

MANAGEMENT, ORGANIZATIONAL PERFORMANCE, AND TASK CLARITY: EVIDENCE FROM GHANA'S CIVIL SERVICE *

IMRAN RASUL

DANIEL ROGGER

MARTIN J. WILLIAMS[†]

AUGUST 2020

*We gratefully acknowledge financial support from the International Growth Centre [1-VCS-VGHA-VXXXX-33301] and the World Bank's i2i trust fund financed by UKAID. We thank Dan Honig, Julien Labonne, Anandi Mani, Don Moynihan, Arup Nath, Simon Quinn, Raffaella Sadun, Itay Saporta-Eksten, Chris Woodruff, three anonymous referees, and seminar participants at Oxford, Ohio State, PMRC, and the EEA Meetings for valuable comments. Jane Adjabeng, Mohammed Abubakari, Julius Adu-Ntim, Temilola Akinrinade, Sandra Boatemaa, Eugene Ekyem, Paula Fiorini, Margherita Fornasari, Jacob Hagan-Mensah, Allan Kasapa, Kpadam Opuni, Owura Simprie-Duncan, and Liah Yecaló-Tecle provided excellent research assistance, and we are grateful to Ghana's Head of Civil Service, Nana Agyekum-Dwamena, members of the project steering committee, and the dozens of Civil Servants who dedicated their time and energy to the project. This study was approved by UCL's Research Ethics Committee. All errors remain our own.

[†]Rasul: University College London and the Institute for Fiscal Studies [i.rasul@ucl.ac.uk], Drayton House, 30 Gordon Street, London WC1H 0AX, United Kingdom; Rogger: World Bank Research Department [drogger@worldbank.org], 1818 H Street NW, Washington, D.C., United States; Williams: Blavatnik School of Government, University of Oxford [martin.williams@bsg.ox.ac.uk], Blavatnik School of Government, Radcliffe Observatory Quarter, Oxford OX2 6GG, United Kingdom.

Management, Organizational Performance, and Task Clarity: Evidence from Ghana's Civil Service

Abstract

We study the relationship between management practices, organizational performance, and task clarity, using observational data analysis on an original survey of the universe of Ghanaian civil servants across 45 organizations and novel administrative data on over 3600 tasks they undertake. We first demonstrate that there is a large range of variation across government organizations, both in management quality and task completion, and show that management quality is positively related to task completion. We then provide evidence that this association varies across dimensions of management practice. In particular, task completion exhibits a positive partial correlation with management practices related to giving staff autonomy and discretion, but a negative partial correlation with practices related to incentives and monitoring. Consistent with theories of task clarity and goal ambiguity, the partial relationship between incentives/monitoring and task completion is less negative when tasks are clearer *ex ante* and the partial relationship between autonomy/discretion and task completion is more positive when task completion is clearer *ex post*. Our findings suggest that organisations could benefit from providing their staff with greater autonomy and discretion, especially for types of tasks that are ill-suited to pre-defined monitoring and incentive regimes.

1 Introduction

The relationship between the management practices under which public servants operate and organizational performance is a central question for public administration [e.g. Lynn *et al.* 2000;

Meier and O'Toole 2002; Ingraham *et al.* 2003; Honig 2018]. Miller [2000] and Miller and Whitford [2016] describe the debate between Carl Friedrich and Herman Finer to distinguish between two broad schools of thought. On one hand, if bureaucrats are viewed as agents whose preferences diverge from their principals or who shirk their duties, then they should be managed through top-down tools of control such as monitoring and rewards/sanctions in order to elicit effort and minimize moral hazard [Finer 1941]. On the other hand, if bureaucrats are viewed as professionals trying to do their best for the public good, then public bureaucracies ought to delegate significant autonomy and discretion to bureaucrats, relying on their professionalism and expertise to deliver public services [Friedrich 1940]. The broad contours of these two approaches manifest themselves in many subsequent debates, with some authors (particularly, but not exclusively, in public administration) emphasizing the value of autonomy and discretion [Simon 1983; Rose-Ackerman 1985; Rainey and Steinbauer 1999; Carpenter 2001; Andersen and Moynihan 2016; Miller and Whitford 2016] while others (particularly, but not exclusively, in economics) focus on the importance of top-down monitoring and incentives [e.g. Duflo *et al.* 2012; various in Finan *et al* 2015]. Other authors have argued that the relationship between management and performance might depend on the nature of the agency's tasks or goals [Wilson 1989; Chun and Rainey 2005].

We contribute to this debate by studying the relationships between a broad spectrum of management practices and the full range of bureaucratic tasks, across 45 ministries and departments in the central government of Ghana. Ghana is a medium-sized, lower-middle income democracy, and ranks just below the global median (46th percentile) in government effectiveness by the Worldwide Governance Indicators.¹ The civil servants we study are largely in mid-level bureaucratic policymaking and oversight roles, rather than front-line implementation. To measure management practices, we conduct in-person surveys with the universe of professional-grade civil servants in Ghana's central government - nearly 3,000 individuals - and construct measures of management quality, adapting the methodological innovations of Bloom and Van Reenen [2007] and Bloom *et al.* [2012] from organizational economics to the public sector. This enables us to construct an overall index of management quality as well as sub-indices related to the use of incentives and

¹Worldwide Governance Indicators 2015 update, available at <https://info.worldbank.org/governance/wgi/Home/Reports>. See Kaufmann *et al* (2010) for methodological details.

monitoring, autonomy and discretion, and other practices. These are based not on subjective self-reported perceptions or whether specific *de jure* rules are in place, but on probing interviews benchmarked to an absolute scale that seek to capture the *de facto* management practices being used in reality, not just the *de jure* practices prescribed on paper.

To measure task completion and clarity, we collect, digitize, and hand-code quarterly and annual progress reports of each organization's planned activities against their actual achievements. This yields a database on the characteristics and completion of 3,620 tasks covering the entire range of bureaucratic activity, from procurement and infrastructure to policy development, advocacy, human resource management, budgeting, and regulation. Task completion is a fundamental aspect of bureaucratic performance and public service delivery, and allows us to compare different organizations against a common performance metric which is widely available even in contexts with relatively little existing performance information. We validate our measure against a sub-sample of audited projects to ensure that this self-reported bureaucratic data is truthful.

We first document that there is substantial variation across organizations on both management quality and task completion. This is despite the fact that all organizations operate under the same civil service law, regulations, and pay structure, are overseen by the same authorities, draw from the same pool of potential hires, and are located proximately to each other in the capital - in some cases in the same building. The existence of significant and systematic variation in both process-based (management practices) and output-based (task/project completion) measures of performance across organizations within government has implications for the theory and measurement of organizational performance, state capacity, and governance.

We then estimate the relationships between management practices and task completion, exploiting the fact that multiple organizations conduct each task type. We find that overall management quality and each management sub-index on its own is positively correlated with task completion, with varying degrees of significance, and that the sub-indices are positively correlated with each other. However, organizations do not implement management practices in isolation, they implement a portfolio of practices, so it is important to estimate their relationships with task completion jointly. When we do this, we show that autonomy/discretion and incentives/monitoring

have *opposing* signs: a one standard deviation increase in autonomy/discretion-related practices is associated with a 12 percentage point increase in the likelihood a task is fully completed; in contrast, a one standard deviation increase in incentives/monitoring-related management practices is associated with a decrease of 4 percentage points in the likelihood it is fully completed.

We further investigate the relationships between management practices and task completion by examining how these relationships vary with the *ex ante* clarity of task definition and the *ex post* clarity of actual achievement. Bureaucratic tasks are *ex ante* clear when the task can be defined in such a way as to create little uncertainty about what is required to complete the task, and are *ex post* clear when a report of the actual action undertaken leaves little uncertainty about whether the task was effectively completed. We hypothesize that the top-down control strategies of incentives and monitoring should be relatively more effective when tasks are easy to define *ex ante*, because it is easier to specify what should be done and construct an appropriate monitoring regime. On the other hand, empowering staff with autonomy and discretion should be relatively more effective when tasks are unclear *ex ante*, as well as when the actual achievement of the task is clear *ex post* (because *ex post* clarity makes it easier to detect abuse of discretion). We find strong evidence consistent with this mechanism: for tasks with below-median *ex ante* clarity, one standard deviation increases in autonomy/discretion- and incentives/monitoring-related practices are associated with a 21 percentage point increase and 14 percentage point decrease in task completion, respectively. However, there is no significant difference between management practices for tasks with above-median *ex ante* clarity. Similarly, for tasks with above-median *ex post* clarity, a one-standard deviation increase in autonomy/discretion (incentives/monitoring) is associated with a 34 percentage point increase (1 percentage point decrease) in task completion, consistent with our hypotheses.

Our findings are consistent with theories that emphasize that monitoring and incentive systems can backfire in contexts (such as core civil service policymaking tasks) intensive in multi-tasking, coordination, and instability, and where tasks are differentially observable (Dixit 2002; Honig 2018). Since these coefficients have to be interpreted relative to each other, the implication is that Ghana's civil service organizations are under-providing autonomy and discretion to their staff

relative to their use of monitoring and incentives, particularly on tasks that are *ex ante* unclear or *ex post* clear. That is, organizations appear to be over-balancing their management practice portfolios towards top-down control measures at the expense of entrusting and empowering the professionalism of their staff. These results are particularly striking in the context of a lower-middle income country such as Ghana, where concerns about abuse of discretion among public officials are especially salient among academics and citizens alike.

We also make a methodological contribution by showing that, since the underlying management practice sub-indices are positively correlated, estimating the relationships between a single set of management practices and performance without accounting for other practices - as is common in the empirical literature - would lead to significant omitted variable bias. We regard this as empirical justification for our approach of seeking to understand these relationships simultaneously across a broad spectrum of management practices, rather than focusing more narrowly on the effects of a single type of management practice, with important implications for other studies of bureaucratic management and service delivery. Although our context does not allow for clean causal identification of these effects, the empirical support we find for the task clarity mechanism makes it unlikely that our findings are driven entirely by reverse causality. Indeed, our goal of investigating these relationships across such a broad spectrum of management practices under actual implementation conditions would be limited in the more controlled or narrow settings that would enable clean causal identification of results. Our findings are thus an important complement to (quasi-)experimental studies in expanding our understanding of the relationship between management and performance in state bureaucracies.

The remainder of this article proceeds as follows. Section 2 presents a theoretical framework for studying the relationship between management practices and output, with a focus on the longstanding debate between incentives/monitoring-led approaches and autonomy/discretion-led approaches. Section 3 discusses our empirical context and data, and Section 4 shows our descriptive results on variation in management and productivity within Ghana's Civil Service. Section 5 presents our empirical method and main results, Section 6 investigates mechanisms, and Section 7 examines external validity. Section 8 concludes by discussing implications for research and policy.

2 Theory: Management, Task Clarity, and Performance

As Miller [2000] and Miller and Whitford [2016] discuss, one common view of bureaucrats (following Finer 1941) is that they need to be monitored and rewarded and sanctioned. Accordingly, the relationship between performance and the use of top-down strategies of management control like incentives and monitoring has long been a central question for public administration and economics [Wilson 1989; Dixit 2002; Boyne and Hood 2010]. The relationship between these approaches and bureaucratic output is theoretically ambiguous. On one hand, management through incentives and monitoring may elicit agent effort and discourage behavior that is misaligned with a principal's preferences. On the other hand, the nature of public sector work - involving multiple goals, difficult-to-measure outputs and outcomes, extensive coordination, and environmental uncertainty - may limit the scope for the effective use of incentives and monitoring or result in efforts at top-down control backfiring. A large literature focuses in specifically on the effect of performance-related pay on performance, documenting both successes and failures [e.g. Miller and Whitford 2007; Dahlstrom and Lapuente 2009; Perry *et al* 2009; Hasnain *et al* 2014], and another set of studies documents how the actual ways in which performance information in bureaucracies is used often deviate from its idealized form [e.g. Heinrich 1999; Moynihan *et al* 2011]. Quantitative evaluations of the effects of incentive schemes or rigid agent monitoring schemes with attached sanctions [e.g. Duflo *et al* 2012; various in Finan *et al* 2015] on performance have tended to focus solely on frontline agents and on the implementation of specific schemes in controlled conditions. Studies on the use of incentives and monitoring more broadly in the core civil service [e.g. Bevan and Hood 2006; Kelman and Friedman 2009; various in Boyne and Hood 2010] have tended to focus on specific agencies or performance measures, resulting in questions about the generality of their findings.

On the other hand, perspectives (following Friedrich 1940) that emphasize the professionalism of bureaucrats are bifurcated between studies of the effects of autonomy of government organizations from politicians and studies of the effects of discretion of frontline bureaucrats. While there is broad consensus that organizational autonomy from political influence is important for performance [Rainey and Steinbauer 1999; Carpenter 2001; Moe 2013; Andersen and Moynihan

2016; Miller and Whitford 2016], quantitative studies of individual-level discretion in economics and political science have mainly focused on its potential downsides in terms of discrimination [Einstein and Glick 2017] or corruption [Olken and Pande 2012]. On other hand, a large literature on bottom-up policy implementation [e.g. Thomann et al 2018] emphasizes the potential positive effects of discretion on policy implementation among frontline bureaucrats, and Rasul and Rogger [2018] find a positive association between autonomy and project completion in Nigeria. However, we are aware of few studies that attempt to measure the relationship between performance and the extent to which mid-level, core civil servants are able to use local information and make flexible decisions *within organizations* and apply it to tasks across the full spectrum of bureaucratic activity, which is the empirical focus of this study. We thus refer to autonomy/discretion throughout our study, to indicate that our operationalization falls in between the conventional use of autonomy as an organization-level characteristic and discretion as pertaining to individual frontline bureaucrats. Importantly, however, our perspective on autonomy/discretion shares its central hypothesis with these more common approaches: the nature of many core public sector tasks may require bureaucrats to make flexible judgments using local information, and thus organizations must rely on bureaucrats' professionalism to make good decisions and perform effectively.

We conceptualize management in public organizations as a portfolio of practices that correspond to different aspects of management, each of which may be implemented more or less well. Bureaucracies may differ in their intended management styles, i.e. what bundle of management practices they are aiming to implement, and may also differ in how well they are executing these practices. An organization may execute a given set of management practices in a consistent and coherent fashion, or in an disorganized and *ad hoc* way. While research on the effects of management practices often focus on a single practice in isolation, in reality bureaucracies are executing a wide range of management practices simultaneously. Whether deliberately or through inaction, organizations adopt policies on the (non-)use of targets, (non-)monitoring of key performance indicators, (dis-)allowance of discretion in decisionmaking, (non-)reward of staff performance, and so on. This implies that management can be viewed as a joint set of choices across a range of practices.

Finally, our approach assumes that what matters for determining organizational performance is not the *de jure* practices by which an organization avows itself to be managed, but the *de facto* practices that are actually used on a day-to-day basis within the organization. Management thus differs across organizations both in style and in quality, so debates over the merits of a particular management practice or style thus also require an understanding of how well the practice(s) are likely to be implemented under real-world conditions, so questions of the choice of management practices are not separable from issues of policy implementation.

