for critical functions within an agency at the discretion of the minister (again, within a budget envelope). The allowance can be discontinued at any time by the agency head or minister.

PRP was also a vehicle for delegation in Chile. The introduction of PRP at the individual and unit level was always implemented through delegation to agency heads or ministers. Its management has also gradually moved from a centralized approach to partial delegation to the ministry in charge of the policy area, but has not yet been delegated to the agency level.

There was fair degree of administrative delegation in Minas Gerais through results

agreements that granted autonomy to those entities agreeing to contractual accountability in developing remuneration criteria linked to performance targets. In fact, the results-based management reforms and PRP triggered delegation of human resource management authority, since agencies were able to convince the central finance and personnel authorities that autonomy was necessary for them to be able to deliver on their contracted outputs.

Conclusions and Policy Recommendations

How did the hypotheses hold up?

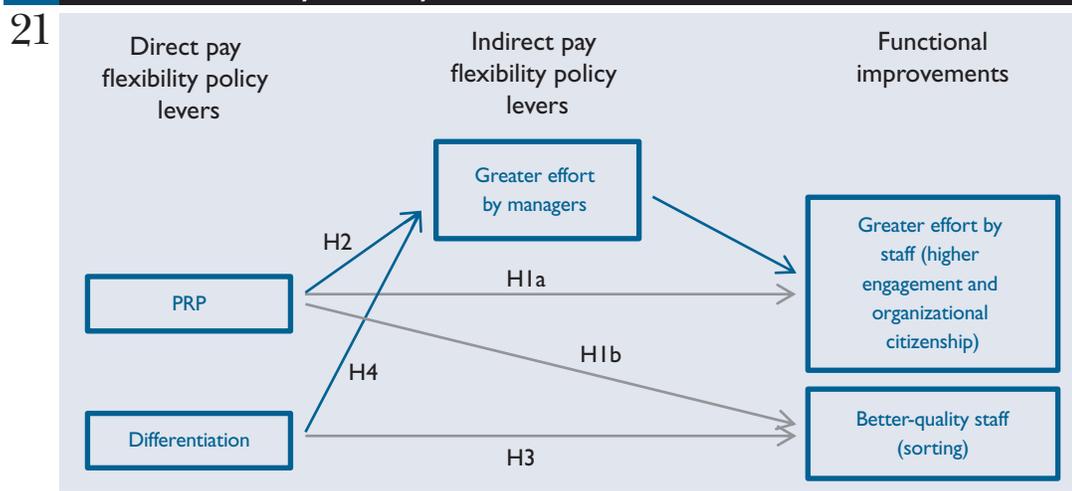
The review of the empirical evidence exhibits the diversity of experiences with pay flexibility, which means that the answer to the central question of this study, “Can pay flexibility improve the performance of public bureaucracies?” is yes, under some conditions and in certain contexts. The earlier figure is repeated below to show how PRP and differentiation are hypothesized to impact directly and indirectly on individual productivity/organizational citizenship and better staff through sorting (figure 21).

With Hypothesis 1a (PRP can directly affect individual effort), there are reasonable grounds for concluding that in low-income countries PRP can have a direct incentive effect on improving effort for teaching, health care, revenue collection, and other craft jobs that have more measurable outputs. The evidence

from the literature review and from Brazil, Chile, Indonesia, and Malaysia is reasonably conclusive, though the effect depends on the nature of the PRP scheme. The main enablers are a combination of an individual- and group-based option and a significant financial incentive—at least a month’s salary for the concerned staff as a rule of thumb. The findings suggest that the risks of too large a financial incentive and problems of free-riding in group-based bonus schemes may be overstated.

Some perverse incentives and unforeseen consequences are inevitable in PRP schemes. Since people respond to incentives, at a minimum some effort distribution will take place toward tasks that are measured and away from tasks that are not measured. Overt manipulation of targets, and other gaming behavior, is also likely. The extent to which these problems

Figure 21 The Link between Pay Flexibility and Performance



The extent to which manipulation and gaming can be mitigated depends on the validation and review processes in place by managers and central oversight agencies

can be mitigated depends on the validation and review processes in place by managers and central oversight agencies, and on learning and improvements in performance indicators so as to encourage sustained, long-term performance gains. In Minas Gerais the evolution of the scheme in the revenue authority and the innovative design of the general PRP scheme, in which the size of the incentive increases with the number of good performance evaluations, are instructive. In Korea senior civil servants are required to identify several key performance targets rather than one or two performance goals, and the general performance appraisal system for middle- and lower-level officials reviews performance goals in three aspects—task completeness, achievement timeliness, and job difficulties—in order to minimize gaming or evaluation errors. In Chile the performance targets have evolved over time to be more robust and to be more result related as opposed to process related.

For core policy and administration jobs, the direct effects of PRP are weak and limited to identifying egregiously underperforming staff. The lack of objective and defensible output measures discourages frontline managers from distinguishing between staff using subjective assessments. As a result, PRP becomes a salary supplement with no incentive effect. The question then becomes a different one of whether this form of salary supplement is better than others.

