

# CAN TRAINING IMPROVE ORGANIZATIONAL CULTURE? EXPERIMENTAL EVIDENCE FROM GHANA'S CIVIL SERVICE \*

MICHEL AZULAI      IMRAN RASUL      DANIEL ROGGER  
MARTIN J. WILLIAMS<sup>†</sup>

OCTOBER 2020

## Abstract

Organizational culture is an important driver of organizational performance, but evidence on how to improve performance-oriented organizational cultures is scarce – especially in the public sector. We partnered with Ghana's Civil Service to design a new innovation training module geared towards such culture change and deliver it on a randomized basis to mid-level bureaucrats in central government. The intervention was delivered at full scale by integrating it into the Civil Service's standard training routine for one year. We find that the training improved organizational culture and performance 6-18 months post-training. Our design was split between an individual-focused training arm and one in which officials from an entire organizational unit were trained together. Our results are completely driven by the individual-focused arm, with the team-based treatment arm having no impact on culture or performance. We discuss potential explanations for this difference in effectiveness. Simple and scalable training interventions can thus have significant impacts on culture and performance, but their design matters. JEL CODES: H83, M53

---

\*We gratefully acknowledge funding from the International Growth Centre, Economic Development and Institutions, and World Bank i2i. We thank 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 Simprii-Duncan, and Liah Yecaló-Teclé for excellent research assistance, and Head of Civil Service Nana Agyekum-Dwamena, Mrs. Dora Dei-Tumi, Mr. Godwin Brocke, members of our project steering committee, Andrew Wyatt, and Stefan Dercon for invaluable input, and seminar audiences at the Paris School of Economics and BI Norwegian Business School for feedback. Special thanks are due to over 80 civil servants from across the service who generously and energetically led trainings, conducted interviews, collected data, and shared their insights. This project received ethics approval from the Blavatnik School of Government Departmental Research Ethics Committee (SSD-CUREC1A-BSG-C1A-17-001). All errors are our own.

<sup>†</sup>Azulai: University College London and the Institute for Fiscal Studies [m.azulai@ucl.ac.uk]; Rasul: University College London and the Institute for Fiscal Studies [i.rasul@ucl.ac.uk]; Rogger: World Bank Development Impact Evaluation Research Department [drogger@worldbank.org]; Williams (corresponding author): Blavatnik School of Government, University of Oxford, Radcliffe Observatory Quarter, Oxford OX2 6GG, United Kingdom [martin.williams@bsg.ox.ac.uk].

# 1 Introduction

Organizational culture is widely recognized as crucial for organizational performance, but how can organizations build performance-oriented cultures at scale? A large body of theory and evidence has demonstrated that organizational culture is associated with better performance across a range of dimensions (e.g. Schein 1985; Barney 1986; Gibbons and Henderson 2012; Guiso *et al* 2015; Martinez *et al* 2015; Graham *et al* 2017; Blader *et al* 2019; Gartenberg *et al* 2019). But while numerous case studies document how particular high-performing organizations have built performance-oriented cultures (Grindle 1997; Tandler 1997; Hoffer Gittel 2002; Madsen *et al* 2006; various in Gibbons and Henderson 2013; McDonnell 2017, 2020), often in idiosyncratic ways and with visionary leadership, there exists little causally identified research about whether and how organizations can build performance-oriented cultures through large-scale, systematized interventions.<sup>1</sup>

We address this question by partnering with Ghana’s Civil Service to design and deliver a new innovation training module across the entire Civil Service as a randomized controlled trial. This new training module was designed to encourage mid-level civil servants to identify potential work process improvements and put them into action. It was delivered at full scale by trainers from Ghana’s Civil Service Training Centre (CSTC), the Civil Service’s in-house training providers, with approximately one-quarter of all professional-grade civil servants participating in at least one training during the year-long intervention period. The training comprised three main elements: 1) basic problem-solving skills related to identifying productivity constraints, such as fishbone diagrams; 2) a motivational video that featured real Ghanaian civil servants talking about productivity routines in their team and giving examples of how they came up with and introduced innovative ideas to improve performance; and 3) applying these skills and motivations to the development of an innovation action plan relevant to their own work. Thus, only a small part of the training was focused on conveying discrete skills, with the main emphasis being on changing

---

<sup>1</sup>A partial exception is Thomas *et al*’s (2005) study finding some improvement in safety climate from randomized executive walk-rounds in hospitals. Blader *et al* (2019) find that randomizing the performance information provided to truck drivers interacts with a non-randomized values intervention, suggesting some effect of the latter. Guadelupe *et al* (2020) find that a randomized values affirmation experiment increases employee engagement.

motivations and norms around bottom-up work process improvement. We study the effects of this training on the type of mid-level bureaucrats engaged in the core civil service work of policymaking, oversight, and administration (as distinct from frontline public sector workers such as teachers and nurses, who are not included in our study).

As an empirical case, core civil service organizations in low- and middle-income countries are arguably one of the organizational types for which building strong cultures is both most challenging and most important. Such organizations produce outputs that are typically non-priced, difficult to measure, and distal from the desired outcomes (Wilson 1989), and individual-level performance measurement and management is difficult because individual bureaucrats tend to have to allocate effort via team production across multiple goals and discretion-intensive tasks (Holmstrom and Milgrom 1991; Dixit 2002; Prendergast 2003; Bandiera *et al* 2009). This typically leads civil servants to have substantial discretion over public action, as in other knowledge-intensive complex production settings. The potential importance of culture as a lever for improvement is further reinforced because the use of incentives and other common private-sector personnel management practices is often restricted in government, and the importance of intrinsic motivations among public servants makes their fit with organizational culture even more salient (Perry 1996; Benabou and Tirole 2006; Ashraf *et al* 2014; Andersson *et al* 2017). Consistent with this, while governments in low- and middle-income countries in particular are widely perceived to be characterized by poor performance and lack of performance orientation, a large body of qualitative evidence has shown that high-performing government organizations do exist in low- and middle-income countries and that cultivating positive organizational cultures has been a crucial part of shaping them (Grindle 1997; Tandler 1997; Owusu 2006; Leonard 2010; Roll 2014; McDonnell 2017, 2020). Such organizations therefore provide an ideal context to study whether and how organizational cultures can be improved with simple interventions.

As part of their routine training cycle trainees at CSTC were randomized to receive either the new training module or a pre-existing module on productivity (the 'status quo' training). We gathered data on trainees' level of knowledge about productivity concepts and skills immediately pre- and post-training, and analyzed the "action plans" they developed to improve work processes

in their team. We then conducted an in-person endline survey (6-18 months post-training) of the universe of eligible professional-grade civil servants in Ghana’s Civil Service, comprising 3,302 individuals in 57 ministries and departments, to measure the longer-term impacts of this training on trainees and their colleagues in their work team. To measure organizational culture, we adapt the widely used Safety Attitudes Questionnaire (Provonost *et al* 2006; Martinez *et al* 2015) from its health-sector context to the government setting. We supplement this survey with the collection and coding of administrative data on compliance with bureaucratic processes and task completion, and qualitative interviews conducted in parallel with the training intervention.<sup>2</sup>

In the short-term, we find that trainees randomized into the new training module improved their knowledge of productivity concepts relative to trainees randomized into the status quo training module. These learning gains persisted at endline. We also find some improvements in trainees’ understanding of what good organizational management practices look like.

Most importantly, we find that work teams that had at least one member randomized into the new training module showed large and significant improvements in our organizational culture index at endline. These improvements were concentrated in the sub-indices which were most directly targeted by the training (teamwork climate, performance climate, fostering new ideas, and relative performance perceptions). We also find that these trainings led to improvements in objective measures of team performance coded from administrative data, which suggests that our survey-based measure of culture is capturing real changes in how teams operate rather than just changed perceptions.

To understand the mechanism through which training affected team norms, a second cross-randomized treatment arm delivered the same new training module to an entire work team at the same time (as opposed to the first treatment arm, in which participants came from a range of organizations and were trained in a normal classroom setting). This was motivated by the theoretical view that organizational culture and performance are equilibrium outcomes from social interactions among team members, which thus might be difficult for a single trained individual to

---

<sup>2</sup>Our trial was pre-registered (<https://doi.org/10.1257/rct.1889-4.0>) and our pre-analysis plan is reproduced as Online Appendix F. This plan pre-specifies our intervention design, envisioned analytical strategy, and preliminary hypotheses, but due to the complex nature of the intervention and data we opted not to fully pre-specify all details of our analysis.

shift (Gibbons and Henderson 2013). We find weaker short-term learning gains from this second training arm, and no long-term effects on organizational culture. In addition to the weaker initial learning, this could potentially have been caused by trainings in the team setting inadvertently reinforcing rather than reshaping existing team culture, or by diffusing responsibility for leading change due to free-rider problems. Taken together, our findings demonstrate that while a simple, scalable intervention can improve organizational culture, the complex social dynamics through which culture is produced mean that the design of such interventions is non-trivial.

Our study connects with several bodies of academic literature. First, we provide experimental evidence that simple and low-cost intervention can improve organizational culture. This complements the large existing literatures on the linkages between performance and organizational culture (e.g. Schein 1985; Barney 1986; Gibbons and Henderson 2012; Guiso *et al* 2015; Martinez *et al* 2015; Graham *et al* 2017; Blader *et al* 2019; Gartenberg *et al* 2019) and case studies of efforts to build performance-oriented cultures in particular organizations (Grindle 1997; Tandler 1997; Madsen *et al* 2006; various in Gibbons and Henderson 2013; McDonnell 2017, 2020). Second, we contribute to a growing body of quantitative studies of bureaucratic effectiveness (Banerjee *et al* 2014; Ashraf *et al* 2014; Khan *et al* 2015; Linos 2018; Rasul and Rogger 2018; Leaver *et al* 2019; Rasul *et al* 2019). While these studies have focused on selection, monitoring, incentives, and management as tools for improvement, we focus on organizational culture and whether it can be shifted through training, providing some of the first field-experimental evidence on large-scale attempts to improve culture and performance across a range of organizations, personnel categories, and task types.

Section 2 briefly reviews the relevant theory and evidence on organizational culture and training as a tool to shape it. Section 3 details our context, the training intervention, and our randomization strategy, and Section 4 discusses our data sources. Section 5 presents our main analysis and results, and Section 6 discusses implications and limitations of our study.