The conceptual distinction we make (following Miller 2000 and Miller and Whitford 2016) between top-down incentives and monitoring and professionalism-oriented autonomy and discretion is a different one than that made by another big-picture debate on public sector management: that between hierarchy-oriented Weberian bureaucracy and market-oriented New Public Management (NPM) styles of governance. While the Weberian-versus-NPM debate has received prominence in the literature due to its correspondence to the historical evolution of management in public organizations, our distinction corresponds instead to two divergent views of how best to manage bureaucratic agents: to what extent should they be managed with the carrot and the stick, and to what extent should they be empowered with the discretion associated with other professions? We view this question as cross-cutting the Weberian/NPM distinction. Throughout the article, we focus on our functional distinction without trying to situate it within the Weberian-versus-NPM debate.

Task clarity and related concepts have long been an important element of public administration theory, because the differing nature of bureaucratic tasks implies that different management approaches might be effective in different situations. Wilson [1989, Ch. 9] makes a widely referenced distinction between the visibility of agency outputs and outcomes, and constructs a two-by-two typology of agency types: production agencies, which have high output and outcome visibility; procedural agencies, which have high output visibility but low outcome visibility, and thus are managed using tight top-down controls on activities and processes; craft agencies, which have low output visibility but high outcome visibility, and thus can be delegated with significant autonomy; and coping agencies, which have low visibility of both outputs and outcomes and are thus difficult

to manage effectively. Relatedly, Chun and Rainey [2005] define four types of goal ambiguity (mission comprehensiveness, directive, evaluative, and priority), and show that goal ambiguity is negatively associated with various performance measures, and Romzek [2000] discusses how different accountability mechanisms (e.g. hierarchical vs. professional) are appropriate for different bureaucratic scenarios.

In our analysis we focus at the level of particular tasks (rather than assuming agencies have a single main task type), and distinguish between *ex ante* and *ex post* task clarity. We hypothesize that *ex ante* and *ex post* task clarity may impact the ability of organizations to use different types of management practices to effectively deliver it. In particular, some types of tasks might be *ex ante* easy to specify, and others might be *ex post* easy to measure. For tasks that are *ex ante* unclear, the design of effective incentives and monitoring schemes is likely to be harder, all else equal, since it is unclear what bureaucrats should be aiming for and how to measure it. On the other hand, when task achievement is clear *ex post*, granting bureaucrats discretion over how to implement tasks is likely to be relatively more effective since abuse of that discretion will be easier to detect *ex post*. While focused on tasks rather than agencies, our distinction between *ex ante* and *ex post* clarity is conceptually similar to Chun and Rainey’s [2005] directive and evaluative goal ambiguity distinction, and to Wilson’s [1989] distinction between procedural and craft agencies.

Our theoretical discussion can be summarized in two sets of hypotheses. The first set refers to whether top-down monitoring- and incentive-based management approaches are more strongly associated with task completion on average than bottom-up autonomy- and discretion-based approaches. Since there are theoretical arguments on either side of this debate, we capture this with two opposing hypotheses:

- *H1a: Intensive use of management practices related to monitoring and incentives is more strongly associated with task completion than intensive use of management practices related to autonomy and discretion, all else equal.*
- *H1b: Intensive use of management practices related to autonomy and discretion is more strongly associated with task completion than intensive use of management practices related to monitoring and incentives, all else equal.*

Our second hypothesis refers to how these relationships are moderated by *ex ante* and *ex post* task clarity:

- *H2: Monitoring- and incentive-based management approaches will be more strongly associated with completion for tasks with relatively high ex ante clarity, and autonomy- and discretion-based management approaches will be more strongly associated with completion for tasks with relatively high ex post clarity, all else equal.*

A priori, either H1a or H1b (but not both) could be true, so we test these hypotheses against each other empirically. Given its underlying theoretical rationale, we expect H2 to hold regardless of which of H1a or H1b we find greater support for.

Our article is related to important literatures in public administration, economics, and political science on organizational performance and state capacity. Within public administration, most of this literature is focused on high-income countries, and due to the well-known challenges of measuring organizational performance [Boyne and Walker 2005; Talbot 2010] is focused on one type of organization, such as school districts or police departments [e.g. Meier and O’Toole 2002; Nicholson-Crotty and O’Toole 2004; Andersen and Mortensen 2010]. Investigations of performance across a fuller range of public sector activities has often relied on subjective measures of organizational performance, leading to concerns about common source bias [Meier and O’Toole 2012]. Our measure of task completion aims to build on this existing literature by establishing an output-based organizational performance metric that is comparable across organizations and not based on subjective surveys, while also enabling us to cover the entire spectrum of sectors and activities undertaken by government - much as we aim to measure organizations’ full portfolios of management practices rather than just the use of one specific practice.

A large literature in public administration, economics, political science, and sociology examines variation in bureaucratic quality across and within states. While many studies analyze cross-national variation in state capacity and overall public sector management [e.g. Kaufmann *et al.* 2010], more relevant to our study is literature which examines within-country, cross-agency variation in performance [e.g. Ingraham *et al* 2003]. Outside of OECD contexts, this has taken the form primarily of case study-based literature that documents the existence of “islands of

excellence” or “pockets of effectiveness” through rich case studies [Tendler 1997; Leonard 2010; McDonnell 2017]. However, these small-N analyses lead to concerns about the generalizability of their findings. Among quantitative studies of within-country variation outside OECD contexts, there is a small but growing literature that documents variation across organizations in capacity using input-based measures [Gingerich 2013; Bersch *et al* 2016]. A limitation of this existing literature is that these measures of capacity are typically divorced from consideration of the type of management practices each organization is implementing, so that there is little evidence on the potential links between the style and the quality of management postulated by this framework. Owusu [2006] comes closer to our focus by using expert perception surveys to measure variations in performance of government organizations in Ghana, but with the usual potential measurement issues associated with subjective measures. Using our novel measures of task completion and management practices, we contribute to this literature on the extent of variation in organizational performance within governments in low- and middle-income countries, and also provide new tools that can be used in high-income contexts.

3 Context and Data

Ghana is a lower-middle income country home to 28 million individuals, with a central government bureaucracy that is structured along lines reflecting both its British colonial origins and more presidentialist post-independence reforms. We study the universe of 45 Ministries and Departments in the Civil Service. The headquarters of these organizations are all located in Accra, but have responsibility for public projects and activities implemented nationwide.² Ministries and Departments are overseen by the Office of the Head of Civil Service (OHCS), which is responsible for personnel management and performance within the civil service. OHCS coordinates and decides on all hiring, promotion, transfer, and (in rare circumstances) firing of bureaucrats across the service. OHCS develops and promulgates official management regulations and processes, but Min-

²Ghana distinguishes between the Civil Service and the broader Public Service, which includes dozens of autonomous agencies under the supervision but not direct control of their sector ministries, as well as frontline implementers such as the Police Service, Education Service, etc. Our sample is restricted to the headquarters offices of Civil Service organizations.

istries' and Agencies' compliance with these is imperfect, with the result that actual management practices are highly variable across organizations - as in many countries worldwide [Tendler 1997; Ingraham *et al* 2003; Gingerich 2013; Bersch *et al* 2016]. All these Ministries and Departments have the same statutory levels of autonomy and political oversight structures, which facilitates our analysis (without limiting the scope of the theoretical predictions). While all countries' civil services differ in certain aspects, the basic structure of Ghana's civil service is common across many Anglophone countries, and the types of challenges civil service organizations in Ghana face (e.g. lack of resources, environmental uncertainty, political-bureaucratic tensions, employee abuse of discretion) are widely shared ones worldwide.

Our analysis focuses on the professional grades of technical and administrative officers within these Ministries and Departments. We therefore exclude grades that cover cleaners, drivers, most secretaries, etc. On average, each organization employs 64 bureaucrats of the type we study (those on professional grades). We designate bureaucrats as being at either a senior or non-senior level. Senior bureaucrats are those that classify themselves as a 'Director (Head of Division) or Acting Director' or as a 'Deputy Director or Unit Head (Acting or Substantive)'. By this definition, the number of non-senior bureaucrats overseen by each senior bureaucrats (i.e. the span of control) is 4.52 on average, but again there is considerable variation across Ministries.³

Around 45% of bureaucrats are women, 70% have a university education, and 31% have a postgraduate degree (senior officers are more likely to be men, and to have a postgraduate degree). As in other state organizations, civil service bureaucrats enjoy stable employment once in service: the average bureaucrat has 14 years in service, with their average tenure in the current organization being just under 9 years. Appointments are made centrally by OHCS, bureaucrats enjoy security of tenure, and transitions between bureaucracies are relatively infrequent.

Our analysis is based on two data sources. First, we hand-coded quarterly and annual progress reports from Ministries and Departments, covering tasks ongoing between January and December 2015. As detailed below, these reports enable us to code the individual tasks under the remit of

³In Ghana, grades of technical and administrative bureaucrats are officially referred to as 'senior' officers while grades covering cleaners, drivers etc. are referred to as 'junior' officers, regardless of their tenure or seniority. While we restrict our sample to 'senior' officers in the formal terminology, throughout we use the terms senior and non-senior in their more colloquial sense to refer to hierarchical relationships within the professional grades.

each organization, and the extent to which they are initiated or successfully completed.

Second, we surveyed 2971 bureaucrats from all 45 civil service organizations over the period August to October 2015. As detailed below, civil servants were questioned on topics including their background characteristics and work history in service, job characteristics and responsibilities, engagement with stakeholders outside the civil service, perceptions of corruption in the service, and their views on multiple dimensions of management practices.

3.1 Coding Task Completion

Worldwide, civil service bureaucracies differ greatly in whether and how they collect data on their performance, and few international standards exist to aid cross country comparisons. To therefore quantify the delivery of public sector task completion in our context, we exploit the fact that each Ghanaian civil service organization is required by OHCS to provide quarterly and annual progress reports. While organizations differed in their reporting formats and coverage, most reports included a comprehensive table of all tasks, outputs, and projects that were to be undertaken by the organization during the reporting period. We are thus able to use the progress reports of 30 of the 45 Ministries and Departments (our civil servant survey covers 2247 bureaucrats in these 30 organizations). Figure A1 provides a snapshot of a typical progress report and indicates the information coded from it, and the Online Appendix discusses the coding process in detail.⁴ Of the 15 organizations for which we were unable to code progress reports, the barrier in most cases was that the organizations produced reports in a purely narrative form (i.e. without a table) which made it impractical to separate the work undertaken into discrete tasks that could be used for coding. We discuss below how we validated the accuracy of reports we were able to code.

Progress reports cover the *entire* range of bureaucratic activity. While some of these tasks are public-facing outputs, others are purely internal functions or intermediate outputs. For brevity, we refer to the activities, outputs, projects, functions, and processes reported on collectively as ‘tasks’.⁵ We were able to use these progress reports to identify 3620 tasks underway during

⁴Where an organization produced multiple reports during this time (e.g. a mid-year report and an annual report), we selected the latest report produced during the year to include in our sample.

⁵Unfortunately, the reports do not distinguish between activities and outputs in the theoretical sense in which the terms are used by some logical frameworks or theories of change, instead just reporting them as lists of tasks

2015. The tasks undertaken by each organization in a given year are determined through an annual planning and budgeting process jointly determined between: the core executive, mainly the Ministry of Finance and the sector minister representing government priorities; the organization’s management, based in large part on consultatively developed medium-term plans; and ongoing donor programs. This schedule of tasks is formalized in the organization’s annual budget (approved by Parliament) and annual workplan. The quarterly and annual reports that we use to code task completion thus detail the task that the organization’s workplan committed the organization to working on during the time period of study.

The Online Appendix describes how we hand-coded and harmonized the information to measure task completion across the Ghanaian civil service. Three key points are of note in relation to this process. First, each quarterly progress report was codified into task line items using a team of trained research assistants and a team of civil servant officers seconded from the Management Services Department (MSD), an organization under OHCS tasked with analyzing and improving management in the civil service. MSD officers are trained in management and productivity analysis and frequently review organizational reports of this nature, making them ideally suited to judging task characteristics and completion.

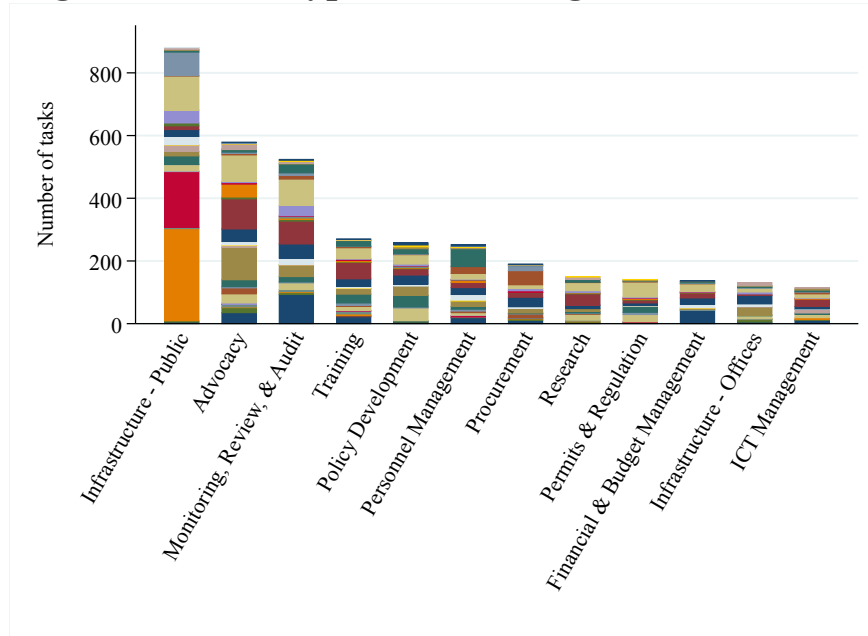
Second, coders were tasked to record task completion on a 1-5 scoring grid, where a score of one corresponds to, “No action was taken towards achieving the target”, three corresponds to, “Some substantive progress was made towards achieving the target. The task is partially complete and/or important intermediate steps have been completed”, and a score of five corresponds to, “The target for the task has been reached or surpassed.” Tasks can be long-term or repeated (e.g. annual, quarterly) tasks. There were at least two coders per task.⁶ For example, a score of 1 was given for the task “Final accounts on 2nd tranche of loan component prepared and endorsed” with associated completion status of “Not undertaken”, and a score of 5 was given for the task “Prepare 2015 1st quarter Progress Report” with associated completion status of “2015 1st Quarter report prepared and submitted to CAGD Management and Ministry of Finance”.

the organization has to undertake.

⁶Given the tendency for averaging scores across coders to reduce variation, for our core analysis we use the maximum and minimum scores to code whether tasks are fully complete/never initiated respectively. Figure A6 shows that alternative approaches to aggregating scores yield similar results.

Third, as progress reports are self-compiled by bureaucracies, an obvious concern is that low performing bureaucracies might intentionally manipulate their reports to hide the fact. To check the validity of progress reports, we matched a sub-sample of 14% of tasks from progress reports to task audits conducted by external auditors through a separate exercise undertaken by OHCS. Auditors are mostly retired civil servants, overseen by OHCS, and they obtain documentary proof of task completion. For matched tasks, 94% of the completion levels we code are corroborated based on the qualitative descriptions of completion in audits.⁷

Figure 1: Task Types Across Organizations



Notes: The task type classification refers to the primary classification for each output. Each colour in a column represents an organization implementing tasks of that type, but the same colour across columns may represent multiple organizations. Figures represent all 30 organizations with coded task data.