Implementing PRP in coping jobs through a forced differentiation of individual performance ratings is risky and can harm staff morale. Korea has done reasonably well under the system, probably because of relatively high trust levels and the general high functionality of the Korean bureaucracy. The experience of Malaysia highlights the risks of this approach and the considerable turmoil in a bureaucracy that PRP in coping jobs can engender, and it may be more representative of developing country contexts. Similarly, evidence from the Philippines points to the overall negative impact on staff motivation of forced distribution because of the lack of credibility of supervisor performance assessments and concerns of favoritism.

With Hypothesis 1b (that PRP can have a direct effect in improving the recruitment and retention of better-quality staff), the literature review suggested some modest effect in OECD settings and some support from the Brazil (Minas Gerais) case study.

Hypothesis 2 (that PRP can act indirectly by providing incentives for greater effort by managers) was supported by evidence from Korea, Minas Gerais, and the Philippines. The cases show that PRP encourages frontline managers to pay attention to the mission and goals of their organization and to bring staff to work together to achieve these goals, and it can complement other reforms like performance-based budgeting and results-based management. Managers can of course use other vehicles, including just becoming more effective managers, but the evidence suggests that monetary incentives encourage managers and staff to have more communication around the performance agenda and that these conversations help renegotiate the “effort bargain” (see box 2).

These management changes are highly conditional on context, as the contrast between the experiences of Malaysia and Minas Gerais shows, and they are also gradual. In Malaysia the imposition of an individual forced ranking for PRP coupled with powerful unions and contentious staff-management relations limited the potential of these managerial improvements. In Minas Gerais management and staff developed a much more cooperative and accommodative relationship that was helped by a more nuanced PRP scheme that encouraged longer-term performance improvements. In the Philippines staff clearly perceive that PRP has improved management, but there is as yet no discernible effect of this improved management on individual productivity and organizational citizenship.

Hypothesis 3 (that differentiation can directly improve staff quality) found support. Higher average pay and greater pay decomposition, so that pay increases significantly over the course of a career, have a clear effect on staff quality in the privileged agencies through improved recruitment and retention. The tradeoff is increased inequity in pay across

the public sector. Inequity and lack of transparency is, however, the norm in many developing and emerging market countries and so the performance improvements for high-pay groups of staff may not necessarily be offset by commensurate performance declines in low-pay groups.

Hypothesis 4 (that differentiation can act indirectly by providing incentives for greater effort by managers) found some support from the case studies. The evidence from Indonesia in particular suggested that management in the Ministry of Finance was under the national spotlight to make performance improvements because of its higher pay, which was widely justified and publicized as being linked to performance improvements even though there was no such explicit linkage in design. This pay change then opened up space for human resource improvements, and this greater management focus was complemented by greater acceptance by staff of organizational changes such as restructuring.

Implications for pay policy in developing countries

What are the broader lessons for pay policy, particularly in the more challenging low-income country contexts in Africa, South Asia, and elsewhere?

While this study has focused on the *impact* of pay flexibility, set out as a series of hypotheses for testing, an equally significant question concerns *political feasibility and motivation*. The case studies have shown that improving performance may not be the primary motivation for introducing PRP. Instead, PRP may be a politically feasible, and fiscally less costly, way of increasing public sector salaries. In Brazil at the federal level, for example, the phrase “variable pay” is often used interchangeably with the term “performance pay,” pointing to the fiscal motivations behind its introduction. With the aging profile of civil servants (the average age has increased from 43 to 46 over the past decade) and increasing pension liabilities, the government has taken a harder line with the employee unions on basic salary increases and has instead increased pay by using allowances that may or may not be

pensionable, depending on the particular career. The performance element is therefore of secondary importance, and it may be a way of justifying salary increases under the guise of performance improvements. A similar motivation existed in Chile, where steady economic growth created pressures to raise real public sector wages, which are more easily sold to the public as PRP.

Even when the government motivation is performance rather than cost savings, political resistance from public sector unions has often proved fatal to PRP. This was most obviously the case in Malaysia, but has also been an important factor in Brazil and Chile. The strength of public sector unions is therefore an important variable for PRP implementations, with countries with weaker unions, such as Korea (unions were banned under the authoritarian regime until the late 1990s) having more success. Interestingly, even in the presence of strong unions governments can have some bargaining space. Staff interviewed in Minas Gerais reported a relatively harmonious relationship with the relevant unions, indicating that part of the success of public sector reforms was also linked to a different attitude toward the role of unions in government. Staff attributed cordial relations in part to the proactive role taken, in particular, by successive secretaries of finance, who customarily met personally at least twice a year with union representatives.

Taking into account both political factors and technical considerations, six main messages are evident. First, *pay flexibility can improve performance*. Given the dearth of success stories in public administration reform in developing countries, and the generally negative view of PRP in particular that has prevailed in public administration academic and policy circles, this is a powerful finding of the study.