## 2 Theory and Related Literature

### 2.1 Theories of organizational culture

The idea that shared norms, expectations, and cognitive frames within organizations are important determinants of individual behavior (and consequently of organizational performance) has deep roots in several disciplines (Van Maanen and Schein 1979; Schein 1985; Barney 1986; Rousseau 1995; DiMaggio 1997; Ashford *et al* 2009; Gibbons and Henderson 2012). While these literatures differ in the terms they use (culture, relational contracts, psychological contracts, values, etc.) and in the precise theoretical mechanisms they posit, they share a common theoretical core which revolves around the ideas that: individual behavior within organizations cannot be completely determined by formal rules and processes alone; many individual behaviors that influence collective performance require individuals to go “above and beyond” their formal duties in non-contractible and non-verifiable ways; and that shared norms and expectations develop within organizations that shape individuals’ willingness necessary to achieve consummate rather than perfunctory levels of performance (Hart and Moore 2008; Baron and Kreps 2013).

For the purpose of our analysis, we refer generically to such shared norms as “performance-oriented organizational cultures”. An extensive body of quantitative and qualitative empirical literature has developed to test the idea that organizational culture, relational contracts, and similar theoretical constructs are important for organizational performance, and while the extent and mechanisms of this relationship are much-debated there is increasing acceptance that shared culture within organizations is important for organizational performance (Riley *et al* 2011; Weaver *et al* 2013; Guiso *et al* 2015; Martinez *et al* 2015; Srivastava *et al* 2018; Blader *et al* 2019; Gartenberg *et al* 2019). Outside of academia, organizational leaders also emphasize the importance of culture, with Graham *et al* (2017, 2) finding that “91% of executives consider corporate culture to be ‘very important’ or ‘important’ at their firm, and 79% rank culture as at least a ‘top 5’ factor among all of the things that make their firms valuable”. In our analysis, we take it as given that performance-oriented organizational cultures are desirable and do not seek to test this link directly, instead focusing on whether and how it is possible to shape such cultures within

organizations.

A key feature of most theories of performance-oriented organizational culture is that they arise as equilibrium outcomes from the interactions of many agents. As such, an agent's actions depend on expectations about other agents' behavior, and changes in organizational norms often involve coordinated changes in expectations by multiple agents (Kreps 1990; Hermalin 2001; Chassang 2010; Gibbons and Henderson 2012; Baron and Kreps 2013). This closely approximates the dynamics of agents within an organization engaged in team production deciding whether to exert the minimal effort required to achieve perfunctory performance (which the organization can compel of them) or the greater effort required to achieve consummate performance (which depends on agents both valuing the outcome individually and sharing a collective norm of exerting the extra effort required to achieve it). If a group of agents in a poorly performing organization are all exerting minimal effort levels, this might constitute an equilibrium since no individual agent can improve team performance by increasing their effort without their colleagues also increasing theirs.

How can these equilibria be changed? Theory on this topic can be broadly split into two strands: 1) work that recognizes the importance of individual leadership in changing organizational culture; and 2) work that emphasizes coordinated group action.

The first strand on leadership emphasizes that individual workers in the organization, possibly outside positions of authority, can seek to lead culture change by engaging in costly actions or providing new information to signal new norms (Hermalin 1998; Bolton *et al* 2013).<sup>3</sup> Individual leadership in this sense can come from any member of the organization, not just those in formal leadership positions. Theories focused on the role of individual leadership in changing culture thus highlight the need to focus interventions on specific individuals who can lead changes in norms (who may or may not be workers in positions of formal authority).

On the other hand, the self-reinforcing dynamics of organizational culture might make it difficult for any one individual to shift the collective equilibrium. This is particularly true since the costs of such efforts are borne by individual agents, while the realization of benefits depends on

---

<sup>3</sup>This is consistent with the literature on voice and bottom-up work process improvements in organizations (e.g. Ashford *et al* 2009; Williams and Yecaló-Teclé 2020), which (while not explicitly focused on culture formation) emphasizes the role that lower- and middle-level individuals who are not in positions of formal authority can play in improving organizational practices and performance.

the actions of other agents. Thus, not only might a single individual not be able to shift collective norms on their own, but they may not have an incentive even to try. Efforts to change organizational culture may therefore require simultaneously inducing change across all (or at least a critical mass) of agents simultaneously.

While both the individual leadership-focused and collective-focused approaches to the development of organizational culture are theoretically plausible, empirical evidence on their relative importance is scarce - particularly for the public sector organizations that provide our empirical context. Understanding the extent to which each theoretical mechanism can be leveraged to improve organizational culture, and which approach is more effective in different contexts, is thus an important topic.

## 2.2 Training and culture

How might training be a tool for cultivating performance-oriented cultures? While trainings of various sorts are used by most organizations in both the private and public sectors, evidence on their effectiveness in the public sector is very limited. For example, Finan *et al*'s (2017) review of research on the personnel economics of the state includes no studies of training interventions with core (i.e. non-frontline) civil servants, and Hansen and Tummers's (2020) systematic review of field experiments in public administration identified no examples of training interventions for mid-level civil servants apart from studies with daycare and school managers (Andersen *et al* 2018; Jensen 2018; An *et al* 2019; Jakobsen *et al* 2019).<sup>4</sup>

Training is normally viewed as a way to impart specific skills to individual trainees (Baser and Morgan 2008). In this sense it can be understood as a way to improve the human capital of bureaucrats who already make up part of a government organization. However, whether and how these skills are actually used post-training, especially when bureaucrats work in teams, may also depend on the work norms of the bureaucrats and their colleagues. While these norms may be

---

<sup>4</sup>One organizational setting in which training has been frequently evaluated is small business training on management practices for micro- and small enterprises in low and medium income countries. While one review found that few such training programs had statistically significant effects (McKenzie and Woodruff 2014), a recent meta-analysis found statistically significant and economically substantive average effects (McKenzie 2020), suggesting that the small scale of many training interventions may have limited researchers' ability to detect small but meaningful effects.



determined in large part by fixed worker characteristics, at the margin it may be possible to shift or activate them. Trainings may thus bundle technical skills training with efforts to shape trainees' norms about work by presenting role models, setting expectations, or trying to socialize employees into the organizational culture (Van Maanen and Schein 1979; Bernard *et al* 2017). Blader *et al* (2019) illustrate this indirectly in their study of truck drivers, by showing how an organizational values intervention shifts individual employees' understandings of organizational culture in ways that interact with firm-level management practices.

But even if trainings can improve individual skills and norms, is that enough to improve team performance? Most mid-level civil servants (as with complex production settings in the private sector) are engaged primarily in team-based rather than individual-based modes of production, in which the production function exhibits strong complementarities across the skill and effort of different workers (Alchian and Demsetz 1972; Holmstrom 1982). At the extreme, a training that improved the skill and effort of one worker might have no impact on team output. This suggests that even if training is effective at shifting individual trainees' skills and motivations, it may be necessary for training to be directed at treating an entire team, in order to simultaneously shift their skills and collective norms. This approach is theoretically consistent with the view of organizational culture that views it as an equilibrium within a group (Chassang 2010; Gibbons and Henderson 2012; Baron and Kreps 2013). Thus the training *cohort* might matter as much as the training *content*.

Alternatively, by targeting training at specific individuals, this can provide the scope for individuals to lead their teams and avoid miscoordination and expectation errors across workers in the organization (Hermalin 1998; Bolton *et al* 2013). Training entire teams may also not help teams coordinate on new norms if individuals in higher hierarchical positions use their formal authority to undermine change efforts proposed by lower-ranking staff (Aghion and Tirole 1997; Baker *et al* 1999; Williams and Yecaló-Teclé 2020), thus reinforcing rather than changing organizational culture. Targeting training at individuals rather than entire teams could thus be more effective by allowing the trained agents to pursue change in ways that were less likely to be vetoed by higher-ups.

There is therefore a need to empirically test this theoretical framework. First, can training be an effective instrument to shift individual skills and norms related to performance orientation? Second, are such trainings more effective at shifting collective norms when they intervene on an entire team or when they target individuals within a team? The next section outlines our research design for addressing these questions in our empirical context of the Ghana Civil Service.

## 3 Context and Intervention

### 3.1 Context

Ghana’s Civil Service consists of 57 central government ministries and departments that primarily perform core bureaucratic functions of policymaking, administration, and oversight of service delivery. However, it excludes most frontline public service delivery agencies (e.g. the Ghana Education Service) as well as all local government employees and staff of semi-autonomous agencies (e.g. the Bank of Ghana, Ghana Audit Service). The Civil Service thus includes the staff and organizations that are responsible for developing new policies and strategic directions, overseeing the implementation and delivery of policy, monitoring and evaluating frontline service delivery agencies and staff, conducting public procurement, and advising government, as well as some direct service delivery, regulation, and permit processing.

Each ministry and department (henceforth referred to as “organizations”, for brevity) has defined responsibility for a particular sector and set of functions, some of which are recurring and some of which are project-based. The exact definition of the tasks and outputs to be undertaken in a given year happens via an annual planning, budgeting, and reporting cycle, in which long-term strategic plans and other government priorities are translated into specific programs of work. Responsibility for these tasks is distributed among the organization’s divisions, which are the primary sub-unit into which organizations are divided. Organizations are typically comprised of four to ten divisions, each headed by a director; we refer to “divisions” and “teams” interchangeably throughout the paper.

Civil Service organizations in Ghana are staffed almost exclusively by career civil servants, with

the exception of politically appointed ministers and a handful of special advisors. Recruitment, staff assignment, transfers, promotions, and termination are all handled centrally by the Office of the Head of Civil Service (OHCS), so while organizations have a say in personnel matters and can make requests of OHCS, there is substantial uniformity of human resource matters across organizations (at least at a *de jure* level). Though a new presidential administration took office in December 2016, preceding the delivery of the training intervention from February 2017 to March 2018, there was little personnel turnover in the Civil Service.

Civil servants are formally divided into “senior” and “junior” staff. Rather than referring to the colloquial meaning of individuals’ age or tenure, in Ghana’s Civil Service (as in many former British colonies) this distinction depends primarily on the individual’s educational qualifications and the career track into which they were hired. Senior staff comprise the core of officials in professional and administrative positions, usually with a university degree or post-graduate qualification, while junior staff comprise support staff like drivers, secretaries, and cleaners. Since “senior” staff fill the core bureaucratic roles responsible for task completion and service delivery we focus our analysis on this set of staff and exclude junior staff from our data collection, although since these distinctions refer to career tracks rather than tenure, our population of senior staff includes the full spectrum of rank and tenure within the service, from recent entry-level hires to heads of divisions.