The types of tasks included in the data are revealing of the full scope of activity of bureaucracies. Figure 1 shows the most common task type in Ghanaian central government bureaucracies relates to the construction of public infrastructure such as roads, boreholes, and schools (e.g. "Identify bungalows and initiate procurement process", "Rehabilitation of Bosomkyekye-Ouagadugu [sic]" road, "Construction of Secondary Data Centre at Kumasi"), comprising 24%

⁷Among the handful of non-corroborated tasks, the lowest "true" completion rate was a 3 out of 5, indicating that the rare instances of misreporting were relatively minor.

of tasks. Other common task types are advocacy (e.g. "Sensitize printers and suppliers on the procurement law and packaging", "Kaizen Forum Organized", "Talk shows in four rural district markets in the region on the GIPC act held"), comprising 16% of tasks, and monitoring, review, and audit (e.g. "Collate 2nd quarterly reports on the Ministry's work plan for Management meetings", "Preparation of 2014 Annual Report", "Conduct second phase of Housing Audit"), comprising 14%.

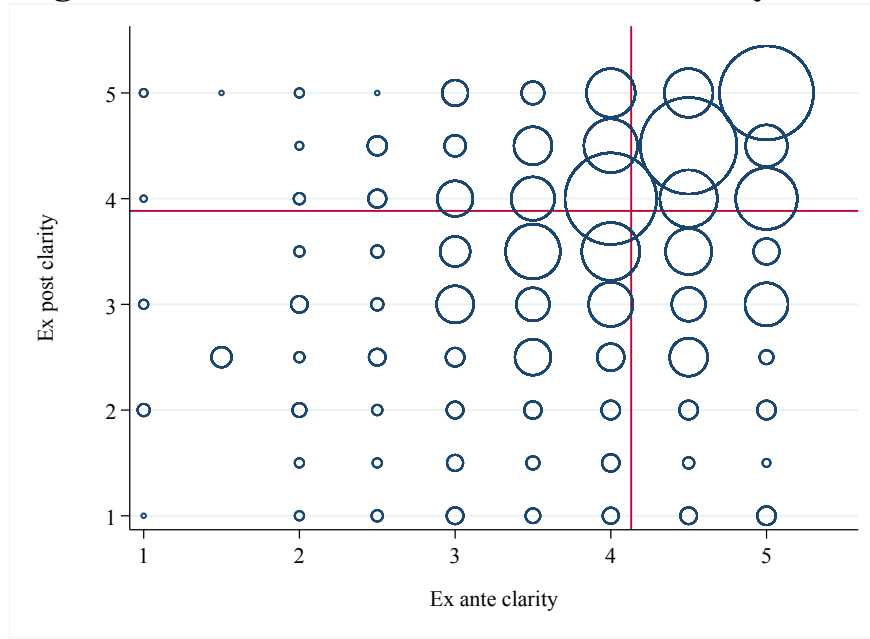
Figure 1 demonstrates two salient facts that motivate our analysis. First, each task type is implemented by many different organizations and each organization implements multiple task types, allowing us to disentangle the performance of bureaucracies from the types of tasks they undertake. Second, prominent activities on which previous studies of management and organizational performance have been based, such as procurement or public infrastructure development, comprise only a small minority of the full range of tasks undertaken by the public sector. While there are of course analytical advantages to focusing on a single type of task, we view this evidence as motivation for our focus on the full range of tasks undertaken by the public sector.

In addition to coding task completion, our coders also rated the clarity of the expected level of achievement for the task in the reporting period, i.e. the *ex ante* task clarity. They also rated the clarity of the actual description of what was done, i.e. the *ex post* task clarity. Their rating was based on the clarity of the information on the *ex ante* target and *ex post* actual achievement contained in the report itself, which likely corresponds to clarity for managers themselves since these reports are designed primarily as management tools. As with task completion, coders scored these variables on a 1-5 scale. A score of one corresponded to an *ex ante* target that was "undefined or so vague it is impossible to assess what completion would mean", a score of three corresponded to a target that "is defined, but with some ambiguity", and a score of five corresponded to "no ambiguity over the target - it is precisely quantified or described". An example of a task coded as a 1 for *ex ante* clarity was "Improved compliance with environmental laws and regulations", while "Organise review meeting on 2014 APR" was coded as a 5. These benchmarks were analogously defined and coded for *ex post* clarity.

Figure 2 plots the variation in *ex ante* and *ex post* task clarity. While the mean task was coded

approximately four out of five on both measures, only 22.7 percent of tasks were given the highest rating of five for *ex ante* clarity and 19.1 percent were rated five for *ex post* clarity, indicating that the vast majority of tasks were not perfectly clear. As one would expect, *ex ante* and *ex post* clarity are positively correlated ($\rho = 0.36$), but not perfectly so, so there are significant numbers of tasks that were relatively clearly specified *ex ante* but not *ex post* (and vice versa).

Figure 2: *Ex Ante* and *Ex Post* Task Clarity



Notes: Circle size is proportional to the number of tasks that fall within each bin of width 0.5. Red lines indicate mean values for each measure of task clarity.

Coders also coded a range of other task characteristics, such as the routineness of a task, whether the task is a single activity or a bundle of interconnected activities, and the level of coordination with external stakeholders the task requires, which we use as control variables in our analysis.

3.2 Measuring Management

To measure the quality of management practices, we draw on the methodological innovations of Bloom and Van Reenen [2007, 2010] and Bloom *et al.* [2012] (BSVR henceforth) which have

transformed the empirical study of management practices in organizational economics. BSVR use structured telephone interviews with managers to score management quality on a 1-5 scale across 15 different management practices, incorporating a number of methodological innovations order to mitigate the recognized biases associated with the common practice of using subjective, self-reported measures of organizational performance [Meier and O’Toole 2012]. This methodology was initially developed for use in measuring management in private firms in the manufacturing sector, but has subsequently been adapted and extended to organizational settings as diverse as hospitals, schools, and NGOs, in both developed and developing countries [Bloom *et al* 2014].

We adapted BSVR’s methodology to be administered as an in-person survey, and to cover fourteen practices across six dimensions of management practice that are relevant for the public sector: roles, flexibility, incentives, monitoring, staffing and targets. Following Rasul and Rogger [2018], we aggregate these into the three sub-indices of autonomy/discretion, incentives/monitoring, and other (which includes the residual practices, staffing and targets). The autonomy/discretion sub-index comprises the topics of role and flexibility, which cover the extent of discretion bureaucrats have to decide how to go about implementing and achieving tasks, the extent to which they have flexibility to adapt their work to contextual specificities, and their flexibility in being able to generate and adopt new work practices.⁸ The incentives/monitoring sub-index covers the extent to which poor or good performance is rewarded or punished (financially or non-financially), and the extent of the use of both individual- and team-level performance metrics. Our residual ‘other’ sub-index covers the remaining questions, which relate to staffing and the use of targets in the organization.⁹ We use this residual sub-index of other practices as a control only, and do not attach a substantive interpretation since it falls outside our theoretical focus. Table 1 summarizes the construction of these management sub-indices, and Table A1 gives full details of each management related question, by topic, as well as the scoring grid used by our enumerators for each question.

⁸Note that autonomy in this sense refers to autonomy within an organization (of individuals and teams from senior management) rather than the autonomy of the organization (from political principals), and that all organizations in our sample have the same degree of legal and procedural autonomy in the organizational sense.

⁹While the use of targets is often associated with top-down management approaches that are also intensive in incentives and monitoring, target-setting is potentially also an important element of management through autonomy and discretion. We therefore leave it in our residual category.

Table 1: Summary of Management Sub-Indices

Management Sub-Index	Topic	Practice Description
Autonomy/ Discretion	<i>Roles</i>	<ul style="list-style-type: none"> • The extent to which senior staff make substantive contributions to the policy formulation and implementation process • The extent to which senior staff are given discretion to carry out assignments in their daily work
	<i>Flexibility</i>	<ul style="list-style-type: none"> • The extent to which the division makes efforts to adjust to the specific needs and peculiarities of communities, clients, or other stakeholders • The extent to which the division is flexible in terms of responding to new and improved work practices
Incentives/ Monitoring	<i>Performance Incentives</i>	<ul style="list-style-type: none"> • The extent to which under-performance would be tolerated, given past experience • The extent to which staff are disciplined for breaking the rules of the civil service • The extent to which the performance of individual officers is tracked using performance, targets, or indicators and rewarded (financially or non-financially)
	<i>Monitoring</i>	<ul style="list-style-type: none"> • The extent to which the division tracks how well it is performing and delivering services using indicators
Other	<i>Staffing</i>	<ul style="list-style-type: none"> • The extent to which efforts are made to attract talented people to the division and retain them • The extent to which people are promoted faster based on their performance • The extent to which the burden of achieving targets is evenly distributed across different officers • The extent to which senior staff try to use the right staff for the right job
	<i>Targeting</i>	<ul style="list-style-type: none"> • The extent to which the division has a clear set of targets derived from the organization's goals and objectives that affect individuals' work schedules • The extent to which staff know what their individual roles and responsibilities are in achieving the organization's goals when they arrive at work each day

Notes: Interviews were conducted using open questions to initiate discussion on each practice, followed by probing follow-up and requests for examples, after which the interviewer would score each practice on a 1-5 scale. See text for further details of interview method and Appendix A for full details of practices and scoring grid.

Following BSVR, for each question enumerators would first ask what practices were used in an open-ended way, then probe respondents' responses and ask for examples in order to ascertain what practices are actually in use (much as a qualitative interview would), as opposed to

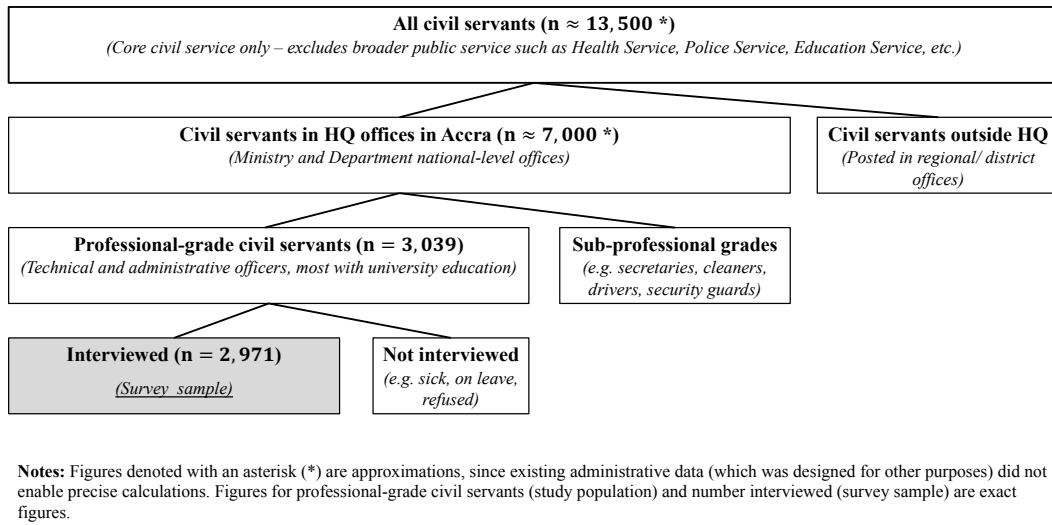
simply asking for respondents’ perceptions of management quality. Interviewers would then use this information to score each practice on a continuous 1-5 scale, where 1 represents non-use or inconsistent/incoherent use of that practice within the organization, and 5 represents strong, consistent, and coherent use of that practice. To further anchor the scores and provide comparability across organizations, the scoring grid for each practice is benchmarked to actual descriptions of the practices in use. This improves on the more commonly used Likert-style measures of perceptions of management practices, which are vulnerable to differential anchoring across respondents and organizations.

To undertake these in-person, interview-style surveys, we collaborated closely with Ghana’s Office of Head of Civil Service (OHCS). We first recruited survey team leaders from the private sector, with an emphasis on previous experience of survey work in Ghana. We worked closely with the team leaders to give them an appreciation and understanding of the practices and protocols of the public service. OHCS then seconded a group of 34 junior public officials with pre-existing experience of public sector work to act as enumerators. They worked on rotation across the survey period, so that at on any given day approximately two-thirds were conducting interviews. The Head of Service ensured their commitment to the survey process by stating the research team would monitor interviewer performance and that these assessments would influence future posting opportunities. We trained the team leaders and public officials jointly, including intensive practice interview sessions, before undertaking the first few interviews together in order to harmonize and calibrate assessments.

Figure 3 summarizes the participant flow and study sample. Within the Civil Service as a whole, we first excluded officers working outside of each organization’s headquarters staff in regional or district offices (officers working in office annexes located in physically separate buildings from the main headquarters but administratively part of the main headquarters were included). We then excluded support staff such as cleaners, drivers, secretaries, and security guards who are classified as “sub-professional” grades by the Civil Service. To compile the list of eligible professional-grade civil servants based in headquarters offices, our survey team worked with each organization’s human resources directorate. This yielded a total of 3039 staff eligible to be inter-

viewed across 45 organizations. Over the period from August to November 2015¹⁰, the enumerators interviewed 2971 of these civil servants. This constitutes 98% of all eligible staff in these organizations, with the remainder mostly having been out of the office during the survey period. Interviews were conducted in person, but were double-blind in the sense that we ensured that interviewers had never worked in the organizations in which they were interviewing and did not know their interviewees, and likewise interviewees did not know their interviewers.

Figure 3: Study Population and Sample



We convert the scores on each management practice into normalized z-scores by taking un-weighted means of the underlying z-scores (so are continuous variables with mean zero and variance one by construction), using the most senior bureaucrat in each division since these officers have an awareness of management practices at senior management level as well as their day-to-day implementation.¹¹ Greater autonomy/discretion for staff within an organization thus corresponds to higher scores on this sub-index, and for the incentives/monitoring measure the provision of stronger incentives/monitoring corresponds to higher scores. Of course, while these scores do correspond to real variation in the qualitative management practices being used, and our quantitative

¹⁰This survey period was a time when the Civil Service was operating normally, so we do not expect that there was anything peculiar about this time period that would have influenced our results. The country's economy was stable, there were no major structural reforms of the Civil Service ongoing or planned, and elections were not scheduled until December 2016.

¹¹The median (mean) number of senior managers per organization is 13 (20).