There is a long tradition of skepticism about the introduction of seemingly OECD-like reforms in developing countries, and pay flexibility could be placed in that category. Some technical criticisms of performance pay go back to their use for teachers in British schools in the nineteenth century (Gratz 2009). Arguments are made that performance measures can induce tunnel vision, myopia, and measure

PRP may be a politically feasible, and fiscally less costly, way of increasing public sector salaries

Pay flexibility can complement and provide the behavioral impetus for budgeting and other performance management reforms

fixation (Propper and Wilson 2003). Pink (2009) has developed the critique of monetary and other extrinsic incentives into a broader theory, hypothesizing that they are both counterproductive, since they frequently undermine intrinsic incentives, and unnecessary, since intrinsic incentives can be harnessed and used to maximize individual productivity. These arguments are not without merit, of course, and there is little doubt that intrinsic motivations are of particular significance in public service (Banuri and Keefer 2013). However, the findings of this study suggest that simplistic universal arguments against pay flexibility are not justified empirically and particularly not when considering self-selection and sorting, since an explicit system of performance pay can attract extrinsically motivated applicants to the civil service.

Pay flexibility reforms are not a silver bullet, as no public sector reforms can be, and involve tradeoffs and risks. Poorly designed PRP schemes can cause more harm than good, and there is the potential for harmful task reallocation and gaming, given that even in craft jobs no performance measures can capture all the dimensions of each job. The evidence reviewed, however, suggests that these tradeoffs can be managed with the necessary sophistication if risks of perverse behavior are correctly identified. The extent of the appropriate pay flexibility measures very much depends on local context. At a minimum, the study suggests that these schemes can be implemented in craft jobs with the necessary monitoring regime in place to detect and respond to gaming behavior. Despite much academic and professional skepticism, there is therefore every reason to keep pay flexibility in the reform toolkit.

Second, *pay flexibility works most strikingly in changing managerial behaviors*. It can focus government attention on management improvements under the broad rubric of “the effort bargain”—vision, strategy, agency strategic objectives and key performance indicators, and better dialogue with staff to achieve these through teamwork and task allocation. It can also complement and provide the behavioral impetus for budgeting and other performance management reforms. The

examples from Korea and Minas Gerais show that it is the combination of reforms—PRP, results-based management, and performance-based budgeting—that is the key to improved performance.

This essential ingredient of management has several implications for the technical design of the PRP scheme and for the sequence of reforms. PRP schemes for coping jobs should encourage such management changes. A large group-based bonus PRP scheme is preferable, despite the difficulties in establishing unit-level performance targets, because it bypasses the problem of distinguishing between individuals’ performance and puts the spotlight on management improvements as the key linking PRP to better performance. The performance measures in such a scheme are likely to be process oriented, and to the extent that these policy jobs also entail providing services to government service delivery units (releases of funds, regulatory oversight, personnel oversight, and so forth), these process indicators can be supplemented with customer satisfaction measures. The risks of gaming, however, will be higher, and the scheme will therefore need to be more vigilantly managed. The incentive has to be large (at least equivalent to a month’s salary) to be sufficient to induce managers to change their practices.

Introducing pay flexibility across the board in a public sector where management is highly dysfunctional is inadvisable. An approach could be to introduce it asymmetrically where there is some basic level of managerial competence—for example, using a PRP scheme having the design features identified above (a group-based bonus only, a financial incentive equivalent to at least a month’s salary, and an incentive designed to encourage longer-term, sustained performance). The focus could be agencies where there has been at least a serious attempt to set goals and to improve human resource management practices that flexible pay policies can build on and strengthen.

In terms of sequencing, it is advisable to launch results-based management (or other variants of performance management) first before introducing PRP. In all the case countries with positive experiences, PRP was

introduced after, and energized, an earlier set of reforms to improve goal setting and monitoring and managing for results. This sequence of reforms enabled the PRP scheme to build on a set of managerial resources—performance indicators, progress review mechanisms, and so forth—that had been established when there was no “money on the table” that could have corrupted the process. Often these management reforms had limited impact precisely because there were no stakes attached to them, and PRP was therefore needed to induce managers to start taking their management responsibilities more seriously. But introducing PRP in an environment with no such history of reforms risks having the money incentive overwhelm everything else.

Third, *the path to improved public sector performance does not necessarily need to go through a stage of “whole of government” pay rationalization or pay simplification.* Many of the cases analyzed in this study—for example, Brazil, Chile, and Indonesia—were symptomatic of the general developing country phenomena of high variance in compensation for similar jobs based on a variety of employer-related and personal factors. Pay inequity and lack of transparency in compensation were ubiquitous. This study suggests that there is no reason to seek to move from these haphazard or asymmetric pay structures to homogenous single pay spine arrangements before contemplating differentiation or PRP to improve performance. Performance improvements are possible through the “purposeful complexity” of pay flexibility even when layered on top of a complex pay regime. This is an encouraging finding given the technical and political challenges of comprehensive pay rationalization and the poor track record of such reforms.

Fourth, in both “messy” and simple pay policy contexts, *flexible pay policy can work with rather than instead of long-term career incentives.* The prospect of continuing salary enhancements over time with a strong component of peer recognition remains important to staff. The individual PRP scheme in Minas Gerais is a good example of this complementarity between short- and long-term incentives, since the size of the incentive increases with

sustained good performance over the employee’s career. Similarly, in Indonesia pay differentiation for BR agencies introduced both higher average pay for all staff but particularly higher pay for senior staff.