The Civil Service has a standard training cycle in which all staff undertake a two-week training at CSTC as part of the professional development process embedded into the promotion cycle. Staff are required to complete the two-week CSTC training (called the “Scheme of Service” [SOS] training) relevant to their grade level in order to become eligible to interview for promotion. Since OHCS specifies the number of years that must be spent at one grade before becoming eligible for promotion, this means that all staff below the level of Director (head of division, the second-highest rank attainable) undertake a SOS training at regular three- to five-year intervals. In practice, though, the timing of when individual staff undertake the CSTC training depends also on the idiosyncrasies of workload and funding availability, so some staff undertake the SOS training slightly earlier or later than their promotion schedule would imply. These SOS training sessions are conducted in person in variably sized cohorts of around 10-40 civil servants (all of the same

level, but from different organizations) at CSTC’s campus in Accra.

The two-week SOS training includes a wide range of content covering topics such as administrative writing, policy development and analysis, and - most importantly for our purposes - one day on productivity improvement. All content is developed by CSTC’s own trainers (themselves civil servants), sometimes with funding from a development partner, and all trainings are delivered as interactive lectures by CSTC trainers. The existing productivity curriculum focused mainly on conveying understanding of abstract concepts, such as the definition of productivity, the importance of productivity in the public sector, and terms associated with the Japanese kaizen approach to productivity. This abstract approach may have limited the practical relevance of these trainings: for example, our baseline survey (discussed below in more detail) found that while 89 percent of civil servants desired more training, only 59 percent thought that their existing trainings were effective, suggesting significant room for improvement.

### **3.2 Training Intervention and Randomization**

In 2016 and early 2017, we worked with OHCS and CSTC to design a new one-day productivity training, called Training for Productivity (TFP). The Civil Service’s decision to develop and deliver a new training module was driven by the main finding of our 2015 baseline survey that organizations delegating and empowering their staff with greater autonomy, discretion, and flexibility was associated with higher rates of task completion (Rasul *et al* 2019). The finding that the creativity and local knowledge of lower- and middle-level bureaucrats was important for productivity thus motivated the idea that a productivity and innovation training intervention could further equip and empower bureaucrats to identify and implement work process improvements. The training curriculum was drafted by an international development consultant but was then extensively adapted and customized by OHCS and CSTC to fit the local context, which is a widely used and highly scalable approach to training delivery used in international development.

The TFP training aimed to be more applied and interactive than the existing productivity training, and comprised three interwoven components: 1) basic problem-solving skills related to identifying productivity constraints, such as fishbone diagrams; 2) a motivational video developed

by the TFP program that featured real Ghanaian civil servants talking about productivity routines in their team and giving examples of how they came up with and introduced innovative ideas to improve performance, as well as exhortations from the Head of Civil Service and other senior officials about the importance of innovation and productivity; and 3) applying these skills and motivations to the development of an action plan, in which trainees develop ideas for small-scale work process innovations to implement in their team. The training thus aimed to improve both the skills and the motivations of participating civil servants, emphasize the importance of individual initiative (in contrast to often-hierarchical bureaucrats norms), and help participants apply them to their day-to-day work - even down to role-playing how they would bring up their new ideas with their colleagues and superiors. Appendix A contains additional details of the training content, and Appendix B contains the action plan template.

CSTC rolled out the new TFP training at full scale as part of its 2017 training calendar, through a randomized controlled trial with two treatment arms. Under the first arm (T1), trainees attending their regular ten-day training at CSTC were randomly selected to receive either the standard one-day productivity training or the new one-day TFP productivity training. On the day of the productivity module (usually the penultimate day of the ten-day training), the training cohort was divided into two groups based on a randomized selection conducted by a member of the research team. One group then received the standard productivity training for the remainder of the day, and the other group received the TFP productivity training. Treatment was blinded in the sense that participants did not know which training they were receiving, as both trainings were delivered by CSTC staff on CSTC-branded materials in CSTC classrooms. During the treatment year of 2017, TFP trainers who were involved in productivity training delivered either the standard training or the new TFP training, but not both, in order to avoid content or teaching style from one module being transposed into the other. This first arm (T1) thus aimed to evaluate the impact of the *content* of the training curriculum, since the delivery of both trainings was identical.

The second treatment arm (T2) was designed in response to CSTC's experience that a major constraint on training effectiveness seemed not necessarily to be what happened on the training day itself, but that trainees often returned to a work environment in which their colleagues maintain a

“business as usual” mentality and show little enthusiasm for change, thus discouraging the newly trained officer. This matches the idea that organizational practices, norms, and expectations often include aspects of collective action. The second treatment arm thus maintained the same content as T1, but varied its *delivery* by conducting the training with an entire division of 5-20 staff simultaneously (as opposed to with a training cohort from other organizations). All other important aspects of training content and delivery were the same, with the exception of minor logistical changes necessitated by practical considerations; see Appendix A for full details. Whereas T1 took the standard approach of delivering training to a group of individuals, the underlying hypotheses tested by T2 was that delivering training to an entire work team simultaneously was likely to increase the probability that the learning from the training lessons would translate into changes in behavior post-training.

These two treatment arms were cross-randomized, as summarized by Table 1. Within each SOS training, for the day of training that is focused on productivity, trainees are randomized into the “old” *status quo* SOS module (the control group) or the “new” TFP module (T1).<sup>5</sup> At the end of the two-week SOS training, approximately 40% of trainees were randomly selected and informed that a day of productivity training would be organized for them and their entire work team of 5-20 people in the same division, including officers of *all* ranks – including, crucially, their director (boss). These team-based productivity trainings (T2) were conducted 3-6 weeks after the end of the SOS training, at the offices of OHCS. While the T1 trainings were conducted entirely during calendar year 2017, this meant that some T2 trainings were conducted in early 2018. By construction, a more minor difference between the trainings is that in the T2 trainings, at least one participant had undergone the SOS training previously (with either the “old” SOS productivity module or the “new” TFP T1 productivity module).

The T2 arm generates treatment and control groups at the division level. In combination with the T1 arm, this allows us to evaluate the effect of the applied productivity training when only individual trainings are undertaken, when the whole division is subject to treatment, and when

---

<sup>5</sup>This within-session randomization approach has the advantage of creating a clean counterfactual, but could create concerns about spillovers if trainees from one group shared ideas or material with trainees from the other group. Our research team and CSTC trainers actively monitored this, and found no evidence of trainees from different groups discussing training content or ideas with each other to a meaningful degree.

both individuals and their divisions are treated.

While this randomization strategy produces a group of individuals who did not undertake training during the intervention period, we do not consider these individuals as a “control” group for several reasons. First, the timing of selection into SOS trainings is not randomized, so that we only cleanly identify the additional impact of being involved in TFP training relative to the *status quo* civil service training. Second, since officers attend scheme of service trainings every three years, many officials will have experienced the *status quo* training in past years, so most individuals in this group had actually undergone training in previous (but not the current) years. We therefore exclude this group from our analysis, and focus on estimating the impacts of the two novel TFP T1 and T2 trainings relative to the counterfactual of the *status quo* SOS productivity training. Treatment groups were balanced on observable characteristics for the randomizations into TFP T1 and TFP T2 (see Tables A2-A4 in Appendix A.1). The training allocation is summarized in Figure 1.

Figure 1: Treatment Assignment

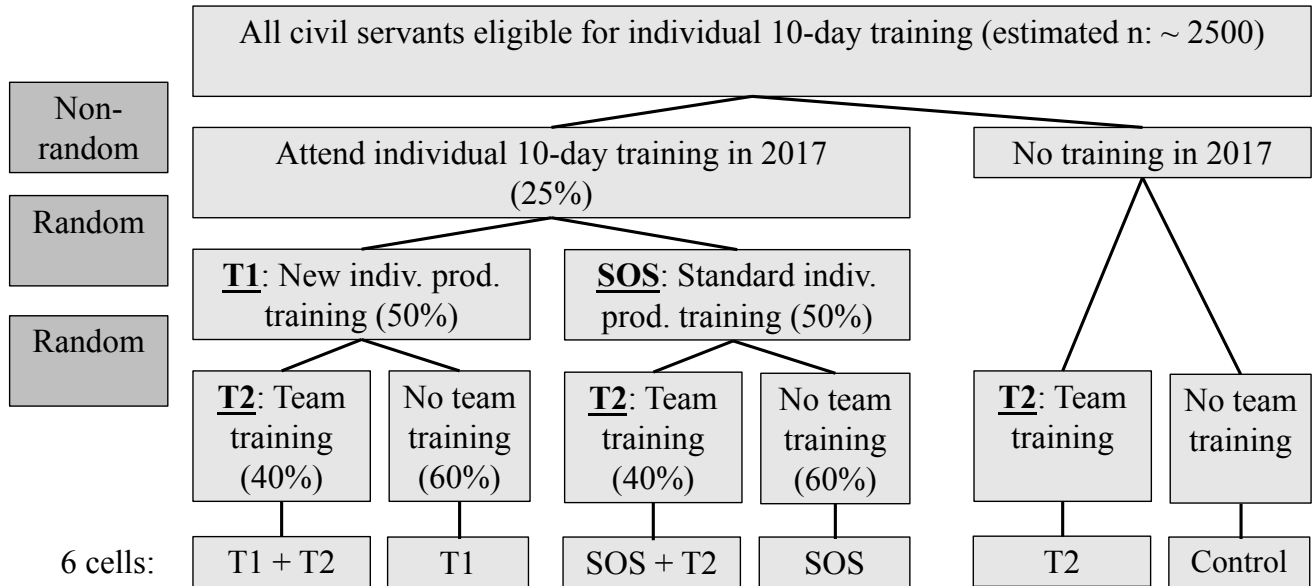


Table 1 presents basic descriptives of the training session numbers, as well as characteristics of the different training types (coded by training facilitators who sat in on each session). A total of 422 individuals from 132 divisions participated in the two individual-based training arms (SOS

and TFP T1), out of which 93 divisions comprising 694 individuals were randomly selected for the follow-up TFP T2 training sessions.<sup>6</sup> As intended, the TFP module (T1 and T2) tended to have higher engagement levels, be more practical and use more empirical examples, and have greater use of team work and role playing to practice applying lessons from training. As expected, the major difference between the two branches of the TFP training was that the team-based T2 had a much greater (73 percent of sessions) use of task-specific language than the individual-based T1 (18 percent of sessions), since the presence of many staff from the same division meant that the training could discuss the actual work of the division rather than hypothetical examples.