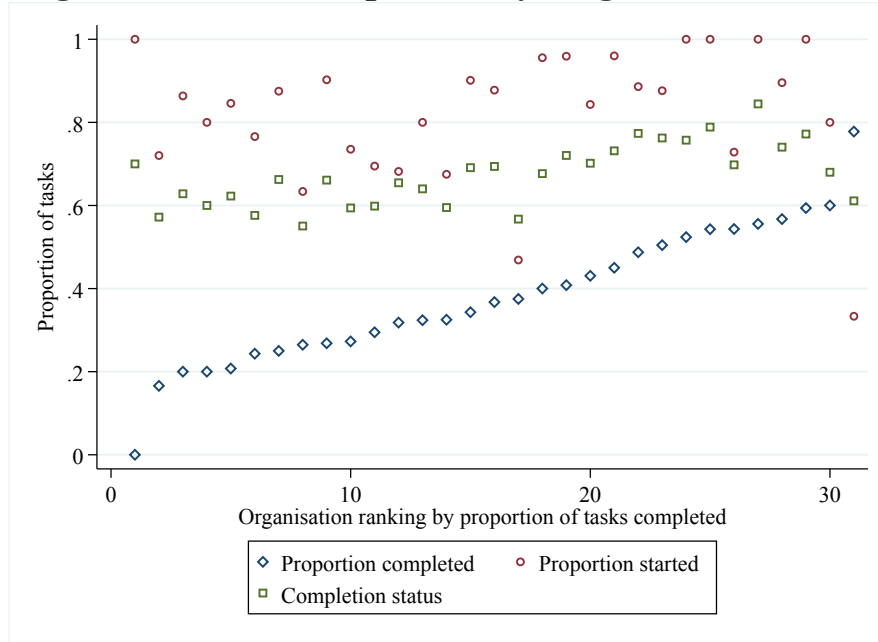
coding represents commonly accepted notions of the quality of each practice, we do not presume that higher scores always correspond to “good management” in the sense that they necessarily improve performance. Rather, these relationships are what we try to estimate empirically in our main results section below.

4 Variation in Task Completion and Management

These datasets allow us to provide rich new descriptive evidence on within-country, cross-organization variation in the completion of tasks, as well as in the management practices with which organizations are run. Since these descriptives are both novel and noteworthy, we first present evidence of this variation in each variable before going on to examine the relationship between them.

Figure 4 shows that there is substantial variation in task completion across civil service organizations, whether measured in the proportion of tasks started or finished or in the average score on our 1-5 scale. To quantify this variation, we note that the 75th percentile organization has an average completion rate 22% higher than the 25th percentile organization. This task-based measure provides powerful evidence of the variation in actual bureaucratic performance across bureaucracies within a government, thus building on and extending the existing input-, survey-, and perception-based measures of this variation [Ingraham *et al.* 2003; Gingerich 2013; Bersch *et al.* 2016].

Figure 4: Task Completion by Organization



Notes: Multiple coders assessed an task such that here we take the minimum assessment of initiation and the maximum assessment of completion, so that it is possible for proportion started to be lower than proportion completed (as it is for one organization). Completion status is a continuous 1-5 score for each task, here rescaled to 0-1.

Figure 5 shows the average completion status of different task types. Average completion rates are clustered between 2.97 (Training) and 3.45 (Personnel Management), and while there are some statistically significant differences across task types, this variation is much less dramatic than the range of variation across organizations shown in Figure 2. This perhaps reflects that organizations aim for some degree of balance across task types in the level of ambition in the setting of tasks. Nonetheless, we will control for task type in our later analysis to ensure that our results are not being driven by the variation in task incidence across organizations.

Figure 5: Task Completion by Task Type

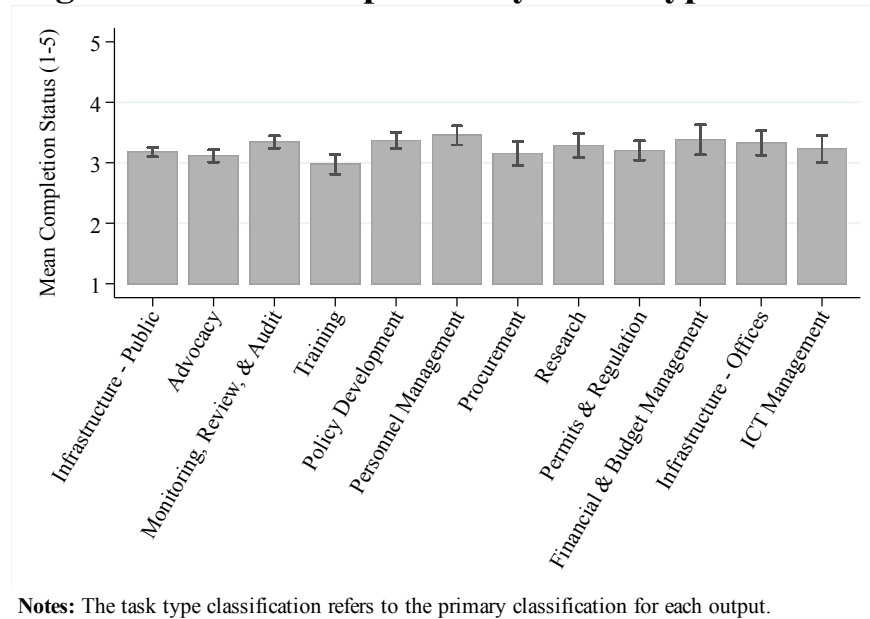
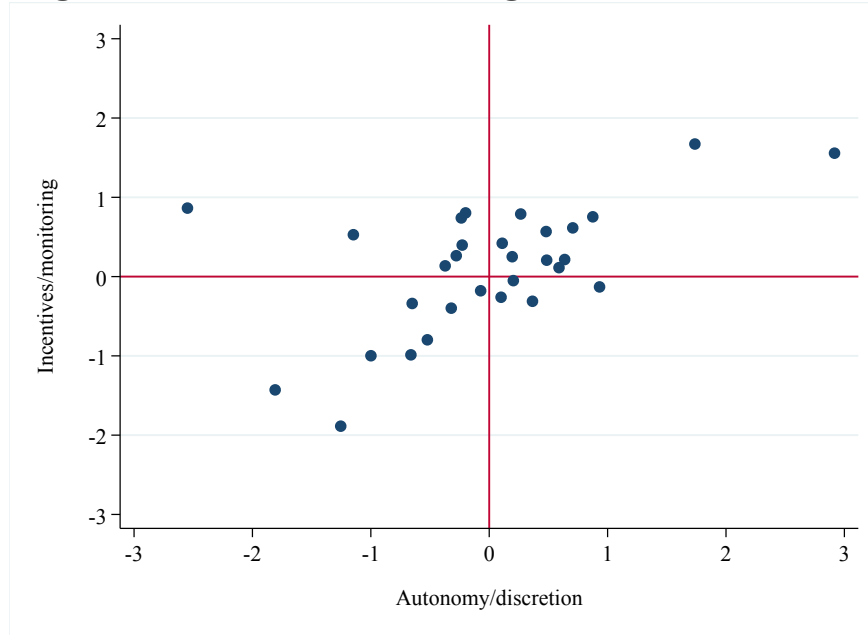


Figure 6 shows that, as with bureaucratic performance on task completion, there is substantial variation in the management practices bureaucrats are subject to across organizations along both of our sub-indices of management practices. Since many of the underlying practices are subject to administrative rules or guidelines aimed at producing uniform practices across organizations, the existence of this variation demonstrates that there is substantial *de facto* deviation from these *de jure* procedures with the Civil Service. To reiterate, this variation occurs despite the fact that all organizations share the same colonial and post-colonial structures, are governed by the same civil service laws and regulations, are overseen by the same supervising authorities, are assigned new hires from the same pool of potential workers, and are located proximately to each other in Accra. While a mainly qualitative literature has previously documented variation in state capacity and performance within states in developing countries, our findings represent large-scale evidence that such within-government variation in management and performance is systematic and does not consist merely of a handful of problem organizations or “pockets of effectiveness” [Tendler 1997; Leonard 2010; McDonnell 2017]. While such variation is often cast in a negative light in

bureaucratic settings, this variation can also derive from organizations going above and beyond formal procedures by implementing supporting informal practices, or by adapting formal rules to fit their circumstances.

Figure 6: Variation in Management Practices



Notes: Organisation z-scores presented for organisations with task completion data available.

A second important feature of organizational management practices illustrated by Figure 6 is that organizations' scores on the two practices are strongly positively correlated with each other ($\rho = .59$). Organizations that give their staff flexibility in day-to-day operations also use stronger monitoring and incentives, on average. While autonomy/discretion and incentives/monitoring as abstract approaches derive from different perspectives on public sector management, in practice organizations that score highly on one also score highly on the other (at least in the Ghanaian context).¹² This empirical finding provides much-needed nuance to the debate between top-down

¹²Comparing raw management scores, we find that the 75th percentile organization has an 8-9 percent higher management score than the 25th percentile organization across both sub-indices as well as in overall management scores. While this variation is smaller in numerical terms than we observe for task completion, quantitative comparison of these differences is difficult since management does not have a natural scale.

and bottom-up approaches to management discussed in Section 2. These are not alternative approaches in the sense that organizations face a binary either-or choice between them; rather, each organization combines these two approaches to one degree or another. Rather than selecting one style or the other, the pertinent question is thus in what proportions or ways they should be combined.

Finally, Figure 6 also illustrates a key feature of the data which enables our research design to answer this question: while there is a positive overall correlation between the autonomy/discretion and incentives/monitoring sub-indices, there is nonetheless considerable variation in the relative balance between them. That is, some organizations lean relatively more heavily towards the use of autonomy/discretion, while others lean relatively more heavily towards incentives/monitoring. In the next section, we explore how this variation in approaches to management is related to the variation we observe in task completion.

5 Empirical Method and Results

To study the relationship between task completion and management, we take as our unit of observation task i of type j in organization n . We estimate the following OLS specification,

$$y_{ijn} = \gamma_1 \text{Autonomy/Discretion}_n + \gamma_2 \text{Incentives/Monitoring}_n + \gamma_3 \text{Other}_n + \beta_1 PC_{ijn} + \beta_2 OC_n + \lambda_j + \epsilon_{ijn} \quad (1)$$

where y_{ijn} is a measure of task completion - either a binary indicator of whether the task is fully completed (Columns 1-5, our preferred measure) or a continuous measure of the task completion rate on the interval $[0,1]$ (Column 6, shown for robustness). Management practices are measured using the autonomy/discretion, incentives/monitoring and other indices, and PC_{ijn} and OC_n are task and organizational controls.¹³ As Figure 1B highlighted, many organizations implement the

¹³Task controls comprise task-level controls for whether the task is regularly implemented by the organization or a one off, whether the task is a bundle of interconnected tasks, and whether the division has to coordinate with actors external to government to implement the task. Organizational controls comprise a count of the number of interviews undertaken (which is a close approximation of the total number of employees) and organization-level controls for the share of the workforce with degrees, the share of the workforce with postgraduate qualifications,

same task type j , so we can control for task-type fixed effects λ_j in (1), as well as fixed effects for the broad sector the implementing organization operates in.¹⁴

The partial correlations of interest are γ_1 and γ_2 , the association between task completion and a one standard deviation change in management practices along the respective margins of autonomy and incentives/monitoring. If $\gamma_2 > \gamma_1$ this would be evidence in support of H1a, whereas if $\gamma_1 > \gamma_2$ it would be evidence in support of H1b. While our specification controls account for a wide array of potential confounding related to the organization, task type, and data collection, it is of course possible that there are other variables that are correlated with both management practices and task completion, so we emphasize that these estimates are partial correlations; we discuss interpretation and the potential for future causal research in the concluding section. To account for unobserved shocks, we cluster standard errors by organization n , the same level of variation as management practices.

Table 2 presents our main results.¹⁵ Before moving on to the main specification presented in Equation (1), Column 1 first shows that there is a positive relationship ($p = 0.034$) between the overall management z-score (which combines all three sub-indices) and task completion. This is evidence that higher overall management scores are indeed associated with higher task completion, even after controlling for an extensive range of organizational, task, and survey noise-related variables.

To illustrate the omitted variable bias that occurs when analyzing just one set of management practices in isolation, Column 2 then estimates the relationship between task and autonomy/discretion without controlling for organizations' other management practice, Column 3 does the same for incentives/monitoring, and Column 4 for our residual sub-index of other practices. The sub-indices are all positively associated with task completion, albeit with varying levels of

and the span of control. Following BVSr, we condition on 'noise' controls related to the management surveys. Noise controls are averages of indicators of the seniority, gender, and tenure of all respondents, the average time of day the interview was conducted and of the reliability of the information as coded by the interviewer.

¹⁴For the purpose of estimation, we aggregate some similar task types based on their primary classification, so that the task fixed effects included are: Advocacy and Policy Development, Financial and Budget Management, ICT Management and Research, Monitoring/Training/Personnel Management, Physical Infrastructure, Permits and Regulation, and Procurement. Sector fixed effects relate to whether the task is in the administration, environment, finance, infrastructure, security/diplomacy/justice or social sector.

¹⁵We do not report the R^2 of our regressions since this statistic does not have its usual interpretation under the linear probability model with binary dependent variables.

statistical significance and more strongly so for autonomy/discretion. Column 5 then presents our core specification from Equation 1, using binary task completion as the dependent variable, while Column 6 presents the same specification with an alternative continuous measure of task completion.

A consistent set of findings emerges across our main specifications in Columns 5-6: (i) management practices providing bureaucrats more autonomy and discretion are *positively* correlated with the likelihood of task completion ($\hat{\gamma}_1 > 0$); and (ii) management practices related to the provision of incentives or monitoring to bureaucrats are *negatively* correlated with the likelihood of task completion ($\hat{\gamma}_2 < 0$). These estimates imply that a one standard deviation increase in the autonomy/discretion sub-index is associated with an increase in the likelihood a task is fully completed by 11.7 percentage points, and a one standard deviation increase in incentives/monitoring is associated with a decrease in the likelihood a task is fully completed by 4.0 percentage points. This evidence is consistent with H1b, and thus is inconsistent with H1a.

Table 2: Management and Task Completion

Dependent Variable:	(1) Task completion [binary]	(2) Task completion [binary]	(3) Task completion [binary]	(4) Task completion [binary]	(5) Task completion [binary]	(6) Completion rate [0-1 continuous]
Management-Overall	0.059 (0.027) [0.034]					
Management-Autonomy/Discretion		0.080 (0.031) [0.013]			0.117 (0.066) [0.086]	0.052 (0.044) [0.243]
Management-Incentives/Monitoring			0.040 (0.034) [0.259]		-0.040 (0.062) [0.521]	-0.076 (0.040) [0.070]
Management-Other				0.047 (0.026) [0.073]	-0.012 (0.059) [0.834]	0.020 (0.038) [0.600]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)					0.152	0.084
Noise Controls	Yes	Yes	Yes	Yes	Yes	Yes
Organizational Controls	Yes	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	3620 (30)	3620 (30)	3620 (30)	3620 (30)	3620 (30)	3620 (30)

Notes: Standard errors are in parentheses with exact p-values in brackets, and are clustered by organization throughout. All columns report OLS estimates. P-values reported for Wald test of coefficient equality. See text for details of dependent variables, controls, and fixed effects. P-values reported for Wald test of coefficient equality. See text for details of controls and fixed effects. Figures are rounded to three decimal places.

These magnitudes are substantively important: recall the backdrop here is that only 34 percent of tasks are fully completed. The different point estimates between Columns 2-4 and Columns 5-6 illustrate that examining autonomy/discretion (incentives/monitoring) in isolation would lead to a downward (upward) bias on the point estimates, due to the underlying positive correlation between these measures. In terms of statistical significance, the difference between these two coefficients is more relevant than the individual difference from zero in terms of our theoretical question; a Wald test of coefficient equality varies in significance depending on the dependent variable. This is suggestive evidence that these two dimensions of management practice are differentially associated with task completion.