Fifth, *the strategy and implementation of pay flexibility reforms have to take into account the extent of fragmentation and complexity of the existing public sector pay structure in the country.* While, as noted above, the evidence suggests that pay simplification is not necessary, or advisable, as a prior step before introducing pay flexibility, the extent of “messiness” of the pay regime has implications for the pay flexibility strategy. The strategy should be different in relatively neat systems (such as the Philippines) where the pay structure is fairly simple and uniform across the core administration, compared with messy systems where pay varies for a whole host of idiosyncratic reasons and where central fiscal control and management coherence is compromised.

In simpler systems there is less risk in more ambitious, across-the-board pay flexibility reforms, if there is an explicit recognition of possible perverse behavior and unintended consequences, and experimentation and learning-as-you-go are built in to the reforms. Ideally, even in these systems pay flexibility would be introduced first for craft jobs, and within coping jobs first in organizations that have already made investments in improving management. This restriction is necessary to limit the administrative burden of the necessary validation and monitoring systems, and the system can then be gradually expanded as the sophistication of this monitoring regime increases.

In complex systems there is the risk that pay flexibility degenerates into yet another element of the messy pay regime with few productivity gains and a further weakening of central fiscal control and management coherence. This risk can be mitigated by limiting pay flexibility to a select few high-priority organizational “islands,” chosen either because they are the highest priority or because they are managed relatively well. These are the staff whose productivity improvements are considered to be the most important for government

The prospect of continuing salary enhancements over time with a strong component of peer recognition remains important to staff

A government's capacity to collect and validate data on performance and to coordinate pay flexibility across the public sector is an important factor determining the success of these schemes

performance. Pay reforms are notoriously sticky and hard to reverse, and limiting the terrain is necessary to prevent the type of uncontrolled lobbying by all groups that has beset Brazil. These agencies should be required to achieve explicit standards in management practices (on recruitment, goal setting and monitoring, and staff performance evaluations) to qualify for pay flexibility.

Sixth, *many questions remain, and much more research is needed*. As noted earlier, the breadth of contextual coverage of this study and the range of evidence used have come at the expense of empirical depth, in part because analyses of the core public administration do not easily lend themselves to rigorous impact evaluations. The study cannot make causal claims, and the use of staff perceptions creates potential biases because respondents may have an incentive to preserve pay flexibility and so may be more positive in their responses. Every effort should therefore be made at the country level to track improvements in deliverables to ensure that the application of pay flexibility is tested against services or outputs that matter.

There is considerable heterogeneity in the impact of pay flexibility reforms based on contextual factors that go beyond the two—type of public sector job and design features of the pay flexibility scheme—that this study has looked at. How pay flexibility interacts with existing formal and informal rules and culture (the

level of trust has been noted as an important factor on numerous occasions in this report) are all key issues that require more investigation. The study could not effectively explore how organizational autonomy interacts with pay flexibility. A government's capacity to collect and validate data on performance and to coordinate pay flexibility across the public sector is also presumably an important factor determining the success of these schemes. The need for a robust monitoring regime raises the question of cost effectiveness and whether the added bureaucratic burden associated with performance incentives can pay for itself though higher productivity.

All these questions that must be analyzed in specific country contexts. In the past, pay flexibility for the core public administration has often been ruled out a priori in World Bank and other donor advice. This study instead calls for a different approach, arguing that these reforms be assessed based on the design features of the scheme, the jobs to which it is being applied, potential tradeoffs in terms of increased pay inequity, the unintended consequences and perverse gaming behavior that these are likely to generate and whether they can be managed, and other contextual factors. We hope that this more nuanced approach to pay reform will assist policy makers and development partners in the critical agenda of improving public sector performance.

Notes

1. Rafferty and others (2005) review the literature on organizational commitment and citizenship, finding empirical associations between commitment and increased job satisfaction (Vandenberg and Lance 1992); increased job performance (Mathieu and Zajac 1990); improved sales (Barber, Hayday, and Bevan 1999); lower employee turnover (Cohen 1991); less intention to leave (Cohen 1993; Balfour and Wechsler 1996); lower absenteeism (Cohen 1993; Barber, Hayday, and Bevan 1999). Osterloh and Frost 2002 note that citizenship entails employee behaviors that maintain the organization's social system and are important to its smooth running (Houston 2009).
2. The literature uses a variety of terms for such financial incentives: performance pay, performance-based pay, performance-based incentives, and pay-for-performance.
3. The expert interviews and perception surveys ask questions on recruitment, retention, career development, performance management, pay levels and dispersion, and performance incentives. The survey in Indonesia was conducted in 14 central government ministries and agencies and covered 4,000 staff. The survey in the Philippines covered 7 central government departments and agencies and approximately 2,500 staff.
4. Wilson had originally used this framework to classify organizations and not jobs, the implicit assumption being that organizations were homogenous in the tasks that they performed.
5. The degree of delegation varies widely across the OECD countries, with delegation largely, but not entirely, correlated with the broad introduction of “new public management” ideas (OECD 2011, 127).
6. Inevitably, the classification of studies is somewhat subjective. Studies were rated as *positive* if they showed general evidence on the basic functionality of incentive schemes, even if additional results qualify the effect—for example, studies on crowding-out of intrinsic motivation generally still find positive effects of explicit incentives.
7. Interpreted as purely theoretical papers or studies with a weak research design (for example, selection on the dependent variable only, no meaningful variation, and no explicit consideration of counterfactuals).
8. Studies that mostly describe reforms implemented in a small number of cases without comparing them with cases without performance pay.
9. Studies based on a small number of cases but having at least an implicit consideration of a counterfactual and with some minimal data analysis.
10. Studies with an explicit counterfactual analysis, using a representative sample of cases with and without treatment and often using statistical techniques to limit threats to causal inference.
11. We think a minimal level of internal and external validity is necessary to draw reliable conclusions from the evidence presented in a study, especially when policy recommendations are concerned. For that