**Table 1: Training descriptives**

	<i>Status quo</i> productivity training (SOS)	T1 (Individual- based)	T2 (Team- based)
<i>Basic Statistics</i>			
<b>Number of Trainings</b>	17	11	85
<b>Number of Trainees</b>	283	139	694
<b>Number of Orgs Represented</b>	40	36	34
<b>Number of Divisions Represented</b>	132	90	93
<i>Training Characteristics (Perceptions of Training Facilitator)</i>			
<b>Level of Engagement of Students</b>	0.60	0.73	0.86
<b>Use of Team Work</b>	0.00	0.82	0.80
<b>Use of Role Plays</b>	0.00	0.27	0.36
<b>Extent to Which Training was Practical</b>	0.40	0.73	0.78
<b>Use of Civil Service Related Examples</b>	0.44	0.64	0.72
<b>Use of Task-Specific Language</b>	0.11	0.18	0.73
<b>Trainee Questions That Implied a Lack of Basic Understanding</b>	0.00	0.18	0.00
<b>Quality of Trainer's Responses to Questions</b>	0.89	0.91	0.81

Notes: The table reports descriptives on the various trainings delivered during the study period. In the second panel, the descriptives refer to the productivity session and team training quality, subjectively evaluated by a training facilitator (who observed the training) on the following margins: (i) "How practical would you say the training was?"; (ii) "How engaged were the students in the training?"; (iii) "Use of role plays in class?"; (iv) "Use of civil service-related examples?"; (v) "Working in teams?"; (vi) "Use of civil service-related working definitions?"; (vii) "Frequency of questions from participants showing a lack of understanding?"; (viii) "Trainer's responses to questions?." The descriptives report the proportion of sessions that scored 3.5 or higher on a five-point Likert scale for each question.

Participants in each of these training sessions were asked by the trainers define action plans outlining the most strategic reform they could undertake in their work team. Perhaps surprisingly, the most common focus of these action plans was to find ways to actually implement rules or

<sup>6</sup>For T2 the number of trainings is slightly lower than the number of divisions represented because a small number of trainings combined two divisions for logistical reasons.



practices that were already supposed to exist but were not being executed (46.8 percent of all action plans), and the second-most common was to expand the scope of an existing practice (21.3 percent). In contrast, only 19.1 percent of all action plans were actually proposing minor or major new innovations, with the remainder being related to seeking additional resources, holding meetings, or other changes.<sup>7</sup> This suggests that most of the ideas civil servants had to improve performance were fairly basic improvements operating within the existing procedural framework of the bureaucracy, mostly on how to implement or expand useful practices that were on the books but were failing to be executed. Interestingly, this was true of action plans from all three training types, with no significant differences across training types (at least in terms of this categorization).

## 4 Data and Outcomes

We collaborated closely with Ghana’s Civil Service over several years to collect a rich range of data on the delivery and impacts of the TFP training intervention, which allows us to examine both the immediate impacts of the trainings as well as their longer-term impacts on bureaucratic behavior and performance. In this section we describe these data sets and the key dependent variables they yield. These data sources fall into three main categories: the immediate pre- and post-training surveys we use to examine short-term effects; our baseline and endline surveys, which we use to examine longer-term effects, in particular on organizational culture and management quality; and administrative data we used to code our measures of administrative process quality and task completion. Appendix C provides further details on data collection.

### 4.1 Pre- and Post-Training Surveys

On the first and final days of each ten-day SOS training, all training participants were administered a short paper-based test in which they answered a set of multiple choice questions bearing on their knowledge about the topics covered in the SOS training, including both the TFP and existing productivity modules as well as the non-productivity-focused modules. These tests also

---

<sup>7</sup>We worked with civil servants to categorize all submitted action plans using this simple classification, for descriptive purposes.

included questions pertaining to the respondents’ perceptions of what constitutes good management practice, using a closed-ended and abbreviated version of the World Management Survey-style questions widely used to measure the quality of organizational management practices (see Appendix C.2 for details), which we also implement in full in the baseline and endline surveys discussed below.

Together, the results of these pre- and post-training tests allow us to measure the immediate impacts of the training on participants’: 1) learning gain related to productivity and other concepts; and 2) understanding of good versus bad organizational management practices relevant to productivity. Appendix C.2 contains further details on these pre- and post-tests.

## 4.2 Baseline and Endline Surveys

Our data collection began with a survey of the universe of senior-grade civil servants across Ghana’s Civil Service, conducted from August to November 2015. Teams of enumerators from the Civil Service, working with members of the research team, conducted in-person surveys with 2,971 senior officials across 45 organizations. This survey was repeated in June through October 2018, with 3,302 senior officials across 57 organizations.<sup>8</sup> Appendix C.1 contains full details of how the survey was designed and administered.

In addition to a wide range of basic characteristics about the respondent and their workplace, our surveys applied a version of the measurement scale of organizational culture used by Martinez *et al* (2015) and originally from Sexton *et al* (2006) from the hospital context which we adapted to our public administration context. This provides a measure of the extent to which individuals feel that their organization’s work culture is performance-oriented, both in terms of demanding performance from staff as well as empowering them to achieve it. While most scale items in our organizational culture scale are near-verbatim from Sexton *et al* (2006), to adapt the scale from the hospital to bureaucratic context we rephrased some items, eliminated some less-relevant scale items, and added two new sub-indices (“Fostering new ideas” and “Relative performance”). Table 2 shows the final set of scale items, all of which respondents were asked to respond to on a five-

---

<sup>8</sup>The larger number of organizations included in the endline survey was due to the creation and splitting of some organizations after the completion of the baseline survey.

point Likert scale. We aggregate these items into an overall index and the component sub-indices by taking the mean of each variable at the division level, z-scoring each variable using the cross-division means and standard deviations, and computing aggregate indices by averaging the z-scores of the different variables (following O'Brien [1984] and Kling, Liebman, Katz [2007]). Throughout the paper, the organizational culture index and its sub-indices are presented as normalized z-scores.

**Table 2: Organizational culture scale and sub-indices**

<i>Sub-index</i>	<i>Scale items</i>	
<i>Teamwork climate</i>	It is easy for personnel here to ask questions when there is something that they do not understand  The managers and other officers here work together as a well-coordinated team	Disagreements in this division are resolved appropriately (i.e., not who is right, but what is best for the service)
<i>Performance climate</i>	The culture in this division makes it easy to learn from the errors of others  You know the proper channels to direct questions regarding correct bureaucratic process in this division  You receive appropriate feedback about your performance	Bureaucratic errors are handled appropriately in this division  You are encouraged by your colleagues to report any work concerns you may have  You would feel happy being served as a Ghanaian citizen by this division
<i>Stress recognition</i>	Fatigue impairs your performance during high-pressure situations (e.g. when there are heavy demands on your division) [reverse-scaled]	
<i>Perceptions of management</i>	Staff (divisional management) doesn't knowingly compromise division services  You get adequate, timely info about events that might affect your work from your division.	Staff (divisional management) supports your daily efforts
<i>Working conditions</i>	All the necessary information for diagnostic and effective decision making is routinely available to you	Trainees in your division are adequately supervised
<i>Fostering new ideas</i>	Your suggestions about work place productivity would be acted upon if you expressed them to management  You can see lots of ways to make your division work better.	Staff (divisional management) in this division are quick to adopt (are open to) new ways of doing things.
<i>Relative performance</i>	Your division works better than others in this organization	Staff (divisional management) in this division are doing a good job

**Notes:** Items in sub-indices "Teamwork climate", "Performance climate", "Stress recognition", "Perceptions of management", and "Working conditions" are drawn/adapted from Sexton *et al* (2006). Text in (parentheses) indicates item phrasing administered to non-management-level staff; all other items are identical for all respondents. See Appendix Table A7 for further details of scale construction and validation.

To validate our adapted organizational culture index, we used our baseline survey data to examine the relationship between organizational culture and task completion (a measure of performance we construct using administrative data; see below for details). We control for an extensive range of

organizational and task characteristics, and find that our organizational culture index is positively and significantly correlated with task completion rates (results not shown for brevity; see Table A6 in Online Appendix C.1.2). While the relationship between culture and performance is not the primary subject of this paper, and the observational nature of the data at baseline means that this positive relationship between culture and performance is not necessarily causal, it is nonetheless reassuring that our adapted organizational culture index does seem to be relevant for performance in the context of public sector bureaucracies.

Our endline survey also included a range of questions that directly probed respondents about their learning from training (to test persistence of learning gain) as well as their follow-up actions after training, in particular with respect to implementing the action plan they developed during training. These are listed in full in Online Appendix Tables A8 and A9. We aggregate these numerous indicators into four indices following Anderson (2008): perceived obstacles to applying learning from the training; steps taken to implement the action plan; perceived obstacles to action plan implementation; and a combined index of overall training follow-up which combines all these indicators into one joint index. We treat these learning persistence and training follow-up variables as intermediate outcomes.

Finally, we included a module that adapted the widely used World Management Survey approach to measuring organizational management quality that has been used across a wide range of organizational types (Bloom and Van Reenen 2007; Bloom *et al* 2012; Bloom *et al* 2014; Rasul and Rogger 2018; Rasul *et al* 2019) to the public sector context. This approach combines a number of methodological innovations to collect high-quality data on management practices and assess organizational management quality that avoids many of the limitations of subjective, Likert-scale measures. To build this measure, enumerators asked respondents to describe the actual state of management practices in their organization across 18 areas of management practice, covering topics such as monitoring, incentives, staff’s understanding of roles, flexibility in adapting new practices, staffing practices, and setting of targets (see Table A7 in Online Appendix C.1.2 for full details).