To examine how these suggestive findings vary with the *ex ante* and *ex post* clarity of the task (H2), Table 3 re-estimates Equation (1) on the sub-sample of tasks that fall below- and above-

median on each measure of task clarity. Columns 1 and 2 show that task completion's positive relationship with autonomy/discretion and negative relationship with incentives/monitoring is far stronger for tasks that have below-median *ex ante* clarity than for those that have above-median *ex ante* clarity. The difference between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ is statistically significant at less than the 1% level for tasks that are below-median on *ex ante* clarity, but there is no substantive or significant difference in coefficients for above-median *ex ante* clarity tasks. Similarly, for tasks with above-median *ex post* clarity (Columns 3-4), a one-standard deviation increase in autonomy/discretion (incentives/monitoring) is associated with a 33.9 percentage point increase (1 percentage point decrease) in task completion. Since 34 percent of tasks are fully completed, this implies that for tasks with high *ex post* clarity an increase of one standard deviation in autonomy/discretion is associated with a near-doubling of completion likelihood. For tasks that are below-median in terms of *ex post* clarity, autonomy/discretion is still positive and statistically significantly different from both zero and the coefficient on incentives/monitoring, but the difference is far smaller. This pattern of results is consistent with H2.

Table 3: Management Practices and Task Clarity

	<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity	
	(1) Below median	(2) Above median	(3) Below median	(4) Above median
Management-Autonomy/Discretion	0.207 (0.059) [0.002]	0.019 (0.107) [0.863]	0.126 (0.052) [0.021]	0.337 (0.077) [<0.001]
Management-Incentives/Monitoring	-0.138 (0.051) [0.011]	-0.020 (0.097) [0.834]	-0.023 (0.047) [0.623]	-0.012 (0.104) [0.907]
Management-Other	-0.007 (0.060) [0.907]	0.027 (0.082) [0.742]	-0.008 (0.050) [0.879]	-0.227 (0.077) [0.006]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.000	0.816	0.043	0.009
Noise and Organizational Controls	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	1851 (29)	1769 (29)	2177 (29)	1443 (27)

Notes: Standard errors are in parentheses with p-values in brackets, and are clustered by organization throughout. All columns report OLS estimates. The dependent variable in all columns is a dummy variable that takes the value 1 if the task is completed and 0 otherwise. The sample for Columns 1-2 and 3-4 is split according to whether the target clarity or actual achievement clarity is above median versus below or equal to the median. P-values reported for Wald test of coefficient equality. See text for details of controls and fixed effects. Figures are rounded to three decimal places.

While our operationalization of task clarity does not correspond precisely with Wilson's [1989] discussion of output and outcome observability, our findings are nonetheless consistent with the distinctions Wilson makes between procedural and craft agencies. Procedural agencies have tasks that are possible to specify *ex ante* and observe agents undertaking but are hard to measure *ex post*, and thus tend to be managed in rigid, rule-bound ways that give agents minimal discretion. Craft agencies have tasks that are difficult to specify or observe *ex ante* but whose results are relatively easy to observe *ex post*, so they tend to give their staff discretion with the knowledge that abuses of it will be relatively easy to detect. Although we focus on tasks rather than agencies as a whole, we provide empirical support for this broad idea: incentives/monitoring-heavy management

approaches are relatively more effective when task clarity is high *ex ante* and low *ex post*, and autonomy/discretion-heavy management approaches are relatively more effective when task clarity is low *ex ante* and high *ex post*.

The Online Appendix provides a battery of robustness checks on our estimates. Appendix Table A2 shows the results of Table 2 to be broadly robust to alternative samples, exclusion of outlier organizations, estimation methods, fixed effects specifications, and codings of completion rates. Appendix Table A3 does likewise for Table 3. Appendix Tables A4 and A5 further show the results to be robust to alternative clusterings of the standard errors, and Appendix Table A6 examines alternative codings of completion rates

One nuance in interpretation of this analysis revolves around our estimation of partial rather than absolute correlations between management practices and task completion. Recall that the coefficient on each index represents a partial correlation of that set of practices with task completion, conditional on the other practices being used in the organization (as well as our other controls and fixed effects). Thus, while we find that management practices related to incentives and monitoring are negatively related to task completion conditional on the level of autonomy and discretion being used in the same organization, this does not imply that all incentives and monitoring are bad for task completion. Rather, it implies that organizations seem to be overbalancing what we describe as their portfolio of management practices inefficiently towards incentives and monitoring at the expense of autonomy and discretion, particularly for tasks with low *ex ante* task clarity and high *ex post* clarity.

A second point relevant for the interpretation of these findings is that these estimates are based on our measurement of *de facto* management practices in an organization. While management in an organization can be described qualitatively both by the style of management (what type of practices the organization is trying to implement) and the quality of implementation of these practices, our measure of management combines both of these into a single dimension for each practice. We thus estimate the relationship between task completion and the management practices that organizations are *actually* using, rather than the management practices they are trying to use or an idealized version of these management practices. While this limits our study's ability to

speak to the potential efficacy of these practices when implemented “correctly”, the relationships between management practices and task completion under real-world implementation conditions is (in our view) a more important question.

6 Conclusion

Our study investigates the question of how two prominent approaches to public sector management - top-down control through monitoring and incentives, versus relying on bureaucratic professionalism through autonomy and discretion - are related to bureaucratic task completion. We find positive conditional associations between task completion and organizational practices related to autonomy and discretion, but negative conditional associations with management practices related to incentives and monitoring. This finding provides new evidence in support of the potential effectiveness of “bottom-up” approaches to management in the running debate about their relative merits vis-à-vis more “top-down”, carrot-and-stick approaches to public sector management [Miller 2000; Miller and Whitford 2016]. This is consistent with recent empirical evidence from a range of other contexts that finds positive effects from management strategies that promote bureaucratic autonomy and discretion [e.g. Honig 2018; Rasul and Rogger 2018; Bandiera *et al* 2019].

Consistent with our theoretical expectations, these relationships vary strongly with task clarity: the positive relationship between task completion and autonomy/incentives (relative to incentives/monitoring) is far stronger for tasks with low *ex ante* clarity and high *ex post* clarity. By refining and operationalizing Wilson’s [1989] classic output-outcome visibility typology in terms of bureaucratic task clarity, we provide empirical support for his key theoretical insight and for the intuition of many practitioners: that while top-down, control-oriented approaches to management may be appropriate for tasks for which it is clear in advance what needs to be done, for tasks that are difficult to fully specify in advance (but for which performance can be measured *ex post*) managers need to find ways to allow and support the exercise of bureaucratic discretion.

These findings are likely to be relevant to a range of other countries. Ranked just below the global median in terms of government effectiveness, empirical findings from Ghana are potentially

more relevant to the wide range of low- and middle-income countries that have been relatively understudied in public administration than findings from high-income countries (which represent the extreme end of the global distribution of wealth, human resource availability, and government effectiveness). For other low- and middle-income countries, our results push back against the widespread assumption that bureaucratic discretion should be minimized in such settings, and point to the ways in which bureaucratic professionalism can be an effective management strategy even in challenging contexts. If anything, we would expect this to be even more true in high-income contexts where such professional norms are even more strongly entrenched. But while the average effectiveness of top-down versus bottom-up management approaches might vary across contexts depending on these underlying conditions, we see an even wider scope of relevance for our theoretical argument about how the relative effectiveness of these approaches varies according to the *ex ante* and *ex post* clarity of the task. While further empirical studies replicating this analysis in high-income countries would be welcome, the fact that the three closest theoretical predecessors of our study [Wilson 1989; Romzek 2000; Chun and Rainey 2005] are all focused on high-income countries hints at the broad relevance of this mechanism.

Two key methodological features of our study are that we measure management across a *broad spectrum of management practices*, rather than just a specific practice or instance of a policy, and that we measure task completion across the *full range of bureaucratic activity* and across many organizations. Not only is this breadth valuable in itself and methodologically innovative, but we show that it matters for our results: since management practices are correlated with each other, estimating the impact of one management practice on task completion without controlling for others leads to significant omitted variable bias. Similarly, our measurement of *de facto* management practices across the full civil service allows us to analyze these practices *as they are actually implemented*, as opposed to what managers say they are doing or how these practices might work under closely controlled experimental conditions. While these features put together make our data unique, the underlying measurement methods are widely applicable and do not rely on idiosyncrasies of the Ghanaian context (although varying institutional features may require different analytical approaches). In particular, our task completion data was contained in simple quarterly

and annual reports that were designed and used for simple monitoring purposes, and similar reports are produced by public sector organizations worldwide. Our findings demonstrate that it is possible to clean and code such data to produce usable data for analysis. While care would need to be taken to validate such self-reported data in each case (as we did in Ghana), this type of administrative data at organizational level seems relatively under-exploited at present (with exceptions such as Ingraham *et al* [2003]).

While we have demonstrated the robustness of our findings against an extensive range of organizational, individual, and task characteristics to rule out many alternative explanations, and shown that many of the mechanism driving this result are consistent with theoretical predictions, it is nonetheless possible that there exist additional unobserved factors that (partially) explain the observed associations. In this sense, we view this study as an important complement to more narrowly focused (pseudo-)experimental studies [e.g. Banerjee *et al* 2014; Andersen and Moynihan 2016] as well as more nuanced qualitative research in advancing our knowledge about how management practices are related to task completion in the public sector. These findings could also serve as motivation and evidence base for future experimental interventions conducted in conjunction with governments to try to improve and evaluate the use of autonomy- and discretion-related management practices.

Finally, our study also builds on the literature examining cross-country differences in bureaucratic effectiveness by pushing forward the frontier in understanding *within-country* variation in effectiveness [Meier and O'Toole 2002; Nicholson-Crotty and O'Toole 2004; Owusu 2006; Andersen and Mortensen 2010; Leonard 2010; Gingerich 2013; Bersch *et al.* 2016; McDonnell 2017]. In its theoretical conceptualization of these issues, its innovative measurement methodology, and its striking empirical findings, we hope that this study can contribute to advancing our understanding of the causes and consequences of within-country variation in public sector management and effectiveness.

Data Availability Statement

The data underlying this article cannot be shared publicly due to the conditions for collecting it agreed with Ghana's Civil Service.

References

- [1] ANDERSEN.S.C. AND P.B.MORTENSEN (2010). "Policy Stability and Organizational Performance: Is There a Relationship?" *Journal of Public Administration Research and Theory* 20(1): 1-22, <https://doi.org/10.1093/jopart/mup005>.
- [2] ANDERSEN.S.C. AND D.P.MOYNIHAN. 2016. "Bureaucratic Investments in Expertise: Evidence from a Randomized Controlled Field Trial." *Journal of Politics* 78(4): 1032-44.
- [3] BANDIERA,O., M.C.BEST, A.Q.KHAN, AND A.PRAT. 2019. "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats." Mimeo, December 9.
- [4] BANERJEE.A, R.CHATTOPADHYAY, E.DUFLO, D.KENISTON AND N.SINGH (2014) Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy and Training, NBER Working Paper 17912, available at: <https://www.nber.org/papers/w17912>.
- [5] BERSCH.K, S.PRAÇA, AND M.TAYLOR. 2016. "State Capacity, Bureaucratic Politicization, and Corruption in the Brazilian State." *Governance* 30(1): 105-124.
- [6] BEVAN.G AND C.HOOD (2006). "What's Measured is What Matters: Targets and Gaming in the English Health Care System." *Public Administration* 84(3): 517-538.
- [7] BLOOM.N AND J.VAN REENEN (2007) "Measuring and Explaining Management Practices Across Firms and Countries," *Quarterly Journal of Economics* 122: 1351-408.
- [8] BLOOM.N AND J.VAN REENEN (2010b). 'New Approaches to Surveying Organizations'. *American Economic Review Papers and Proceedings* 100: 105-109.

- [9] BLOOM.N, R.SADUN AND J.VAN REENEN (2012). “The Organization of Firms Across Countries,” *Quarterly Journal of Economics* 127: 1663-705.
- [10] BLOOM.N, R.LEMOS, R.SADUN, D.SCUR, AND J.VAN REENEN (2014). ‘The New Empirical Economics of Management.’ NBER Working Paper 20102, May.
- [11] BOYNE.G AND C.HOOD (2010). ‘Incentives: New Research on an Old Problem’ *Journal of Public Administration Research and Theory* 20: i177-i180, <https://doi.org/10.1093/jopart/muq023>.
- [12] CARPENTER.D (2001). *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862-1928*. Princeton: Princeton University Press.
- [13] CHUN, YOUNG HAN, AND HAL G. RAINEY. 2005. “Goal Ambiguity and Organizational Performance in U.S. Federal Agencies”. *Journal of Public Administration Research and Theory* 15: 529-557, <https://doi.org/10.1093/jopart/mui030>.
- [14] DAHLSTROM.C AND V.LAPUENTE (2010). “Explaining Cross-Country Differences in Performance-Related Pay in the Public Sector.” *Journal of Public Administration Research and Theory* 20(3): 577-600, <https://doi.org/10.1093/jopart/mup021>.
- [15] DIXIT.A (2002) “Incentives and Organizations in the Public Sector: An Interpretive Review,” *Journal of Human Resources* 37: 696-727.
- [16] DUFLO, ESTHER, REMA HANNA, AND STEPHEN P. RYAN. 2012. “Incentives work: Getting teachers to come to school.” *American Economic Review* 102(4): 1241–78.
- [17] EINSTEIN, KATHERINE LEVINE, AND DAVID M. GLICK. 2017. “Does Race Affect Access to Government Services? An Experiment Exploring Street-Level Bureaucrats and Access to Public Housing.” *American Journal of Political Science* 61(1): 100-116.
- [18] FINAN.F, B.A.OLKEN, AND R.PANDE (2017) “The Personnel Economics of the State,” in A.V.Banerjee and E.(eds.) *Handbook of Field Experiments*, Elsevier.

- [19] FINER.H (1941). "Administrative Responsibility in Democratic Government." *Public Administration Review* 1(4): 335-350.
- [20] FRIEDRICH.C (1940) "Public Policy and the Nature of Administrative Responsibility." *Public Policy* 1: 3-24. Reprinted in Francis Rourke, ed. *Bureaucratic Power in National Politics*, 3rd ed. Boston: Little, Brown. 1978.
- [21] GINGERICH, DANIEL. 2013. "Governance Indicators and the Level of Analysis Problem: Empirical Findings from South America." *British Journal of Political Science* 43 (3): 505-540.
- [22] HASNAIN.Z, N.MANNING AND J.H.PIERSKALLA (2012) Performance-related Pay in the Public Sector, *World Bank Policy Research Working Paper* 6043, available at: <http://documents.worldbank.org/curated/en/666871468176639302/Performance-related-pay-in-the-public-sector-a-review-of-theory-and-evidence> .
- [23] HEINRICH.C. 1999. "Do Government Bureaucrats Make Effective Use of Performance Management Information?" *Journal of Public Administration Research and Theory* 9(3): 363-393, <https://doi.org/10.1093/oxfordjournals.jpart.a024415>.
- [24] HONIG.D. 2018. *Navigation by Judgment: Why and When Top Down Management of Foreign Aid Doesn't Work*. Oxford: Oxford University Press.
- [25] INGRAHAM.P.W., P.G.JOYCE, AND A.KNEEDLER DONAHUE (2003). *Government Performance: Why Management Matters*. London: The Johns Hopkins University Press.
- [26] KAUFMANN.D, A.KRAAY, AND M.MASTRUZZI. 2010. "The Worldwide Governance Indicators Methodology and Analytical Issues." *World Bank Policy Research Working Paper* 5430, available at: <https://openknowledge.worldbank.org/handle/10986/3913> .
- [27] KELMAN.S AND J.FRIEDMAN (2009). "Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service." *Journal of Public Administration Research and Theory* 19: 917-946, <https://doi.org/10.1093/jopart/mun028>.