reason we opted to classify studies as “high quality” only if their analysis was based on quasi-experimental methods and the analyzed sample was somewhat representative of the theoretical population under study.

12. Deconcentrated units of federal executive authorities in the regions are referred to as *territorial authorities*.
13. Comparing only agencies with fingerprint machines or only agencies without them, respondents in agencies with BR status are much less likely to report that at least 20 percent of their coworkers leave work early.

References

- Ahmad, S., and R. G. Schroeder. 2003. "The Impact of Human Resource Management Practices on Operational Performance: Recognizing Country and Industry Differences." *Journal of Operations Management* 21: 19–43.
- Ahmed, Z., F. Shields, R. White, and J. Wilbert. 2010. "Managerial Communication: The Link between Frontline Leadership and Organizational Performance." *Journal of Organizational Culture, Communications & Conflict* 14 (1): 107–20.
- Andrade, N. H. S. 2011. A introdução dos especialistas em políticas públicas e gestão governamental na administração pública Brasileira: Análise comparativa da carreira nas esferas estadual e federal. IV Congresso CONSAD de Gestão Pública, Brasília, DF.
- Armstrong, M. 2010. *Armstrong's Essential Human Resource Management Practice*. London, Kogan Page.
- Asch, B. J. 1990. *Navy Recruiter Productivity and the Freeman Plan*. Santa Monica, CA: RAND Corporation.
- Atkinson, A., S. Burgess, B. Croxson, P. Gregg, C. Propper, H. Slater, and D. Wilson. 2004. *Evaluating the Impact of Performance-Related Pay for Teachers in England*. Working Paper No.04/113. Bristol, UK: Centre for Market and Public Organisation.
- Balfour, D., and B. Wechsler. 1996. "Organisational Commitment: Antecedents and Outcomes in Public Organisations." *Public Productivity and Management Review* 29: 256–77.
- Bandiera, O., I. Barankay, and I. Rasul. 2006. *Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment*. Discussion Paper No. 2062. Bonn, Germany: Institute for the Study of Labor.
- Banuri, S., and P. Keefer. 2013. *Intrinsic Motivation, Effort and the Call to Public Service*. Washington, DC: World Bank.
- Barber, L., S. Hayday, and S. Bevan. 1999. *From People to Profits (IES Report 355)*. London: Institute for Employment Studies.
- Basinga, P., P. J. Gertler, A. Binagwaho, A. L. B. Soucat, J. R. Sturdy, and C. M. J. Vermeersch. 2010. *Paying Primary Health Care Centers for Performance in Rwanda*. Washington, DC: World Bank.
- Behrend, H. 1957. "The Effort Bargain." *Industrial and Labor Relations Review* 10 (4): 503–15.
- Bender, K. A., and R. F. Elliott. 2003. *Decentralised Pay Setting: A Study of the Outcomes of Collective Bargaining Reform in the Civil Service in Australia, Sweden and the U.K.* Ashgate, UK: Farnham.
- Bertelli, A. M. 2006. "Motivation Crowding and the Federal Civil Servant: Evidence from the U.S. Internal Revenue Service." *International Public Management Journal* 9 (23): 3–23.
- Bruns, B., D. Filmer, and H. A. Patrinos. 2011. *Making Schools Work: New Evidence on Accountability Reforms*. Washington, DC: World Bank.
- Burgess, S., and P. Metcalfe. 1999. *Incentives in Organisations: A Selective Overview of the Literature with Application to the Public Sector*. Bristol, UK: University of Bristol, CMPO and CEPR.
- Burgess, S., C. Propper, M. Ratto, S. Scholder, and E. Tominey. 2010. "Smarter Task Assignment or Greater Effort: What Makes the Difference on Team Performance?" *Economic Journal* 120 (547): 968–89.