To improve the accuracy of responses and minimize social desirability bias, enumerators ask

these questions in a style more akin to conducting qualitative interviews than to administering a closed-ended questionnaire, starting by asking an open-ended question about how that practice works in the organization and responding with probing follow-up questions and requests for illustrative examples. This allows the enumerator to gain an understanding of the actual practice used in the organization, not what the practice is supposed to be on paper or what the respondent thinks is the “correct” practice. Based on this understanding, the enumerator then uses their own judgment to score each practice on a continuous scale from one to five, where one corresponds to very poor practice and five corresponds to excellent practice, benchmarking against a pre-defined scoring grid. While efforts to define international “best practices” are rightly frowned upon, the benchmark practices in the scoring grid are not prescriptive of exactly what an organization should do, but rather measure whether an organization has a defined and coherent management strategy for each practice that is actually executed in reality. This approach to measuring management quality has been extensively validated in various types of organizations around the world (Bloom *et al* 2014), and our adaptation of it to the Ghanaian public sector context was pre-validated with Ghana’s Civil Service.

We then normalized each practice score into a z-score, and took the unweighted average of these z-scores to define overall management quality z-scores at the individual level; we discuss different ways to aggregate these individual-level scores into organization-level scores in the next section. Following Rasul and Rogger (2018) and Rasul *et al* (2019), we also define three sub-indices that capture different “styles” of management: 1) a monitoring/incentives sub-index that captures the “top-down” levers of control; 2) an autonomy/discretion sub-index that measures the use of “bottom-up” efforts to empower staff and provide them with flexibility; and 3) a residual sub-index of the remaining practices. This overall management quality index and the thematic sub-indices serve as one set of outcome variables in our analysis; we discuss their interpretation in the following section.

### 4.3 Administrative Data on Performance

As additional non-survey-based measures of team performance, we also undertook two extensive exercises in collecting and coding administrative data. While measuring bureaucratic performance objectively is challenging and each measure in isolation has some limitations, the use of administrative data sources allows us to create measures of actual performance that are separate from our surveys and that are not subject to common concerns with survey-based measures such as social desirability bias.

First, we collected and coded official administrative reports of organizations' completion of their workplans in order to obtain a measure of task completion. Each organization is required by OHCS to submit quarterly and annual reports that detail their completion of the tasks that comprised their annual workplan. These tasks span the full spectrum of bureaucratic activity, from procurement to personnel management, infrastructure development, policy development and analysis, permitting and regulation, financial management, and other task types. Unlike other studies which focus on the delivery of a narrow type of service or the execution of specific functions like procurement (e.g. Khan *et al* 2015; Best *et al* 2018), we are thus able to study bureaucratic performance across the full range of tasks undertaken by governments using a consistent metric. Our research team worked with a group of civil servants from OHCS to compile these, extract and standardize data (often from hard-copy reports), and code the completion of each task on continuous scale from one to five, with a score of one representing "No action was taken towards achieving the target", a score of three representing a task for which "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 corresponding to "The target for the task has been reached or surpassed." We were thus able to code the completion of 1,473 tasks at division-level across 47 organizations for the endline year 2018. Appendix C.3 contains further details on the task completion data and coding, and Appendix D contains a sample organizational report from which data was extracted.

Second, while task completion is an output-based measure of bureaucratic performance, another important type of performance measure for core civil service organizations is the quality of

basic administrative processes. Effectively managing administrative processes - in particular, filing and record-keeping - is a crucial and basic task for governments, but one which governments in weakly institutionalized settings often fail to perform adequately. We therefore collaborated with OHCS and the Public Records and Archives Administration Department (PRAAD) to measure the quality of administrative processes in divisions across the Civil Service in 2018. We randomly sampled 763 recent files from 256 divisions across 55 organizations and worked with retired senior civil servants to code the quality and completeness of each file, according to a scoring grid designed in conjunction with expert civil servants. For analysis in this paper, we used only the subset of files that were open (i.e. in use) during or after the training period and which thus could have been affected by the training, and excluded files that had been closed prior to the start of the intervention. Appendix C.4 contains full details of this coding process. This allowed us to create division-level measures of administrative process along several dimensions of quality. We use this to construct two main sub-indices of the quality of administration: the quality of procedure (capturing the extent to which files complied with rules about how files should be handled, compiled, and circulated) and the quality of content (measuring the extent to which the information compiled, analysis undertaken, and decisions made within files reflected good administrative practice as expected by the Civil Service).

Altogether, the data collected from these sources provides us an extremely rich picture of bureaucrats' work lives and the potential impacts of training. However, collecting such rich data in an ever-changing Civil Service with limited existing administrative data infrastructure posed some practical challenges. In particular, matching of individuals and divisions across different survey waves and administrative datasets was challenging, as name spellings and division names and structure were all reported variably and with some errors and missing data, leading to attrition in tracing individuals across datasets. This challenge was compounded by the normal creation, merging, and splitting of organizations and internal reorganizations which occurred during the several years of data collection. Appendix E discusses the data linking process in more detail and examines how matching challenges affected sample balance and composition. On the whole, these challenges somewhat reduce our sample size, but we find little consistent evidence that they bias

its composition in ways that might affect our analysis.

## 5 Analysis

Our analysis proceeds in four steps. First, we use data from pre- and post-training tests to investigate the immediate impacts on officers' conceptual knowledge and their perceptions of what constitutes good management practice. Second, we use data from the endline survey to examine learning persistence and follow-up actions among trainees. Third and most importantly, we examine the training's impacts on division-level (i.e. team-level) organizational culture at endline (6-18 months post-training). Finally, we examine whether the training also had long-term impacts on measures of division-level performance (the quality of administrative process, task completion, and management quality).

In each of these sub-sections the aim of our analysis is the same: to understand 1) the impacts of different training *content*, by comparing TFP T1 to the standard SOS productivity training; and 2) the impacts of different training *delivery*, by comparing the impacts of delivering the TFP module via team-based (T2) as opposed to individual-based (T1) trainings.

### 5.1 Short-term Training Impacts

To measure the short-run impacts of the training, we administered short, paper-based, closed-ended questionnaire to each training cohort. For the individual-level SOS and TFP T1 trainings, these were administered on the first and final days of the ten-day training, and for the team-based TFP T2 training they were administered first thing in the morning and then upon completion of the training in the afternoon. These questionnaires covered two sets of topics. First, they included multiple-choice questions on key productivity-related concepts covered in both the SOS and TFP modules, as well as non-productivity-related concepts that were covered in other days of the SOS training (but not at all in the TFP training). This enables us to examine direct learning gain from the trainings. Second, we converted the standard WMS-style coding grid for measuring management quality (Rasul and Rogger 2018; Rasul *et al* 2019) into a set of multiple-choice



questions that asked respondents to select which description of management they felt represented good practice.<sup>9</sup> While changing management perceptions was not directly targeted by the training and was not an explicit part of the training content, these questions nonetheless give a sense of the extent to which the trainings led to broader changes in outlook as well as to greater topic-specific knowledge. See Appendix C.2 for full details of questions and scoring.

We first estimate the difference in learning and management perceptions between officers who undertook either the standard SOS productivity module or the TFP T1 productivity module as part of their ten-day SOS training. To do this, we estimate a simple OLS regression of the form:

$$\Delta test\ score_{i,j} = \beta_0 + \beta_1 T1_i + \tau_i + \mathbf{X}_j\boldsymbol{\beta} + \epsilon_{i,j} \quad (1)$$

where  $\Delta test\ score_{i,j}^{indiv}$  is the difference between the post- and pre-training test scores for individual  $i$  in division  $j$  for the ten-day individual training, measured according to the share of questions in each module answered correctly,  $T1_i$  is an indicator variable equal to 1 if the individual was randomized into the TFP productivity module and 0 if the individual was randomized into the SOS productivity module,  $\tau_i$  is a fixed effect for the question set taken by the individual (see Appendix C.2 for details),  $\mathbf{X}_j$  is a vector of de-meaned division-level controls including dummies for the type of division (using the Civil Service’s classification of divisions) and an indicator for whether the division sent high-ranking individuals for training, (type of division, number of workers in division attending some training, presence of high/low payscale workers) controls, and  $\epsilon_{i,j}$  is an error term.

Given this, the constant  $\beta_0$  represents the mean difference in scores for individuals who took the SOS productivity module and  $\beta_0 + \beta_1$  represents the mean difference in scores for individuals who took the TFP productivity module, so  $\beta_1$  represents the average difference in outcomes between

---

<sup>9</sup>This differs from the standard WMS approach (which we use in our endline survey), in which respondents are asked to describe the actual practice used in their organization through open-ended questions and the enumerator then scores the organization’s practice on a 1-5 scale. In this closed-ended adaptation, respondents are instead given three short descriptions of management practices on a particular aspect of management (each corresponding to the bad, average, and good practice descriptions associated with one of the questions in the standard WMS coding grid), and asked to select which they think represents the best practice.

the TFP and SOS treatment groups. Because training timing was staggered across a twelve-month period, in some cases individuals had undergone their team-level T2 training prior to undertaking their individual-level SOS/T1 training; in order to avoid confounding of the treatment effects, we exclude such individuals and restrict the sample to the set of individuals who undertook their SOS/TFP T1 training prior to undertaking team training or who did not undertake T2 at all. Standard errors are clustered at the division level.

To estimate the impact of T2, we estimate a simpler specification that simply measures the extent of trainees' learning gain from the pre- to post-tests:

$$\Delta test\ score_{i,j} = \beta_0 + \tau_i + \mathbf{X}_j\boldsymbol{\beta} + \epsilon_{i,j} \quad (2)$$

where all variables are as in Equation 1, and the constant  $\beta_0$  is the coefficient of interest representing learning gain from pre- to post-test, which can be compared in magnitudes to the effects estimated by  $\beta_0$  and  $\beta_1$  in Equation 1. Analogously to Equation 1, our estimation sample is the set of individuals who took TFP T2 prior to undertaking either SOS or TFP T1.