- [28] LEONARD.D.K (2010) “Pockets’ of Effective Agencies in Weak Governance states: Where are they Likely and Why does it Matter?,” *Public Administration and Development* 30: 91-101.
- [29] LYNN, LAURENCE, CAROLYN HEINRICH, AND CAROLYN HILL. 2000. “Studying Governance and Public Management: Challenges and Prospects.” *Journal of Public Administration Research and Theory* 10(2): 233-261, <https://doi.org/10.1093/oxfordjournals.jpart.a024269>.
- [30] MCDONNELL.E (2017) “Patchwork Leviathan: How Pockets of Bureaucratic Governance Flourish within Institutionally Diverse Developing States.” *American Sociological Review* 82(3): 476-510.
- [31] MEIER, KENNETH J., AND LAURENCE J. O’TOOLE JR. 2002. “Public Management and Organizational Performance: The Effect of Managerial Quality.” *Journal of Policy Analysis and Management* 21(4): 629-643.
- [32] MEIER, KENNETH J., AND LAURENCE J. O’TOOLE JR. 2012. “Subjective Organizational Performance and Measurement Error: Common Source Bias and Spurious Relationships.” *Journal of Public Administration Research and Theory* 23: 429-456, <https://doi.org/10.1093/jopart/mus057>.
- [33] MILLER.G (2000). “Above Politics: Credible Commitment and Efficiency in the Design of Public Agencies.” *Journal of Public Administration Research and Theory* 10(2): 289-327, <https://doi.org/10.1093/oxfordjournals.jpart.a024271>.
- [34] MILLER, GARY, AND ANDREW B. WHITFORD (2007). “The Principal’s Moral Hazard: Constraints on the Use of Incentives in Hierarchy.” *Journal of Public Administration Research and Theory* 17(2): 213-233, <https://doi.org/10.1093/jopart/mul004>.
- [35] MILLER, GARY, AND ANDREW B. WHITFORD. 2016. *Above Politics: Bureaucratic Discretion and Credible Commitment*. Cambridge: Cambridge University Press.
- [36] MOE, TERRY. 2013. “Delegation, Control, and the Study of Public Bureaucracy.” In Gibbons, Robert, and John Roberts (eds) *The Handbook of Organizational Economics*, Oxford: Princeton University Press, 1148-1181.

- [37] MOYNIHAN, D.P., S.FERNANDEZ, S.KIM, K.M.LEROUX, S.J.PIOTROWSKI, B.E.WRIGHT, AND K.YANG. 2011. "Performance Regimes Amidst Governance Complexity." *Journal of Public Administration Research and Theory* 21: i141-i155, <https://doi.org/10.1093/jopart/muq059>.
- [38] NICHOLSON-CROTTY.S AND L.O'TOOLE (2004). "Public Management and Organizational Performance: The Case of Law Enforcement Agencies." *Journal of Public Administration Research and Theory* 14(1): 1-18, <https://doi.org/10.1093/jopart/muh001>.
- [39] OLKEN, BENJAMIN, AND ROHINI PANDE. 2012. "Corruption in Developing Countries." *Annual Review of Economics* 2012(4): 479-509.
- [40] OWUSU, FRANCIS. 2006. "Differences in the Performance of Public Organisations in Ghana: Implications for Public-Sector Reform Policy." *Development Policy Review* 24(6): 693-705.
- [41] PERRY.J, T.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.
- [42] RASUL.I AND D.ROGGER (2018) "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," *Economic Journal* 128: 413-46.
- [43] ROMZEK.B. 2000. "Dynamics of public sector accountability in an era of reform." *International Review of Administrative Sciences* 66: 21-44.
- [44] ROSE-ACKERMAN.S (1986) "Reforming Public Bureaucracy Through Economic Incentives?," *Journal of Law, Economics and Organization* 2: 131-61.
- [45] SIMON.W (1983) "Legality, Bureaucracy, and Class in the Welfare System," *Yale Law Journal* 92: 1198-269.
- [46] TALBOT.C (2010) *Theories of Performance: Organizational and Service Improvement in the Public Domain*. Oxford: Oxford University Press.
- [47] TENDLER, JUDITH. 1997. *Good government in the tropics*. Baltimore: Johns Hopkins University Press.

- [48] THOMANN.E, N.VAN ENGEN, AND L.TUMMERS (2018). “The Necessity of Discretion: A Behavioral Evaluation of Bottom-Up Implementation Theory.” *Journal of Public Administration Research and Theory* 28(4): 583-601, <https://doi.org/10.1093/jopart/muy024>.
- [49] WILSON.J (1989) *Bureaucracy: What Government Agencies Do and Why They Do It*, New York: Basic Books.

A Online Appendix

A.1 Measuring Management Practices

Table A1 below presents the practice groupings, topics, indicative questions, and benchmark management scores for the interview items used to construct the BSVR-style management scores.

Table A1: Defining Management Practices

Management Practice	Topic	Indicative Question	Score 1	Score 3	Score 5
Autonomy/Discretion	Roles	Can most senior staff in your division make substantive contributions to the policy formulation and implementation process?	Senior staff do not have channels to make substantive contributions to organisational policies, nor to the management of their implementation.	Substantive contributions can be made in staff meetings by all senior staff but there are no individual channels for ideas to flow up the organisation.	It is integral to the organisation's culture that any member of senior staff can substantively contribute to the policies of the organisation or their implementation.
		When senior staff in your division are given tasks in their daily work, how much discretion do they have to carry out their assignments? Can you give me an example?	Officers in this division have no real independence to make decisions over how they carry out their daily assignments. Their activities are defined in detail by senior colleagues or organisational guidelines.	Officers in this division have some independence as to how they work, but strong guidance from senior colleagues, or from rules and regulations.	Officers in this division have a lot of independence as to how they go about their daily duties.
Flexibility		Does your division make efforts to adjust to the specific needs and peculiarities of communities, clients, or other stakeholders?	The division uses the same procedures no matter what. In the face of specific needs or community/client peculiarities, it does not try to develop a 'better fit' but automatically uses the default procedures.	The division makes steps towards responding to specific needs and peculiarities, but stumbles if the specific needs are complex. Often, tailoring of services is often unsuccessful.	The division always redefines its procedures to respond to the needs of communities/clients. It does its best to serve each individual need as best as it can.
		How flexible would you say your division is in terms of responding to new and improved work practices?	There is no effort to incorporate new ideas or practices. When practice improvements do happen, there is no effort to disseminate them through the division.	New ideas or practices are sometimes adopted but in an ad hoc way. These are sometimes shared informally or in a limited way, but the division does not actively encourage this or monitor their adoption.	Seeking out and adopting improved work practices is an integral part of the division's work. Improvements are systematically disseminated throughout the division and their adoption is monitored.

Table A1 Continued: Defining Management Practices

Management Practice	Topic	Indicative Question	Score 1	Score 3	Score 5
Incentives/Monitoring	Performance Incentives	Given past experience, how would under-performance be tolerated in your division?	Poor performance is not addressed or is inconsistently addressed. Poor performers rarely suffer consequences or are removed from their positions.	Poor performance is addressed, but on an ad hoc basis. Use of intermediate interventions, such as training, is inconsistent. Poor performers are sometimes removed from their positions under conditions of repeated poor performance.	Repeated poor performance is systematically addressed, beginning with targeted intermediate interventions. Persistently poor performers are moved to less critical roles or out of the organisation.
		Given past experience, are members of [respondent's organisation] disciplined for breaking the rules of the civil service?	Breaking the rules of the civil service does not carry any consequences in this division. Guilty parties do not receive the stipulated punishment.	An officer may break the rules infrequently and not be punished. An officer who regularly breaks the rules may be disciplined, but there would be no other specific actions beyond this. The underlying drivers of the behaviour can persist indefinitely.	Any officer who breaks the rules of the civil service is punished; the underlying driver is identified and rectified. On-going efforts are made to ensure the issue does not arise again.
Monitoring	Monitoring	Does your division use performance, targets, or indicators for tracking and rewarding (financially or non-financially) the performance of its officers?	Officers in the division are rewarded (or not rewarded) in the same way irrespective of their performance.	The evaluation system awards good performance in principle (financially or non-financially), but awards are not based on clear criteria/processes.	The evaluation system rewards individuals (financially or non-financially) based on performance. Rewards are given as a consequence of well-defined and monitored individual achievements.
		In what kind of ways does your division track how well it is delivering services? Can you give me an example?	Measures tracked are not appropriate or do not indicate directly if overall objectives are being met. Tracking is an ad hoc process and most processes aren't tracked at all. Tracking is dominated by the head of the division.	Performance indicators have been specified but may not be relevant to the division's objectives. The division has inclusive staff meetings where staff discuss how they are doing as division.	Performance is continuously tracked, both formally with key performance indicators and informally, using appropriate indicators and including many of the divisional staff.

Table A1 Continued: Defining Management Practices

Management Practice	Topic	Indicative Question	Score 1	Score 3	Score 5
Other	Staffing	Do you think about attracting talented people to your division and then doing your best to keep them? For example, by ensuring they are happy and engaged with their work.	Attracting, retaining and developing talent throughout the division is not a priority or is not possible given service rules.	Having top talent throughout the division is seen to be a key way to effectively deliver on the organisations mandate but there is no strategy to identify, attract or train such talent.	The division actively identifies and acts to attract talented people who will enrich the division. They then develop those individuals for the benefit of the division and try to retain their services.
		If two senior level staff joined your division five years ago and one was much better at their work than the other, would he/she be promoted through the service faster?	The division promotes people by tenure only, and thus performance does not play a role in promotion.	There is some scope for high performers to move up through the service faster than non-performers in this division, but the process is gradual and vulnerable to inefficiencies.	The division would certainly promote the high-performer faster, and would rapidly move them to a senior position to capitalise on their skills.
		Is the burden of achieving your division's targets evenly distributed across its different officers, or do some individuals consistently shoulder a greater burden than others?	A small minority of staff undertake the vast majority of substantive work within the division.	A majority of staff make valuable inputs, but it is by no means everyone who pulls their weight.	Each member of the division provides an equally valuable contribution, working where they can provide their highest value.
		Would you say that senior staff try to use the right staff for the right job?	Often tasks are not staffed by the appropriate staff. Staff are allocated to tasks either randomly, or for reasons that are not associated with productivity.	Most jobs have the right staff on them, but there are organisational constraints that limit the extent to which effective matching happens.	The right staff are always used for a task.
Targeting	Targeting	Does your division have a clear set of targets derived from the organization's goals and objectives? Are they used to determine your work schedule?	The division's targets are very loosely defined or not defined at all; if they exist, they are rarely used to determine our work schedule and our activities are based on ad hoc directives from senior management.	Targets are defined for the division and its individual officers (managers and staff). However, their use is relatively ad hoc and many of the division's activities do not relate to those targets.	Targets are defined for the division and individuals (managers and staff) and they provide a clear guide to the division and its staff as to what the division should do. They are frequently discussed and used to benchmark performance.
		When you arrive at work each day, do you and your colleagues know what their individual roles and responsibilities are in achieving the organisation's goals?	No. There is a general level of confusion as to what the organisation is trying to achieve on a daily basis and what individual's roles are towards those goals.	To some extent, or at least on some days. The organisation's main goals and individual's roles to achieve them are relatively clear, but it is sometimes difficult to see how current activities are moving us towards those.	Yes. It is always clear to the body of staff what the organisation is aiming to achieve with the days activities and what individual's roles and responsibilities are towards that.

A.2 Measuring Task Completion

In Ghana each civil service organization is required to provide quarterly and annual progress reports. These detail targets and achievements for individual tasks. The process of measuring completion for each organization then comprised two steps. First, extracting the data from organizations' reports (which differed slightly in their formats) into a standardized template. Second, coding variables based on the standardized data.

Figure A1 shows a snapshot of a typical quarterly progress report. The unit of observation is the task, defined as the most disaggregated activity reported. For each quarterly progress report, we codified task line items using a team of trained research assistants and a team of civil servant officers seconded from the Management Services Department in the Civil Service. Each task was thus assigned to an organization (ministry or department) and to a division within that organization.

Figure A1: Quarterly Report, an Example

MINISTRY OF TRADE AND INDUSTRY – DRAFT 2015, SECOND QUARTER PROGRESS REPORT

PROGRAMMED ACTIVITIES	BUDGET PERFORMANCE	EXPECTED OUTPUT/IMPACT	ACTUAL OUTPUT	REMARKS
MANAGEMENT AND ADMINISTRATION				
ADMINISTRATION				
LEGAL DIVISION				
Output 1: Competition Policy Developed and Approved				
1. Develop Draft Competition Policy		Draft Competition Policy developed	Evaluation of technical proposals completed. Two firms have been shortlisted. The two firms have been invited to present financial proposals.	Procurement process is still on-going and expected to be completed by end of Quarter 3.
2. Organize stakeholder meeting to validate draft Competition Policy		Stakeholder meeting held	Yet to be done	
3. Prepare Cabinet Memo and submit Competition Policy to Cabinet for approval		Competition Policy submitted to Cabinet	Work on the bill is still frozen due to on-going work on the policy	
Output 2: Competition Bill prepared and approved				
1. Review the Draft Competition Bill taking into		The draft bill reviewed	Yet to be done	Work on the bill will begin after the Policy has been

Notes: Key information used in coding is highlighted; Appendix A provides details on all output data variables.

A.3 Extracting and Standardizing

Although organizations' reports differed in their format and variable coverage, we extracted the following standard variables for each organization (leaving them blank where the variable was

missing).

Task Level 1 The name or short description of the task specifying the action to be taken during the time period, at the most disaggregated or fine-grained level available. For instance, in Figure A1, this is ‘Develop draft competition policy’. This variable defines the unit of observation, and by definition, cannot be missing.

Task Level 2 The name or short description of the task, aggregated to one level higher than in *Task level 1*. Many organizations reported tasks that were nested into broader outputs, or whose completion required multiple sequential or simultaneous smaller tasks to be completed. For example, in Figure A1 the *Task level 2* for ‘Develop draft competition policy’ is ‘Competition Policy Developed and Approved.’ Multiple tasks can thus share the same *Task level 2*.

Task Level 3 The same as Task level 2, but one level of aggregation higher. As in Figure A1, this level of aggregation was frequently unreported, but was extracted where relevant.