- Cadsby, C. B., F. Song, and F. Tapon. 2007. "Sorting and Incentive Effects of Pay-for-Performance: An Experimental Investigation." *Academy of Management Journal* 50 (2): 387–405.
- Christensen, D. B., N. Neil, W. E. Fasset, D. H. Smith, and G. A. S. Holmes. 2000. "Frequency and Characteristics of Cognitive Services Provided in Response to a Financial Incentive." *Journal of the American Pharmaceutical Association* 40 (5): 609–17.
- Clotfelter, C., R. A. Diaz, H. Ladd, and J. Vigdor. 2004. "Do School Accountability Systems Make It More Difficult for Low-Performing Schools to Attract and Retain High-Quality Teachers?" *Journal of Policy Analysis and Management* 23 (2): 251–71.
- Clotfelter, C., E. Glennie, H. Ladd, and J. Vigdor. 2008. "Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina." *Journal of Public Economics* 92: 1352–70.
- Coase, R. H. 1937. "The Nature of the Firm." *Economica* 4 (16): 386–405.
- Cohen, A. 1991. "Career Stage as a Moderator of the Relationship between Organisational Commitment and Its Outcomes: A Meta-Analysis." *Journal of Occupational Psychology* 64: 253–68.
- . 1993. "Age and Tenure in Relation to Organisational Commitment: A Meta-Analysis." *Basic and Applied Social Psychology* 14: 143–59.
- Courty, P., and G. Marschke. 2004. "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives." *Journal of Labor Economics* 22 (1): 23–56.
- de Assis, L. O. M. 2012. *Efeitos organizacionais da implantação de metas e remuneração variável por desempenho: O caso da segurança pública em Minas Gerais*. São Paulo: Fundação Getúlio Vargas.
- Dewatripont, M., I. Jewitt, and J. Tirole. 1999a. "The Economics of Career Concerns, Part 1: Comparing Information Structures." *Review of Economic Studies* 66: 183–98.
- . 1999b. "The Economics of Career Concerns, Part 2: Application to Missions and Accountability of Government Agencies." *Review of Economic Studies* 66: 199–217.
- Dickens, W. T., and L. F. Katz. 1987. *Interindustry Wage Differences and Industry Characteristics*. NBER Working Paper 2014. Washington, DC: National Bureau of Economic Research.
- Dowling, B., and R. Richardson. 1997. "Evaluating Performance-Related Pay for Managers in the National Health Service." *International Journal of Human Resource Management* 8 (3): 348–66.
- Duflo, E., R. Hanna, and S. P. Ryany. 2010. *Incentives Work: Getting Teachers to Come to School*. Cambridge, MA: Department of Economics and J-PAL, MIT, and Kennedy School of Government.
- Eldridge, C., and N. Palmer. 2009. "Performance-Based Payment: Some Reflections on the Discourse, Evidence and Unanswered Questions." *Health Policy and Planning* 1 (7): 160–66.
- Fairbrother, G., K. L. Hanson, S. Friedman, and G. C. Butts. 1999. "The Impact of Physician Bonuses, Enhanced Fees, and Feedback on Childhood Immunization Coverage Rates." *American Journal of Public Health* 89 (2): 171–75.
- Fairbrother, G., M. J. Siegel, S. Friedman, P. D. Kory, and G. C. Butts. 2001. "Impact of Financial Incentives on Immunization Rates in the Inner City: Results of a Randomized Controlled Trial." *Ambulatory Pediatrics* 1 (4): 206–12.
- Figlio, D. N., and L. W. Kenny. 2007. "Individual Teacher Incentives and Student Performance." *Journal of Public Economics* 91: 901–14.
- Figlio, D. N., and J. Winicki. 2005. "Food for Thought: The Effects of School Accountability Plans on School Nutrition." *Journal of Public Economics* 89: 381–94.
- Frey, B. S., and M. Osterloh. 1999. *Pay for Performance: Immer Empfehlenswert?* Münster, Germany: Zeitschrift für Führung und Organisation.
- Fryer, R. G. 2011. *Teacher Incentives and Student Achievement: Evidence from New York City Public Schools*. NBER Working Paper 16850. Washington, DC: National Bureau for Economic Research.
- Glewwe, P., N. Ilias, and M. Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (3): 205–27.