**Table 3: Short-term learning gain and changed perceptions of good management practice**

	(1)	(2)	(3)	(4)	(5)	(6)
	Training material learning gain			Perceptions of good management		
	TFP (T1 & T2) productivity modules	SOS productivity modules	Non- productivity modules	Incentives/ Monitoring	Autonomy/ Discretion	Full Management Index
<i>Panel (a): Learning from individual trainings (TFP T1 vs SOS)</i>						
Mean learning, status quo training (SOS)	0.104*** (0.026)	0.272*** (0.033)	0.085*** (0.027)	-0.046 (0.046)	0.036 (0.041)	-0.005 (0.032)
Difference in learning, T1 (individual) - SOS	0.075* (0.042)	-0.043 (0.049)	-0.045 (0.043)	0.084 (0.062)	0.119* (0.069)	0.102** (0.047)
R <sup>2</sup>	0.364	0.447	0.282	0.024	0.271	0.164
N	241	241	241	241	241	241
<i>Panel (b): Learning from team training (TFP T2)</i>						
Learning (T2, team-based)	0.128*** (0.017)	0.139*** (0.020)	-0.023 (0.035)	-0.078** (0.034)	0.055* (0.028)	-0.012 (0.020)
R <sup>2</sup>	0.365	0.379	0.036	0.078	0.359	0.099
N	475	475	475	475	475	475

\* p<0.10, \*\* p<0.05, \*\*\* p<0.01

Notes: Standard errors clustered at the division level in parenthesis. Dependent variable is individual training learning gains or changes in perceptions of good management practices. Sample for Panel (a) is individuals who either had no team training or who had team training after individual training; Panel (b) is individuals who either had no individual training or who had individual training after team training. Learning gain is evaluated as the change in the share of questions answered correctly in each module from the pre-training to post-training tests. Test scores are re-scaled so that maximum feasible score in each SOS and TFP test score component is the same as the maximum feasible score in each T2 component. Results control for division characteristics (dummies for type of division, dummy stating that the division sent for training workers with high paygrade), and for the type of exam taken. See Appendix C.2 for details of productivity and management knowledge questions.

Panel (a) of Table 3 shows the results. Column 1 shows that while SOS trainees showed some learning on the TFP module content (as there was some overlap in concepts), TFP T1 trainees improved their test scores on this module by 7.5 percentage points more on average. This is a quantitatively large improvement relative to the individual training entry exams, where officials, on average, answered 35.0 percent of the TFP productivity module correctly. The Column 2 shows that SOS trainees also achieved significant improvements of 27.2 percentage points on the SOS-related content, but TFP T1 trainees did not perform significantly worse on the SOS-related content. Column 3 shows that SOS and TFP T1 trainees both experienced small and statistically indistinguishable improvements in non-productivity-related test scores, as would be expected since they undertook these modules together during the other nine days of the ten-day training. Overall, these results suggest that learning from the TFP training improved knowledge across a greater breadth of productivity-related issues than SOS training.

Columns 4-6 of Table 3 Panel (a) then show that while SOS productivity training has no significant effect on trainees' perceptions of good management practice, TFP T1 trainees were more likely to answer these management questions "correctly" - that is, according to the commonly understood definition of good management captured within the WMS coding grid - across both the monitoring/incentives and autonomy/discretion modules, and significantly so for the monitoring/incentives sub-index and the full management practice index.<sup>10</sup> This is despite the fact that the TFP training module did not explicitly discuss either monitoring/incentives or autonomy/discretion, or target changes in them (although there is a natural relationship between autonomy/discretion-related practices and the type of bottom-up work process innovations encouraged in the TFP module). These results suggest that the TFP T1 training content not only helped workers learn more about the abstract productivity concepts and skills which they were being directly taught, but also improved their understanding of what good management process look like.

Panel (b) of Table 3 examines the impacts of the TFP T2 team training on these outcomes. Columns 1-3 show that delivering the TFP training module in a team setting (T2) leads to similar and statistically significant patterns of productivity-related learning as delivering it in the normal individual-based format through which CSTC normally delivers trainings, but with magnitudes about half as large. However, Columns 4-6 show that TFP T2 leads to a slightly different pattern of change in perceptions of good management practice: T2 trainees score lower on the monitoring/incentives module, while scoring higher on the autonomy/discretion module, leading to unchanged learning on the full management index.

Overall, it thus seems that the TFP module content was successful in the sense that it shifted both productivity knowledge and management knowledge, without sacrificing knowledge of the productivity concepts being taught in the existing SOS training.<sup>11</sup> However, the individual-based

---

<sup>10</sup>While we cannot say definitively whether these changed perceptions are "good" in the sense that they would lead to higher performance if they were to be implemented, the improvement in the autonomy/discretion index is consistent with the overall thrust of what the TFP training module aimed to achieve, and actual use of autonomy/discretion-related practices has been found to be positively associated with task completion in Rasul and Rogger (2018) and Rasul *et al* (2019). We do not assume that higher management scores are necessarily associated with better outcomes, and indeed Rasul *et al* (2019) show that higher monitoring and incentives scores at organizational level have a partial negative association with task completion rates (and higher autonomy and discretion scores have a partial positive associated with task completion rates).

<sup>11</sup>We find little evidence of complementarities between individual- and team-based trainings, for individuals who

training approach of T1 (with individuals trained in cohorts of officers from many different organizations) led to greater learning gains than the team-based training approach of T2 (in which individuals were trained in cohorts of officers from their own division, who they work with on a day-to-day basis).

## 5.2 Learning Persistence and Training Follow-up

To understand the persistence of learning and other direct follow-up from training, we estimate individual-level regressions of the form:

$$y_{i,j} = \beta_0 + \beta_1 T1_i + \beta_2 T2_j + \mathbf{X}_j \boldsymbol{\beta} + \epsilon_{i,j} \quad (3)$$

where  $y_{i,j}$  is one of several indices of learning persistence and training follow-up collected in the endline survey;  $T1_i$  and  $T2_j$  are indicator variables capturing whether the individual participated in TFP T1 and their division participated in TFP T2, respectively;  $\mathbf{X}_j$  is the same vector of division-level controls as in the previous equations; and  $\epsilon_{i,j}$  is an error term. We restrict the sample to individuals who appear both in our training and endline datasets, and who responded in the endline survey that they attended at least one of the relevant trainings and thus were tracked into responding to the training follow-up questions. The parameters  $\beta_1$  and  $\beta_2$  thus capture the partial impacts of the two TFP treatment arms, relative to the base group of officers who received only the standard SOS productivity training. Standard errors are clustered at division level.

Columns 1 and 2 of Table 4 show that learning persistence was higher among TFP T1 trainees than SOS trainees, on the SOS-related productivity concepts as well as the TFP curriculum itself. Officers who participated in team-based TFP T2 also maintained greater knowledge on both sets of questions, albeit with a lesser magnitude and statistical significance. Thus the short-term learning gains shown in Table 3 seem to persist 6-18 months after training.

---

take both trainings (results not shown for brevity, but available upon request).

**Table 4: Learning persistence and training follow-up**

	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Learning persistence</i>		<i>Training follow-up</i>			
	TFP-related (T1 & T2) productivity knowledge	SOS-related productivity knowledge	Index: obstacles to applying learning (indiv. training only)	Index: steps taken to implement action plan	Index: perceived obstacles to action plan implementation	Index: overall training follow- up
<b>T1 (individual)</b>	0.086** (0.036)	0.079** (0.037)	0.444*** (0.163)	-0.098 (0.139)	0.165 (0.130)	0.245** (0.118)
<b>T2 (team)</b>	0.044* (0.025)	0.029 (0.026)	- -	0.072 (0.105)	-0.009 (0.101)	0.098 (0.098)
<b>Controls</b>	Yes	Yes	Yes	Yes	Yes	Yes
<b>R<sup>2</sup></b>	0.029	0.028	0.027	0.057	0.067	0.084
<b>N</b>	624	624	135	462	331	462

\* p<0.10, \*\* p<0.05, \*\*\* p<0.01

Notes: Standard errors clustered at the division level in parenthesis. Sample includes all officials who participated in any training session and could be matched to the endline survey; sample size changes across columns due to survey skip patterns that mean not all respondents are asked every question. Results control for division characteristics (dummies for type of division, dummy stating that the division sent for training workers with high paygrade).

Columns 3-6 then examine officers' follow-up on the training, using a battery of nine questions with 18 potential outcomes (see Table A7, Appendix C.1). For brevity and power, we aggregate these outcomes into four normalized indices following Anderson (2008); on each index, positive outcomes indicate outcomes that should have been more likely if the TFP module were more useful and practical than the status quo productivity module. These indices capture: for attendees of the ten-day training, whether trainees were less likely to cite lack of training material relevance or lack of support from their manager or colleagues as obstacles to applying lessons from training (Column 3); whether trainees were more likely to say they discussed their action plan with superiors or colleagues, or set up a team or did additional feasibility research to implement their action plan (Column 4); whether trainees were less likely to cite as obstacles to implementing their action plan that the idea was not a good one, that they lacked resources (since TFP focused on identifying cost-free improvements), or that their manager or division were not supportive of the new idea (Column 5); and an overall index that captures these three follow-up sub-indices as well as a set of other questions on the usefulness of the productivity training and its application in their work (Column 6).

We find strong positive results of TFP T1 on avoiding obstacles related to applying learning and on the overall training follow-up index, but null effects on the other sub-indices and for TFP T2. While we cannot reject equality of coefficients between the two treatments, these longer-term results are nonetheless consistent with the stronger short-term effects seen in the individual- rather than team-based approach to delivering the TFP productivity module.

Together, our findings imply that the new training intervention was effective at inducing long-term learning and attitudinal change, with larger and more persistent learning effects when it was targeted at individuals rather than work teams as a whole. Our results indicate that while trainees were not necessarily more likely to implement the exact action plan they envisioned during the training than participants in the status quo productivity training, at a minimum the training was successful in shifting trainees' attitudes and perceptions of how to make change within their workplace.

### 5.3 Impacts on Organizational Culture

We now turn to understanding how the trainings affected division-level measures of norms related to productivity and innovation, management quality, and performance. Throughout this subsection, we restrict our sample to the sub-set of divisions that had at least one member attend the ten-day SOS training and were thus eligible for randomization into TFP T1 or T2.

We estimate these division-level impacts using a specification of the form:

$$y_j = \beta_1 T1_j + \beta_2 T2_j + \mathbf{SOS}_j \boldsymbol{\beta} + \epsilon_j \quad (4)$$

where  $y_j$  is a measure of organizational culture for division  $j$ ,  $T1_j$  is a binary treatment indicator that equals 1 if the division had any member undertake the TFP T1 training,  $T2_j$  is a binary treatment indicator that equals 1 if the division undertook the TFP T2 training,  $\mathbf{SOS}_j \boldsymbol{\beta}$  is a vector of indicator variables controlling flexibly for the number of division members who participated in the ten-day SOS training, and  $\epsilon_j$  is an error term. Since the treatments are well-balanced we

omit the division-level controls used in sections 4.1 and 4.2 as they slightly reduce the precision of estimates; including them produces results of similar magnitude and significance.<sup>12</sup> We calculate robust standard errors throughout.