Budget Allocation/Cost The budgeted cost of the task. This was reported infrequently.

Baseline Completion Level Where reported, the level of attainment on the task at the start of the time period.

Actual Achievement The actual attainment or work done during the time period. Together with the target level of achievement for the time period (from Task level 1) and (where relevant) the baseline level of completion, this is used to code task completion (as described in more detail below).

Remarks Where reported, the organization’s comments about the task. These often explain why the target level of attainment was not achieved during the time period.

A.4 Coding

After extracting the data, our team of civil servants and research assistants coded a fixed list of variables for each task (at the most disaggregated level, *Task level 1*). As the variables to be coded required coders to interpret and judge the information being reported by each organization, coding was undertaken by two independent coders, with reconciliation led by managers where necessary. Below is a list of all variables coded for each task.

Task Type (primary) *Which category best describes this task?* Coders had to select one of the following: (i) Advocacy, outreach and stakeholder engagement/relations; (ii) Financial & budget management; (iii) ICT management and/or development; (iv) Monitoring, review, & audit; (v) Permits and regulation; (vi) Personnel management; (vii) Physical infrastructure – office & facilities; (viii) Physical infrastructure – public infrastructure and projects; (ix) Policy development; (x) Procurement; (xi) Research; (xii) Training.

Task Type (secondary) *If task covers more than one category, select the secondary category here.* Coders had to select one of the same twelve categories as above.

Period/Regular vs. One-off *Is the task repeated (e.g. weekly, quarterly, annually) or one-off (no planned repetition)?* Coders had to select one of: (i) Periodic/ regular (e.g. weekly, quarterly, annually); (ii) One-off (no planned repetition).

Task Scope *How narrowly is the task defined? Does it include multiple tasks, or even multiple broader outputs?* Coders had to select one of: (i) Single activity (one step in a larger activity, has no value on its own; e.g. hold a meeting about writing a policy); (ii) Single task (multiple steps, has value on its own; e.g. write a policy); (iii) Bundle of tasks (multiple tasks that each have their own value; e.g. write four policies)].

Technical Complexity *Does the task require specific technical or scientific knowledge, beyond the level most civil servants would have?* Coders had to select one of: (i) No technical knowledge required (any senior civil servant could do this); (ii) Technical knowledge is required (special education or training needed).

Coordination Required *Does the division have to coordinate or interact with other actors in order to achieve the task?* Coders could select the most applicable option of the following: (i) Requires action from other divisions in the organization; (ii) Requires action from other government organizations; (iii) Requires action from stakeholders outside government.

Ex Ante Target Clarity *How precise, specific, and measurable is the target?* Coders had to answer on a 1-5 scale (where integers and half values were both permitted) using the following scoring guidelines. Score 1: Target is undefined or so vague it is impossible to assess what completion would mean; Score 3: Target is defined, but with some ambiguity; Score 5: There is no

ambiguity over the target – it is precisely quantified or described.

Ex Post Actual Achievement Clarity *How precise, specific, and measurable is what the division actually achieved?* Coders had to answer on a 1-5 scale (where integers and half values were permitted) using the following scoring guidelines. Score 1: Task information is absent or so vague it is impossible to assess completion; Score 3: Task information is given but there is some ambiguity over whether the target was met; Score 5: Task information is clear and unambiguous.

Completion Status *How did actual achievement compare to the target?* Coders had to answer on a 1-5 scale (where integers and half values were permitted) using the following scoring guidelines. Score 1: No action was taken towards achieving the target; Score 3: Some substantive progress was made towards achieving the target. The task is partially complete and/or important intermediate steps have been completed; Score 5: The target for the task has been reached or surpassed.

Completion Remarks *Were any challenges/ obstacles mentioned? Coders could select all that applied from the following:* (i) awaiting action from another division, organization or stakeholder; (ii) 2 = Procurement/sourcing delay or problem; (iii) Sequencing issue (can't start until another task has been completed); (iv) Lack of technical knowledge to complete activity; (v) Delayed/non-release of funds; (vi) Unexpected event; (vii) Activity not due.

There are at least two coders per task. Inter-coder reliability for the completion scores (Cohen's Kappa) was 0.348. Given the tendency for averaging scores to reduce the measured variation, we use the maximum and minimum scores to code whether tasks are fully complete/never initiated respectively. Our main result is robust to alternative methods by which to combine codings.

A.5 Robustness

Appendix Table A2 provides a battery of checks on Table 2's estimates of the relationship between management practices and task completion. These show the basic results to be robust to a range of samples, estimation methods, fixed effects specifications, and alternative codings of completion rates. Column 1 reproduces the main specification from Table 2, Column 5, for reference. Column 2 excludes tasks implemented by the largest organization in terms of number of tasks. Column 3

excludes the five smallest organizations by number of tasks. Columns 4 and 5 exclude organizations at the top and bottom of the autonomy/discretion and incentives/monitoring management scales respectively. Column 6 reports the result of estimation a specification analogous to Equation (1) but using a fractional regression to account for the fact that task completion rates lie between zero and one. In Column 7 we control for task-sector level fixed effects (so allowing for sector specific impacts of task types). Column 8 takes a binary measure of whether a project was initiated at all (i.e. completion status greater than 1 on the 1-5 scale).

The results are qualitatively similar throughout in terms of the signs on autonomy/discretion and incentives/monitoring, and indeed in most specifications the difference between the coefficients on autonomy/discretion and incentives/monitoring is actually more statistically significant than in our main specification. In Column 8 the coefficient on autonomy/discretion is slightly negative (but statistically insignificant), but is still less negative than for incentives/monitoring. The difference in results here may be partly explained by the relatively low variation at organizational level in the proportion of tasks not initiated (see Figure 2). Overall, however, these robustness checks provide evidence that our main results are not driven by outliers or modelling choices.

Appendix Table A3 re-conducts the task clarity analysis of Table 3 across the range of alternative specifications laid out in Table A2. Since the task clarity analysis consists of a set of four regressions (on the samples of below- and above-median values of *ex ante* clarity and *ex post* clarity tasks), Appendix Table A3 takes the form of eight sets of four regressions. The specification of Set 1 in Table A3 corresponds to the specification of Column 1 in Table A2, and so on. Again, the results are fairly consistent across specifications, with the pattern identified in our main analysis emerging across a range of specifications and samples.

Table A2: Robustness - Management and Task Completion

	(1) Core	(2) Excl. Org. With Most Tasks	(3) Excl. Five Orgs. With Smallest No. of Tasks	(4) Excl. Autonomy/ Discretion Outliers	(5) Excl. Incentives/ Monitoring Outliers	(6) Fractional Regression	(7) Alternative Fixed Effects	(8) Project Initiation
Management-Autonomy/Discretion	0.117 (0.066) [0.086]	0.145 (0.053) [0.010]	0.271 (0.066) [<0.001]	0.332 (0.124) [0.013]	0.124 (0.068) [0.081]	0.533 (0.304) [0.080]	0.090 (0.068) [0.195]	-0.003 (0.080) [0.970]
Management-Incentives/Monitoring	-0.040 (0.061) [0.521]	-0.083 (0.062) [0.192]	-0.041 (0.058) [0.483]	-0.061 (0.060) [0.324]	-0.072 (0.076) [0.349]	-0.177 (0.287) [0.537]	-0.035 (0.053) [0.511]	-0.017 (0.068) [0.016]
Management-Other	-0.012 (0.059) [0.834]	-0.150 (0.058) [0.015]	-0.117 (0.046) [0.018]	-0.144 (0.071) [0.053]	-0.004 (0.062) [0.946]	-0.080 (0.274) [0.772]	-0.002 (0.054) [0.972]	0.067 (0.070) [0.341]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.152	0.020	0.009	0.029	0.128	0.156	0.232	0.171
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type Within Sector	Task Type, Sector
Observations (clusters)	3620 (30)	3125 (29)	3593 (26)	3379 (27)	3585 (29)	3620 (30)	3620 (30)	3620 (30)

Notes: Standard errors are in parentheses with p-values in brackets, and are clustered by organization throughout. Columns 1-7 report OLS estimates. In Columns 1 through 7, the dependent variable is a dummy variable that takes the value 1 if the Task is fully completed and 0 otherwise. Column 2 excludes tasks implemented by the largest organization in terms of number of Tasks. Column 3 removes the 5 smallest organizations by number of tasks. Columns 4 and 5 exclude organizations at the top and bottom of the Autonomy/Discretion and Incentives/Monitoring management scales respectively. Column 6 estimates a specification analogous to Column 1 but using a fractional regression to account for the fact that task completion rates lie between zero and one. Column 7 controls for task-sector level fixed effects. Column 8 uses project initiation as a dependent variable. P-values reported for Wald test of coefficient equality. Task Type fixed effects relate to whether the primary classification of the Task is 'Advocacy and Policy Development', 'Financial & Budget Management', 'ICT Management and Research', 'Monitoring, Training and Personnel Management', 'Physical infrastructure', 'Permits and Regulation' or 'Procurement'. Sector fixed effects relate to whether the task is in the administration, environment, finance, infrastructure, security/diplomacy/justice or social sector. Task controls comprise task-level controls for whether the task is regularly implemented by the organization or a one off, whether the task is a bundle of interconnected tasks, and whether the division has to coordinate with actors external to government to implement the task. Organizational controls comprise a count of the number of interviews undertaken, which is a close approximation of the total number of employees, and organization-level controls for the share of the workforce with degrees, the share of the workforce with postgraduate qualifications, and the span of control. Noise controls are averages of indicators of the seniority, gender, and tenure of all respondents, the average time of day the interview was conducted and of the reliability of the information as coded by the interviewer. Figures are rounded to three decimal places.

Table A3: Robustness - Management Practices and Task Clarity (Part 1/2)

Panel (a)								
	Set 1: Core specification				Set 2: Excl. Org. With Most Tasks			
	<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity		<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity	
	(1) Below median	(2) Above median	(3) Below median	(4) Above median	(1) Below median	(2) Above median	(3) Below median	(4) Above median
Management-Autonomy/Discretion	0.208 (0.059) [0.002]	0.019 (0.107) [0.863]	0.126 (0.052) [0.021]	0.337 (0.077) [<0.001]	0.223 (0.049) [<0.001]	0.068 (0.037) [0.076]	0.116 (0.035) [0.002]	0.424 (0.058) [<0.001]
Management-Incentives/Monitoring	-0.138 (0.051) [0.011]	-0.020 (0.097) [0.834]	-0.023 (0.047) [0.623]	-0.012 (0.104) [0.907]	-0.185 (0.057) [0.003]	-0.131 (0.039) [0.003]	-0.076 (0.035) [0.037]	-0.171 (0.052) [0.003]
Management-Other	-0.007 (0.060) [0.907]	0.027 (0.082) [0.742]	-0.008 (0.050) [0.879]	-0.227 (0.077) [0.006]	-0.100 (0.048) [0.045]	-0.293 (0.063) [<0.001]	-0.100 (0.028) [0.002]	-0.519 (0.066) [<0.001]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.000	0.816	0.043	0.009	0.000	0.001	0.001	0.000
Noise and Organizational Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	1851 (29)	1769 (29)	2177 (29)	1443 (27)	1624 (28)	1501 (28)	1889 (28)	1236 (26)
Panel (b)								
	Set 3: Excl. 5 Orgs. With Smallest No. of Tasks				Set 4: Excl. Autonomy/ Discretion Outliers			
	<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity		<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity	
	(1) Below median	(2) Above median	(3) Below median	(4) Above median	(1) Below median	(2) Above median	(3) Below median	(4) Above median
Management-Autonomy/Discretion	0.311 (0.035) [<0.001]	0.336 (0.151) [0.036]	0.204 (0.049) [<0.001]	0.461 (0.078) [<0.001]	0.208 (0.124) [0.106]	0.259 (0.126) [0.051]	0.121 (0.059) [0.052]	0.654 (0.117) [<0.001]
Management-Incentives/Monitoring	-0.100 (0.048) [0.050]	0.022 (0.089) [0.807]	-0.033 (0.054) [0.553]	0.071 (0.101) [0.487]	-0.106 (0.063) [0.103]	-0.038 (0.069) [0.589]	0.001 (0.041) [0.975]	0.021 (0.078) [0.790]
Management-Other	-0.087 (0.047) [0.075]	-0.183 (0.100) [0.080]	-0.072 (0.044) [0.116]	-0.368 (0.072) [<0.001]	-0.025 (0.068) [0.718]	-0.144 (0.104) [0.179]	-0.024 (0.046) [0.603]	-0.478 (0.105) [<0.001]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.000	0.070	0.007	0.004	0.086	0.037	0.092	0.000
Noise and Organizational Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	1838 (26)	1755 (25)	2157 (26)	1436 (24)	1792 (27)	1635 (27)	2069 (27)	1358 (2)

Notes: Standard errors are in parentheses with p-values in brackets, and are clustered by organization throughout. All columns report OLS estimates. The dependent variable in all columns is a dummy variable that takes the value 1 if the task is completed and 0 otherwise. For each set of robustness checks, the sample for Columns 1-2 and 3-4 is split according to whether the target clarity or actual achievement clarity is above median versus below or equal to the median. In sets 1 through 7, the dependent variable is a dummy variable that takes the value 1 if the Task is fully completed and 0 otherwise. Set 2 excludes tasks implemented by the largest organization in terms of number of Tasks. Set 3 removes the 5 smallest organizations by number of tasks. Sets 4 and 5 exclude organizations at the top and bottom of the Autonomy/Discretion and Incentives/Monitoring management scales respectively. Set 6 estimates a specification analogous to Set 1 but using a fractional regression to account for the fact that task completion rates lie between zero and one. Set 7 controls for task-sector level fixed effects. Set 8 uses project initiation as a dependent variable. P-values reported for Wald test of coefficient equality. See text for details of controls and fixed effects. Figures are rounded to three decimal places.