- Goodman, S., and L. Turner. 2010. "Teacher Incentive Pay and Educational Outcomes: Evidence from the NYC Bonus Program." Working paper presented at the PEPG Conference, "Merit Pay: Will It Work? Is It Politically Viable?" Cambridge, MA, Harvard Kennedy School, June 3–4.
- Grady, K., J. Lemkau, L. N., and C. Caddell. 1997. "Enhancing Mammography Referral in Primary Care." *Preventive Medicine* (26): 791–800.
- Gratz, D. B. 2009. *The Peril and Promise of Performance Pay: Making Education Compensation Work*. Lanham, MD: Rowman & Littlefield.
- Groshen, E. L. 1991. "Sources of Intra-Industry Wage Dispersion: How Much Do Employers Matter?" *Quarterly Journal of Economics* 106 (3): 869–84.
- Han, S. 2010. "Government Officials' Responses to the Performance-Related Pay System: The Application of the Grounded Theory." *Korean Public Administration Review* 44 (4): 29–58.
- Hasnain, Z., N. Manning, and J. H. Pierskalla. 2012. *Performance-Related Pay in the Public Sector: A Review of Theory and Evidence*. Policy Research Working Paper 6043. Washington, DC: World Bank.
- Heckman, J., C. Heinrich, and J. Smith. 1997. "Assessing the Performance of Performance Standards in Public Bureaucracies." *American Economic Review* 87 (2): 389–95.
- Hillman, A., K. Ripley, N. Goldfarb, I. Nuamah, J. Weiner, and E. Lusk. 1998. "Physician Financial Incentives and Feedback: Failure to Increase Cancer Screening in Medicaid Managed Care." *American Journal of Public Health* 88 (11): 1699–701.
- . 1999. "The Use of Physician Financial Incentives and Feedback to Improve Pediatric Preventive Care in Medicaid Managed Care." *Pediatrics* (104): 931–35.
- Hochberg, Y. V., and L. Lindsey. 2010. "Incentives, Targeting and Firm Performance: An Analysis of Non-Executive Stock Options." *Review of Financial Studies* 23 (11): 4148–86.
- Holmstrom, B. 1982. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies* 1: 169–82.
- Holmstrom, B., and P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics & Organization* 7: 24–52.
- Houston, D. J. 2009. "Motivating Knights or Knaves? Moving Beyond Performance-Related Pay for the Public Sector." *Public Administration Review* 69 (1): 43–56.
- Hutchinson, S., and J. Purcell. 2003. *Bringing Policies to Life: The Vital Role of Front Line Managers in People Management*. London: Chartered Institute of Personnel and Development.
- . 2007. *Learning and the Line: The Role of Line Managers in Training, Learning and Development*. London: Chartered Institute of Personnel and Development.
- Independent Evaluation Group. 2008. *Public Sector Reform: What Works and Why?* Washington, DC: World Bank.
- Jacob, B. A. 2005. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89: 761–96.
- Jacob, B. A., and S. D. Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 118 (3): 843–77.
- Kahn, C. M., E. C. De Silva, and J. P. Ziliak. 2001. "Performance-Based Wages in Tax Collection: The Brazilian Tax Collection Reform and Its Effects." *Economic Journal* 111: 188–205.
- Kellough, E. J., and H. Lu. 1993. "The Paradox of Merit Pay in the Public Sector: Persistence of a Problematic Procedure." *Review of Public Personnel Administration* 13: 45–64.
- Ketelaar, A., N. Manning, and E. Turkisch. 2007. *Performance-Based Arrangements for Senior Civil Servants: OECD and Other Country Experiences*. OECD Working Papers on Public Governance 2007/5. Paris: OECD.
- Knies, E., and P. Leisink. 2013. "Leadership Behavior in Public Organizations: A Study of Supervisory Support by Police and Medical Center Middle Managers." *Review of Public Personnel Administration* (online), November.
- Koch, R. 1996. "Towards and Entrepreneurial Paradigm in Public Management: International Trends." In *Entrepreneurial Management in the Public Sector*, edited by J. Wanna, J.

- Forster, and P. Graham, 33–40. Melbourne: Macmillan Education.
- Kouides, R., N. Bennett, B. Lewis, J. Cappuccio, W. Barker, and M. LaForce. 1998. “Performance-Based Physician Reimbursement and Influenza Rates in the Elderly.” *American Journal of Preventive Medicine* 14 (2): 89–95.
- Kremer, M., and D. Chen. 2001. “An Interim Report on a Teacher Attendance Incentive Program in Kenya.” Cambridge, MA: Harvard University.
- Krueger, A. B., and L. H. Summers. 1988. “Efficiency Wages and the Inter-Industry Wage Structure.” *Econometrica* 56 (2): 259–93.
- Kular, S., M. Gatenby, C. Rees, E. Soane, and K. Truss. 2008. *Employee Engagement: A Literature Review*. Kingston upon Thames, UK: Kingston Business School. <http://eprints.kingston.ac.uk/4192/1/19wempen.pdf>.
- Lavy, V. 2008. *Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-Based Pay Tournaments among Teachers*. NBER Working Paper No. 14338. Washington, DC: National Bureau for Economic Research.
- . 2009. “Performance Pay and Teachers’ Effort, Productivity and Grading Ethics.” *American Economic Review* 99 (5): 1979–2011.
- Lazear, E. P. 1981. “Agency, Earnings Profiles, Productivity and Hours Restrictions.” *American Economic Review* 71 (5): 606–20.
- . 1989. “Pay Equality and Industrial Politics.” *Journal of Political Economy* 97: 561–80.
- Lee, H. 2010. “The Effect of Public Officials’ Pay for Performance Satisfaction on Salary Satisfaction, Job Satisfaction and Organizational Commitment.” *Korea Governance Review* 17 (2): 139–66.
- Lindauer, D. L., and B. Nunberg, eds. 1994. *Rehabilitating Government*. Washington, DC: World Bank.
- Marsden, D. 2004. “The Role of Performance-Related Pay in Renegotiating the “Effort Bargain”: The Case of the British Public Service.” *Industrial and Labor Relations Review* 57 (3): 350–70.
- . 2009. *The Paradox of Performance-Related Pay Systems: Why Do We Keep Adopting Them in the Face of Evidence That They Fail to Motivate?* London: London School of Economics, Centre for Economic Performance.
- Marsden, D., and S. French. 1998. *What a Performance: Performance-Related Pay in the Public Services*. London: London School of Economics, Centre for Economic Performance.
- Mathieu, J., and D. Zajac. 1990. “A Review and Meta-Analysis of the Antecedents, Correlates, and Consequences of Organisational Commitment.” *Psychological Bulletin of the American Psychological Association* 108: 171–94.
- Muralidharan, K., and V. Sundararaman. 2009. *Teacher Performance Pay: Experimental Evidence from India*. NBER Working Paper 15323. Washington, DC: National Bureau for Economic Research.
- Norrgren, F., and J. Schaller. 1999. “Leadership Style: Its Impact on Cross-Functional Product Development.” *Journal of Product Innovation Management* 16 (4): 377–84.
- Nunberg, B., and R. Taliario. 2012. “Sabotaging Civil Service Reform in Aid-Dependent Countries: Are Donors to Blame?” *World Development* 40 (10): 1970–81.
- Odden, A., and C. Kelley. 2002. *Paying Teachers for What They Know and Do*. Thousand Oaks, CA: Corwin Press.
- OECD (Organisation for Economic Co-operation and Development). 2005. *Performance-Related Pay Policies for Government Employees*. Paris: OECD.
- . 2011. *Government at a Glance*. Paris: OECD.
- Osterloh, M., and J. Frost. 2002. “Motivation and Knowledge as Strategic Resources.” In *Successful Management by Motivation: Balancing Intrinsic and Extrinsic Incentives*, edited by B. S. Frey and M. Osterloh, 27–51. New York: Springer-Verlag.
- Perry, J. L., T. A. Engbers, and S. Y. Jun. 2009. “Back to the Future? Performance-Related Pay, Empirical Research and the Perils of Persistence.” *Public Administration Review* 69 (1): 39–51.
- Pfeffer, J. 1998a. *The Human Equation: Building Profits by Putting People First*. Boston: Harvard Business School Press.
- . 1998b. “Seven Practices of Successful Organizations.” *California Management Review* 40 (2): 96–124.