**Table 5: Effects on organizational culture**

	(1)	(2)	(3)	(4)	(5)
		Selected culture sub-indices (Sexton <i>et al</i> 2006)		Additional culture sub-indices	
	Aggregate Culture	Teamwork climate	Performance climate	Fostering new ideas	Relative Performance
<b>T1 (individual)</b>	0.454** (0.192)	0.599*** (0.188)	0.431** (0.194)	0.389** (0.187)	0.452** (0.191)
<b>T2 (team)</b>	0.004 (0.190)	0.056 (0.191)	0.117 (0.189)	-0.187 (0.198)	0.107 (0.192)
<b>Control for number of officials in indiv. training</b>	Yes	Yes	Yes	Yes	Yes
<b>R<sup>2</sup></b>	0.080	0.093	0.070	0.079	0.108
<b>N</b>	148	148	148	148	148

\* p<0.10, \*\* p<0.05, \*\*\* p<0.01

Notes: Robust standard errors in parenthesis. All specifications include a set of indicator variables flexibly controlling for the number of officers from the division that attended the ten-day SOS training (in either standard SOS or T1 productivity modules). Sample is all divisions with at least one member attending the ten-day individual-level SOS training.

Table 5 examines the effects on organizational culture, in aggregate and for selected sub-indices (see Section 4 and Appendix C.1 for details of index construction). Column 1 shows that TFP T1 leads to an improvement in the overall organizational culture index of 0.454 standard deviations. This is a substantial improvement, especially since this index represents the average perception of culture by all staff in the division, not just those who participated in a training. Columns 2-5 show that this aggregate improvement is driven by similarly large increases in the sub-indices of *teamwork climate*, *performance climate*, *fostering new ideas*, and *relative performance*, which one would expect to be affected by the TFP training and its emphasis on innovation, work process improvement, and teamwork. The improvements in these dimensions are matched by null effects (not shown, for brevity) on other sub-indices of organizational culture (*stress recognition*,

<sup>12</sup>We use binary divisional treatment indicators rather than the share of the division enrolled in each training because the combination of attrition in matching individuals across datasets and the unavailability of definitive division size measures (see Online Appendices C and E) make calculating share variables highly susceptible to measurement error.



*perceptions of management, working conditions*) that are part of the adapted Sexton *et al* index but are less directly targeted by the TFP curriculum.

In line with the more muted impacts on learning and action follow up, the team-based TFP T2 training has no significant impacts either on the overall index or any sub-indices. This provides direct evidence that the intervention targeted at individual reform leaders within work teams (as opposed to at all members of the team collectively) was more effective at changing culture in this context.

## 5.4 Impacts on Performance and Management

We next ask whether these trainings led to improvements in performance as well as improvements in organizational culture. We examine this question using three division-level measures of performance: the quality of administrative process, task completion, and management quality. Though we cannot distinguish whether these effects arise directly from the trainings or indirectly through the changes in culture we document in the last section, our question is whether culture-change oriented training has wider effects on team productivity.

We begin with our measures of process quality and task completion, which were each coded from separate administrative data sources. We estimate the following specification for the  $k^{th}$  task (for task completion) or file (for the quality of process) in division  $j$ :

$$y_{j,k} = \beta_1 T1_j + \beta_2 T2_j + \mathbf{SOS}_j \boldsymbol{\beta} + \mathbf{Z}_k \boldsymbol{\gamma} + \epsilon_{j,k} \quad (5)$$

where  $y_{j,k}$  is an outcome variable related to the quality of process or task completion,  $\beta_1$  and  $\beta_2$  are the coefficients of interest,  $\mathbf{Z}_k$  is a vector of task or file controls, and other variables are the same as in Equation 4.<sup>13</sup> We inverse-weight observations by the number of tasks or files observed

---

<sup>13</sup>For the quality of administrative process, the file controls capture: whether the assessor stated they had access to all the information needed to assess the file; a dummy stating that there were no challenges encountered in judging the quality of the file; the log number of files in the division; dummies for the file assessors; and the day on which the assessment was done. For task completion, controls are: a dummy stating whether the task is a one-off task or a periodic/regular task; a dummy stating whether the reported task is a single task, or a bundle of tasks; a dummy stating whether the task requires coordination with stakeholders outside the organization; and the log

for each division, so that each division receives the same weight (as is implicitly so for the estimates on organizational culture and management quality), and estimate robust standard errors.

**Table 6: Training impacts on performance and management quality**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Administrative process quality			Task completion	Management quality			Combined index
	Quality of procedure	Adherence to procedure	Quality of content	Full completion (binary)	Full management index	Autonomy/discretion sub-index	Incentives/monitoring sub-index	
<b>T1 (individual)</b>	0.298*	0.335*	0.138	0.110*	0.235	0.269	0.238	0.420**
	(0.153)	(0.169)	(0.179)	(0.060)	(0.190)	(0.183)	(0.196)	(0.201)
<b>T2 (team)</b>	0.160	0.136	0.168	-0.039	0.061	0.150	-0.066	0.263
	(0.130)	(0.176)	(0.187)	(0.071)	(0.197)	(0.189)	(0.196)	(0.184)
<b>Control for number of officials in indiv. training</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<b>R<sup>2</sup></b>	0.459	0.361	0.306	0.079	0.083	0.118	0.071	0.118
<b>Number of divisions (clusters)</b>	286 (106)	286 (106)	286 (106)	627 (84)	148 (n/a)	148 (n/a)	148 (n/a)	152 (n/a)

\* p<0.10, \*\* p<0.05, \*\*\* p<0.01

Notes: Robust standard errors in parenthesis. All specifications include a set of indicator variables flexibly controlling for the number of officers from the division that attended the ten-day SOS training (in either standard SOS or T1 productivity modules). Sample is all divisions with at least one member attending the ten-day individual-level SOS training, and with non-missing data for dependent variable. Sample N is number of number of files for columns 1-3, number of tasks for column 4, and divisions for Columns 5-8; number of divisions (clusters) given in parentheses for Columns 1-4. Columns 1-4 are inverse-weighted by the number of observations (files or tasks) per cluster (division). Columns 1-3 include controls for file characteristics (whether the assessor stated they had access to all the information needed to assess the file; a dummy stating that there were no challenges encountered in judging the quality of the file; the log number of files in the division; dummies for the file assessors; and the day on which the assessment was done). Column 4 includes controls for task characteristics (a dummy stating whether the task is a one-off task or a periodic/regular task; a dummy stating whether the reported task is a single task, or a bundle of tasks; a dummy stating whether the task requires coordination with stakeholders outside the organization; and the log number of tasks in the division). Column 8 is a combined division-level index of management, task completion, and process efficiency, compiled following Anderson (2008) using the following components: autonomy/discretion management sub-index; incentives/monitoring management sub-index; other management sub-index; inverse-weighted task completion; inverse-weighted quality of process procedure; and inverse-weighted quality of process content.

Column 1 estimates that a division being exposed to the TFP T1 training leads to an increase of 0.206 standard deviations in the quality of procedure with which audited files were handled ( $p = 0.054$ ). In particular, we find that the TFP T1 training’s strongest effect is on the ‘Adherence to procedure’ sub-component of procedure quality (shown in Column 2), which is consistent with our observation that most trainees’ action plans are focused on implementing or expanding the scope of practices that already exist on paper and that they think would be beneficial if implemented properly (section 3.2).

Interestingly, Column 3 indicates that improved changes in the quality of *procedure* do not necessarily imply changes in the quality of the *content* of the files. We do not see corresponding number of tasks in the division.

impacts of TFP trainees efforts on the content of government activities. As described above, the content indicator relates to the extent to which files had a clear background, course of action, allocation of responsibilities and so on.

As with our culture measures, exposure to team-level T2 training has no significant impacts on the quality of administrative process. Thus, our results indicate that TFP training targeted at individual public officials generated improvements in the areas of work the trainees felt were most strategically in need of reform. The same training curriculum targeted at the division as a whole did not have subsequent impacts on administrative process quality.

To what extent does culture-change oriented training lead to improvements in division-level task completion? Column 4 estimates the impact on a binary dependent variable indicating whether the task was fully completed (i.e. the maximum completion status of 5) or not, and finds that exposure to the TFP T1 training increases the probability of a task being fully completed by 11.0 percentage points - a large effect magnitude, given that the average full completion rate is just 15.2 percentage points. This effect is significant at the 10 percent level ( $p = 0.071$ ). We find no significant effect of TFP T2 on task completion.

Finally, we estimate the impacts of the training programs on management quality at division level, using the same specification as we did for organizational culture (Equation 4) but with the World Management Survey-style management quality indices from the endline survey as dependent variables. Columns 5-7 of Table 6 show that while TFP T1 has substantively large and positive estimated effects on the overall management quality index and the incentives/monitoring and autonomy/discretion sub-indices, these are imprecisely estimated and thus not statistically significant. Note that the TFP training content and action plans did not target division-level management practices explicitly, so these positive effects are secondary implications of the wider shifts in work team culture and procedures.

Across the different management and performance outcomes in Table 6, a clear and consistent pattern emerges: the individual-based TFP T1 is associated with improvements in performance, while the team-based TFP T2 has little effect. However, due to the limited number of divisions in our sample (which is in turn constrained by the total number of divisions in Ghana's Civil

Service), our ability to detect statistically significant effects across these measures is limited. We therefore combine the three underlying constructs (the quality of administrative process, task completion, and management quality) into a single index following Anderson (2008), and estimate the impacts of training on this combined division-level index.<sup>14</sup> The results in Column 8 show that the impact of training on this combined index of management quality and performance is positive and statistically significant ( $p = 0.038$ ), implying that treatment by TFP T1 improves a division's performance index by 0.420 standard deviations. This increase in the overall performance index is of similar magnitude to the increase in the aggregate organizational culture index (0.454 standard deviations) we see as a result of the T1 training.

This pattern of the training's impacts on performance is consistent with the observed impacts on short-term learning (Table 3), long-term learning persistence and follow-up (Table 4), and long-term organizational culture change (Table 5). Taken together, these results suggest that the observed improvements in organizational culture catalyzed by the TFP T1 training also resulted in tangible improvements in management quality and performance.