Table A3: Robustness - Management Practices and Task Clarity (Part 2/2)

Panel (c)								
	Set 5: Excl. Incentives/ Monitoring Outliers				Set 6: Fractional Regression			
	<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity		<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity	
	(1) Below median	(2) Above median	(3) Below median	(4) Above median	(1) Below median	(2) Above median	(3) Below median	(4) Above median
Management-Autonomy/Discretion	0.207 (0.060) [0.002]	0.131 (0.108) [0.236]	0.156 (0.069) [0.033]	0.525 (0.074) [<0.001]	0.903 (0.296) [0.002]	0.104 (0.505) [0.837]	0.781 (0.377) [0.038]	1.551 (0.375) [<0.001]
Management-Incentives/Monitoring	-0.110 (0.063) [0.091]	-0.008 (0.104) [0.938]	-0.092 (0.059) [0.132]	0.087 (0.088) [0.331]	-0.608 (0.246) [0.013]	-0.127 (0.449) [0.777]	-0.130 (0.306) [0.672]	-0.037 (0.445) [0.934]
Management-Other	-0.023 (0.065) [0.732]	0.048 (0.072) [0.511]	-0.003 (0.055) [0.953]	-0.178 (0.038) [<0.001]	-0.011 (0.293) [0.969]	0.168 (0.373) [0.653]	-0.118 (0.336) [0.726]	-1.054 (0.356) [0.003]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.001	0.443	0.036	0.001	0.000	0.764	0.092	0.003
Noise and Organizational Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	1837 (27)	1739 (27)	2152 (27)	1424 (26)	1851 (29)	1769 (29)	2177 (29)	1443 (27)
Panel (d)								
	Set 7: Alternative Fixed Effects				Set 8: Project Initiation			
	<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity		<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity	
	(1) Below median	(2) Above median	(3) Below median	(4) Above median	(1) Below median	(2) Above median	(3) Below median	(4) Above median
Management-Autonomy/Discretion	0.149 (0.063) [0.026]	-0.016 (0.101) [0.874]	0.077 (0.046) [0.101]	0.339 (0.095) [0.002]	0.038 (0.067) [0.577]	-0.015 (0.094) [0.873]	0.062 (0.090) [0.494]	0.157 (0.029) [<0.001]
Management-Incentives/Monitoring	-0.137 (0.060) [0.029]	-0.007 (0.082) [0.928]	-0.033 (0.025) [0.201]	0.039 (0.099) [0.700]	-0.153 (0.066) [0.027]	-0.296 (0.061) [<0.001]	-0.226 (0.055) [<0.001]	-0.092 (0.049) [0.074]
Management-Other	0.021 (0.061) [0.732]	0.033 (0.077) [0.674]	0.019 (0.040) [0.638]	-0.242 (0.080) [0.006]	0.059 (0.061) [0.335]	0.143 (0.066) [0.039]	0.062 (0.077) [0.428]	-0.057 (0.037) [0.133]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.003	0.952	0.044	0.031	0.098	0.029	0.017	0.001
Noise and Organizational Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type Within Sector	Task Type Within Sector	Task Type Within Sector	Task Type Within Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	1851 (29)	1769 (29)	2177 (29)	1443 (27)	1851 (29)	1769 (29)	2177 (29)	1443 (27)

Notes: Standard errors are in parentheses with p-values in brackets, and are clustered by organization throughout. All columns report OLS estimates. The dependent variable in all columns is a dummy variable that takes the value 1 if the task is completed and 0 otherwise. For each set of robustness checks, the sample for Columns 1-2 and 3-4 is split according to whether the target clarity or actual achievement clarity is above median versus below or equal to the median. In sets 1 through 7, the dependent variable is a dummy variable that takes the value 1 if the Task is fully completed and 0 otherwise. Set 2 excludes tasks implemented by the largest organization in terms of number of Tasks. Set 3 removes the 5 smallest organizations by number of tasks. Sets 4 and 5 exclude organizations at the top and bottom of the Autonomy/Discretion and Incentives/Monitoring management scales respectively. Set 6 estimates a specification analogous to Set 1 but using a fractional regression to account for the fact that task completion rates lie between zero and one. Set 7 controls for task-sector level fixed effects. Set 8 uses project initiation as a dependent variable. P-values reported for Wald test of coefficient equality. See text for details of controls and fixed effects. Figures are rounded to three decimal places.

Appendix Table A4 further shows the results on management practices and task completion

to be robust to alternative clusterings of the standard errors, including robust standard errors, allowing them to be clustered by task type within organization (so at the *jn* level), by task type within sector, and by sector. As expected the significance of the estimates varies depending on the choice of clustering, and are in some cases weaker and in some cases stronger than our default of clustering at organization level.

Table A4: Alternative Clustering - Management and Task Completion

	(1) Core: Organization	(2) Robust	(3) Task Type Within Organization	(4) Task Type Within Sector	(5) Sector
Management-Autonomy/Discretion	0.117 (0.066) [0.086]	0.117 (0.050) [0.019]	0.117 (0.069) [0.090]	0.117 (0.073) [0.119]	0.117 (0.043) [0.041]
Management-Incentives/Monitoring	-0.040 (0.062) [0.521]	-0.040 (0.047) [0.394]	-0.040 (0.072) [0.581]	-0.040 (0.078) [0.611]	-0.040 (0.073) [0.606]
Management-Other	-0.012 (0.059) [0.834]	-0.012 (0.046) [0.788]	-0.012 (0.069) [0.857]	-0.012 (0.069) [0.859]	-0.012 (0.053) [0.825]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.152	0.032	0.163	0.207	0.119
Noise Controls	Yes	Yes	Yes	Yes	Yes
Organizational Controls	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	3620 (30)	3620	3620 (167)	3620 (41)	3620 (6)

Notes: Standard errors are in parentheses with p-values in brackets, and are clustered by organization in Column 1, by output type within organization in Column 3, by output type within sector in Column 4, and by sector in Column 5. In Column 2, robust standard errors are reported. All columns report OLS estimates. The dependent variable is a dummy variable that takes the value 1 if the task is fully completed and 0 otherwise. P-values reported for Wald test of coefficient equality. Task Type fixed effects relate to whether the primary classification of the task is 'Advocacy and Policy Development', 'Financial & Budget Management', 'ICT Management and Research', 'Monitoring, Training and Personnel Management', 'Physical infrastructure', 'Permits and Regulation' or 'Procurement'. Sector fixed effects relate to whether the output is in the administration, environment, finance, infrastructure, security/diplomacy/justice or social sector. Task controls comprise task-level controls for whether the output is regularly implemented by the organization or a one off, whether the task is a bundle of interconnected tasks, and whether the division has to coordinate with actors external to government to implement the task. Organizational controls comprise a count of the number of interviews undertaken, which is a close approximation of the total number of employees, and organization-level controls for the share of the workforce with degrees, the share of the workforce with postgraduate qualifications, and the span of control. Noise controls are averages of indicators of the seniority, gender, and tenure of all respondents, the average time of day the interview was conducted and of the reliability of the information as coded by the interviewer. Figures are rounded to three decimal places.

Appendix Table A5 imposes these alternative clustering choices on the task clarity analysis. We omit our baseline specification from Table A5 for brevity, so Column 1 of Table A5 corresponds

to the clustering of Column 2 in Table A4, and so on. Again, the pattern identified in our main analysis emerges strongly regardless of the choice of clustering approach.

Table A5: Alternative Clustering - Management Practices and Task Clarity**Panel (a)**

	Set 1: Robust				Set 2: Task Type Within Organization			
	<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity		<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity	
	(1) Below median	(2) Above median	(3) Below median	(4) Above median	(1) Below median	(2) Above median	(3) Below median	(4) Above median
Management-Autonomy/Discretion	0.208 (0.074) [0.005]	0.019 (0.084) [0.824]	0.126 (0.063) [0.046]	0.337 (0.079) [<0.001]	0.207 (0.079) [0.009]	0.019 (0.088) [0.832]	0.126 (0.061) [0.039]	0.337 (0.090) [<0.001]
Management-Incentives/Monitoring	-0.138 (0.069) [0.047]	-0.020 (0.077) [0.790]	-0.023 (0.059) [0.694]	-0.012 (0.098) [0.901]	-0.138 (0.078) [0.079]	-0.020 (0.091) [0.823]	-0.023 (0.075) [0.757]	-0.012 (0.142) [0.921]
Management-Other	-0.007 (0.073) [0.922]	0.027 (0.073) [0.709]	-0.008 (0.060) [0.898]	-0.227 (0.077) [0.003]	-0.007 (0.082) [0.931]	0.027 (0.082) [0.740]	-0.008 (0.063) [0.904]	-0.227 (0.092) [0.015]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.001	0.748	0.095	0.005	0.004	0.772	0.155	0.030
Noise and Organizational Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	1851 (29)	1769 (29)	2177 (29)	1443 (27)	1851 (29)	1769 (29)	2177 (29)	1443 (27)

Panel (b)

	Set 3: Task Type Within Sector				Set 4: Sector			
	<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity		<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity	
	(1) Below median	(2) Above median	(3) Below median	(4) Above median	(1) Below median	(2) Above median	(3) Below median	(4) Above median
Management-Autonomy/Discretion	0.207 (0.076) [0.010]	0.019 (0.091) [0.839]	0.126 (0.055) [0.027]	0.337 (0.103) [0.002]	0.207 (0.045) [0.006]	0.019 (0.068) [0.795]	0.126 (0.022) [0.002]	0.337 (0.087) [0.012]
Management-Incentives/Monitoring	-0.138 (0.083) [0.103]	-0.020 (0.097) [0.834]	-0.023 (0.071) [0.742]	-0.012 (0.126) [0.923]	-0.138 (0.073) [0.117]	-0.020 (0.115) [0.866]	-0.023 (0.044) [0.621]	-0.012 (0.130) [0.928]
Management-Other	-0.007 (0.088) [0.936]	0.027 (0.081) [0.739]	-0.008 (0.056) [0.892]	-0.227 (0.097) [0.025]	-0.007 (0.068) [0.921]	0.027 (0.097) [0.789]	-0.008 (0.034) [0.830]	-0.227 (0.091) [0.054]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.004	0.792	0.149	0.058	0.008	0.769	0.025	0.033
Noise and Organizational Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	1851 (29)	1769 (29)	2177 (29)	1443 (27)	1851 (29)	1769 (29)	2177 (29)	1443 (27)

Notes: Standard errors are in parentheses with p-values in brackets, and are clustered by output type within organization in Set 2, by output type within sector in Set 3, and by sector in Set 4. In Set 1, robust standard errors are reported. Core specification (omitted for brevity) in Table 3 clusters at organization level. All columns report OLS estimates. The dependent variable is a dummy variable that takes the value 1 if the task is fully completed and 0 otherwise. For each set of robustness checks, the sample for Columns 1-2 and 3-4 is split according to whether the target clarity or actual achievement clarity is above median versus below or equal to the median. P-values reported for Wald test of coefficient equality. See text for details of controls and fixed effects. Figures are rounded to three decimal places.

Appendix Table A6 reproduces our main results using a recoded dependent variable for binary task completion. This recoded dependent variable codes a task as fully completed only if the mean of both coders' scores is a 5 - i.e. both coders adjudge it fully completed. This recoding reduces the variation in the dependent variable and thus is not our preferred specification. As expected, the significance level of the results is reduced and some of the point estimates change, but the overall pattern of the results with respect to our main hypotheses is nevertheless qualitatively similar to our main results.

Table A6: Alternative Coding of Dependent Variable

	(1) Core regression with pooled sample (Table 2, Column 5)	<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity	
		(2) Below median	(3) Above median	(4) Below median	(5) Above median
Management-Autonomy/Discretion	0.013 (0.039) [0.746]	0.126 (0.041) [0.005]	-0.128 (0.054) [0.024]	0.032 (0.026) [0.239]	0.102 (0.049) [0.047]
Management-Incentives/Monitoring	-0.052 (0.039) [0.192]	-0.156 (0.039) [<0.001]	-0.073 (0.037) [0.058]	-0.085 (0.029) [0.006]	0.016 (0.059) [0.788]
Management-Other	0.025 (0.037) [0.510]	0.028 (0.038) [0.469]	0.096 (0.033) [0.007]	0.049 (0.024) [0.051]	0.001 (0.026) [0.965]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.296	0.000	0.467	0.009	0.398
Noise and Organizational Controls	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	3620 (30)	1851 (29)	1769 (29)	2177 (29)	1443 (27)

Notes: Standard errors are in parentheses and p-values in brackets, and are clustered by organization throughout. All columns report OLS estimates. The dependent variable in all columns is a dummy variable that takes the value 1 if the task is completed and 0 otherwise. The sample for Columns 1-2 and 3-4 is split according to whether the target clarity or actual achievement clarity is above median versus below or equal to the median. P-values reported for Wald test of coefficient equality. See text for details of controls and fixed effects. Figures are rounded to three decimal places.

Finally, Appendix Tables A7 and A8 reproduce our main results using two other measures of task completion: whether the task's average completion status between both coders was 3 or higher (i.e. roughly half complete, Table A7); and whether the task's minimum completion status coding between the two coders was higher than 1 (i.e. the task was initiated and at least some work was done, Table A8). While point estimates vary, the same broad patterns of results emerge.

Table A7: Partial Completion as Dependent Variable

	(1) Core regression with pooled sample (Table 2, Column 5)	<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity	
		(2) Below median	(3) Above median	(4) Below median	(5) Above median
Management-Autonomy/Discretion	0.117 (0.078) [0.144]	0.140 (0.064) [0.039]	0.181 (0.104) [0.095]	0.204 (0.097) [0.044]	0.340 (0.032) [<0.001]
Management-Incentives/Monitoring	-0.082 (0.069) [0.245]	-0.116 (0.074) [0.127]	-0.093 (0.113) [0.421]	-0.060 (0.090) [0.516]	-0.100 (0.052) [0.066]
Management-Other	-0.026 (0.066) [0.691]	0.013 (0.060) [0.827]	-0.044 (0.092) [0.636]	-0.096 (0.090) [0.292]	-0.167 (0.040) [<0.001]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.119	0.033	0.151	0.059	0.000
Noise and Organizational Controls	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	3620 (30)	1851 (29)	1769 (29)	2177 (29)	1443 (27)

Notes: Standard errors are in parentheses with p-values in brackets, and are clustered by organization throughout. All columns report OLS estimates. The dependent variable in all columns is a dummy variable that takes the value 1 if the task's completion status was coded 3 or above (i.e. roughly half complete), and 0 otherwise. The sample for Columns 1-2 and 3-4 is split according to whether the target clarity or actual achievement clarity is above median versus below or equal to the median. P-values reported for Wald test of coefficient equality. See text for details of controls and fixed effects. Figures are rounded to three decimal places.

Table A8: Task Initiation as Dependent Variable

	(1) Core regression with pooled sample (Table 2, Column 5)	<i>Ex Ante</i> Task Clarity		<i>Ex Post</i> Task Clarity	
		(2) Below median	(3) Above median	(4) Below median	(5) Above median
Management-Autonomy/Discretion	-0.003 (0.080) [0.970]	0.038 (0.067) [0.577]	-0.015 (0.095) [0.873]	0.063 (0.091) [0.494]	0.158 (0.029) [<0.001]
Management-Incentives/Monitoring	-0.174 (0.068) [0.016]	-0.154 (0.066) [0.027]	-0.298 (0.062) [<0.001]	-0.227 (0.056) [<0.001]	-0.092 (0.050) [0.074]
Management-Other	0.067 (0.070) [0.341]	0.060 (0.061) [0.335]	0.144 (0.066) [0.039]	0.062 (0.077) [0.428]	-0.057 (0.037) [0.133]
Test: Autonomy/Discretion = Incentives/Monitoring (p-value)	0.171	0.098	0.029	0.017	0.001
Noise and Organizational Controls	Yes	Yes	Yes	Yes	Yes
Task Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector	Task Type, Sector
Observations (clusters)	3620 (30)	1851 (29)	1769 (29)	2177 (29)	1443 (27)

Notes: Standard errors are in parentheses with p-values in brackets, and are clustered by organization throughout. All columns report OLS estimates. The dependent variable in all columns is a dummy variable that takes the value 1 if the task's minimum completion status coding was higher than 1 (i.e. some work was done), and 0 otherwise. The sample for Columns 1-2 and 3-4 is split according to whether the target clarity or actual achievement clarity is above median versus below or equal to the median. P-values reported for Wald test of coefficient equality. See text for details of controls and fixed effects. Figures are rounded to three decimal places.