- Pink, D. H. 2009. *Drive: The Surprising Truth about What Motivates Us*. New York: Riverhead.
- Propper, C., and D. Wilson. 2003. *The Use and Usefulness of Performance Measures in the Public Sector*. CMPO Working Paper Series No. 03/073. Bristol, UK: Centre for Market and Public Organisation.
- Purcell, J., and S. Hutchinson. 2007. "Front-Line Managers as Agents in the HRM-Performance Causal Chain: Theory, Analysis and Evidence." *Human Resource Management Journal* 17 (1): 3–20.
- Purcell, J., K. Kinnie, S. Hutchinson, B. Rayton, and J. Swart. 2003a. *People and Performance: How People Management Impacts on Organizational Performance*. London: Chartered Institute of Personnel and Development.
- . 2003b. *Understanding the People and Performance Link: Unlocking the Black Box*. London: Chartered Institute of Personnel and Development.
- Rafferty, A. M., J. Maben, E. West, and D. Robinson. 2005. *What Makes a Good Employer?* Geneva: International Council of Nurses. www.icn.ch/images/stories/documents/publications/GNRI/Issue3_Employer.pdf.
- Rexed, K., C. Moll, N. Manning, and J. Allain. 2007. *Governance of Decentralised Pay Setting in Selected OECD Countries*. OECD Working Papers on Public Governance, 2007/3. Paris: OECD. <http://caliban.sourceoecd.org/vl=7179447/cl=20/nw=1/rpsv/cgi-bin/wppdf?file=514qdfvl56d.pdf>.
- Roski, J., R. Jeddeloh, L. An, H. Lando, P. Hannan, and C. Hall. 2003. "The Impact of Financial Incentives and a Patient Registry on Preventive Care Quality: Increasing Provider Adherence to Evidence-Based Smoking Cessation Practice Guidelines." *Preventive Medicine* 36 (3): 291–99.
- Singh, P. 2010. *Performance Pay and Information: Reducing Child Malnutrition in Urban Slums*. London: London School of Economics.
- Springer, M. G., D. Ballou, L. Hamilton, V.-N. Le, J. R. Lockwood, D. F. McCaffrey, and M. P. B. M. Stecher. 2010. *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching*. Nashville, TN: Vanderbilt University, National Center on Performance Incentives.
- Tirole, J. 1994. "The Internal Organisation of Government." *Oxford Economic Papers* 46 (1): 1–29.
- Vandenberg, R., and C. Lance. 1992. "Satisfaction and Organisational Commitment." *Journal of Management* 18: 153–67.
- Weber, M. 1978. *Economy and Society*, Vol. 2. Berkeley, CA: University of California Press.
- Wilson, J. Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.
- Winters, M. A., G. W. Ritter, J. P. Greene, and R. Marsh. 2009. "Student Outcomes and Teacher Productivity and Perceptions in Arkansas." In *Performance Incentives: Their Growing Impact on American K–12 Education*, ed. M. G. Springer. Washington, DC: Brookings Institution Press.
- World Bank. 2012. *The World Bank's Approach to Public Sector Management 2011–2020: Better Results from Public Sector Institutions*. Washington, DC: World Bank.