One question raised by our results is why the team-based T2 training had no long-term impacts. One possibility is that the weaker initial learning gains we observe led to effects that faded out more rapidly or were too small for us to detect. It is also possible that conducting the training in the presence of an entire work team, including its leaders, led to a reinforcement of existing cultures. This could have occurred if the presence of team leaders inhibited open discussion of issues and potential improvements. Team-based training could also have been less effective due to free-riding problems, if each individual perceived that they had little incentive to take follow-up actions, whereas this responsibility was clear for the individual-based training arm. While our results do not allow us to distinguish empirically between these potential explanations, these possibilities illustrate the nuances and complexities facing culture-change interventions and

---

<sup>14</sup>We compile the index from six underlying components: inverse-weighted quality of process procedure at division-level; inverse-weighted quality of process content at division-level; inverse-weighted average task completion at division-level; the autonomy/discretion management sub-index; incentives/monitoring management sub-index; and the "other" management sub-index. The weights on each sub-component in the combined index are defined by the Anderson (2008) procedure. Alternative approaches to compiling these indices, such as aggregating the administrative process quality and management sub-indices into overall averages prior to including them in the Anderson index or residualizing task completion and process quality on task/file controls prior to including them in the index, produce similar results.

represent an important set of questions for future research.

## 6 Conclusion

This paper evaluated the impact of rolling out an applied productivity and innovation training module among mid-level bureaucrats in Ghana’s Civil Service, across two treatment arms that vary the cohort with whom training is undertaken: a standard, classroom-style, individual-focused approach, and a more innovative team-based approach. We found that both treatment arms led to short-term improvements in learning and management knowledge (relative to the status quo training), and that for the individual-focused treatment arm these gains persisted 6-18 months after the training, were associated with a range of follow-up actions, and led to improvements in organizational culture and performance in the trainees’ work teams. This provides evidence that culture change might be more effectively achieved (at least in our context) by focusing interventions on individuals who can serve as reform leaders, rather than by attempting to simultaneously induce change among an entire team.

Our findings provide among the first experimental evidence on the potential for large-scale, bureaucrat-led change to improve bureaucratic performance. This complements the growing body of observational and experimental evidence that greater autonomy and discretion for mid-level bureaucrats is often associated with improvements in performance and service delivery, even in countries often associated with more challenging governance contexts (Rasul and Rogger 2018; Rasul *et al* 2019; Bandiera *et al* 2020). While many interventions and reforms try to improve performance by increasing control over bureaucrats, our paper contributes to the increasing evidence that better performance can also be achieved by trying to empower (rather than constrain) bureaucrats, even in challenging contexts. A further contribution of this paper has been to adapt to public sector contexts or introduce novel measures of bureaucratic culture, bureaucratic voice, and performance that may be useful to other scholars of bureaucracy in political science, public administration, and other disciplines.

From a policy perspective, these results provide scarce evidence for the potential effectiveness of at least some public sector training programs. The fact that a light-touch, one-day training

module delivered at scale through the Ghana Civil Service's existing training institution can catalyze short- and long-term improvements both in individual learning and team culture and performance is particularly impressive. Delivering the training through the Civil Service's own training mechanisms may also make its take-up, sustainability, and scaling of this intervention more straightforward than externally driven interventions to improve bureaucratic performance. However, our results also suggest that not all training has substantial long-term impacts, so further research is needed to understand the mechanisms underlining training's effectiveness and the generalizability of these findings.

## References

- Alchian, A. A., Demsetz, H. 1972. “Production, information costs, and economic organization”. *American Economic Review* 62(5), 777-795.
- An, Seung-Ho, Kenneth J. Meier, Anne Bøllingtoft, and Lotte Bøgh Andersen. 2019. “Employee Perceived Effect of Leadership Training: Comparing Public and Private Organizations.” *International Public Management Journal* 22(1): 2–28.
- Andersson, Ola, Marieke Huysentruyt, Topi Miettinen, and Ute Stephan. 2017. “Person-Organization Fit and Incentives: A Causal Test.” *Management Science* 63(1): 73-96.
- Andersen, L.B., B. Bjornholt, L. Ladegaardbro, and C. Holm-Petersen. 2018. “Achieving High Quality Through Transformational Leadership: A Qualitative Multilevel Analysis of Transformational Leadership and Perceived Professional Quality,” *Public Personnel Management* 47(1): 51-72.
- Andrews, Matt, Lant Pritchett, and Michael Woolcock. 2017. *Building State Capability: Evidence, Analysis, Action*. Oxford: Oxford University Press.
- Ashford, S., Sutcliffe, K., Christianson, M. (2009). Speaking up and speaking out: The leadership dynamics of voice in organizations. In J. Greenberg M. Edwards (Eds.), *Voice and silence in organizations* (pp. 175–202). Bingley: Emerald Group Publishing Ltd.
- Ashraf, Nava, Oriana Bandiera, and Kelsey Jack. (2014). “No Margin, No Mission? A Field Experiment on Incentives for Pro-social Tasks.” *Journal of Public Economics* 120: 1-17.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2009. “Social connections and incentives in the workplace: Evidence from personnel data.” *Econometrica* 77(4): 1047–1094.
- Bandiera, O., Best, M. C., Khan, A. Q., Prat, A. (2020). “The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats”. National Bureau of Economic Research Working Paper 26733.
- 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>.
- Barney, Jay B. (1986). “Organizational Culture: Can It Be a Source of Sustained Competitive Advantage?” *The Academy of Management Review* 11(3): 656-665
- Baron, James N., and David M. Kreps. (2013). “Employment as an Economic and a Social Relationship.” In *The Handbook of Organizational Economics*, Robert Gibbons and John Roberts (eds), Princeton: Princeton University Press.
- Baser, Heather, and Peter Morgan. 2008. “Capacity, Change and Performance.” Discussion Paper 59B, ECDPM.
- Bénabou, R. Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5), 1652-1678.
- Bernard, Tanguy, Stefan Dercon, Kate Orkin, and Alemayehu Seyoum Taffesse (2017). “Will Video Kill the Radio Star? Assessing the Potential of Targeted Exposure to Role Models through Video.” *World Bank Economic Review* 29(suppl 1): S226-S237.
- Best, Michael, Jonas Hjort, and David Szakonyi (2018). “Individuals and Organizations as Sources of State Effectiveness.” Mimeo, December.
- Blader, Steven, Claudine Gartenberg, and Andrea Prat. 2019. “The Contingent Effect of Management Practices.” *Review of Economic Studies* 87(2): 721–749.
- Bloom, Nicholas, and John Van Reenen. (2007). “Measuring and Explaining Management Practices Across Firms and Countries.” *Quarterly Journal of Economics* 122: 1351-408.
- Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. (2012). “The Organization of Firms Across Countries,” *Quarterly Journal of Economics* 127: 1663-705.
- Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2014. “Does Management Matter in Schools.” NBER Working Paper 20667.



- Chassang, Sylvain. 2010. "Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts." *American Economic Review* 100(1): 448-465.
- Detert, J., Burris, E. (2007). Leadership behavior and employee voice: Is the door really open? *Academy of Management Journal*, 50(4), 869–884.
- DiMaggio, Paul. (1997). "Culture and Cognition." *Annual Review of Sociology* 23: 263-287.
- Dixit, Avinash. 2002. "Incentives and Organizations in the Public Sector: An Interpretative Review." *The Journal of Human Resources*, Vol. 37, No. 4 (Autumn, 2002), pp. 696-727.
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383.
- Finan, Frederico, Benjamin Olken, and Rohini Pande. 2017. "The Personnel Economics of the State." In Banerjee, Abhijit Vinayak, and Esther Duflo (eds) *Handbook of Economic Field Experiments*, 467-514.
- Gartenberg, Claudine, Andrea Prat, and George Serafeim. 2019. "Corporate Purpose and Financial Performance." *Organization Science* 30(1): 1-18.
- Gibbons, Robert, and Rebecca Henderson. 2012. "Relational Contracts and Organizational Capabilities." *Organization Science* 23(5): 1350-1364.
- Gibbons, Robert, and Rebecca Henderson. 2013. "What Do Managers Do?" In *The Handbook of Organizational Economics*, Robert Gibbons and John Roberts (eds), Princeton: Princeton University Press.
- Graham, John R., Campbell R. Harvey, Jillian Popadak, and Shivaram Rajgopal. (2017). "Corporate Culture: Evidence from the Field." *NBER Working Paper* 23255.
- Grindle, Merilee. 1997. "Divergent cultures? When public organizations perform well in developing countries." *World Development* 25: 481-95.
- Guadalupe, Maria, Zoe Kinias, and Florian Schlotter. (2020). "Individual Identity and Organizational Identification: Evidence from a Field Experiment." *AER Papers and Proceedings* 110:

193-198.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. (2015). "The Value of Corporate Culture." *Journal of Financial Economics* 117: 60-76.

Hansen, Jesper Asring, and Lars Tummers. (2020). "A Systematic Review of Field Experiments in Public Administration." *Public Administration Review*.

Hart, Oliver, and John Moore. 2008. "Contracts as Reference Points." *Quarterly Journal of Economics* 123(1): 1-48.

Hoffer Gittell, J. 2002. *The Southwest Airlines Way*. McGraw-Hill, New York.

Holmstrom, Bengt. 1982. "Moral Hazard in Teams." *The Bell Journal of Economics* 13(2): 324-340.

Holmstrom, Bengt, and Paul Milgrom (1991). "Multitask principal-agent analyses: incentive contracts, asset ownership, and job design." *Journal of Law, Economics and Organization* 7: 24-52.

Jakobsen, M., C. Bøtcher Jacobsen, and S. Serritzlew. 2019. "Managing the Behavior of Public Frontline Employees through Change-Oriented Training: Evidence from a Randomized Field Experiment," *Journal of Public Administration Research and Theory*: 1-16.

Jensen, Ulrich T. 2018. "Does Perceived Societal Impact Moderate the Effect of Transformational Leadership on Value Congruence? Evidence from a Field Experiment." *Public Administration Review* 78(1): 48-57.

Khan, Adnan Q., Asim I. Khwaja, and Benjamin Olken. (2015). "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *Quarterly Journal of Economics* 131: 219-71.

Leaver, Clare, Owen Ozier, Pieter Serneels, and Andrew Zeitlin (2019). "Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools". Mimeo, June.

- Leonard, David. 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.
- Linos, Elizabeth. 2018. More Than Public Service: A Field Experiment on Job Advertisements and Diversity in the Police. *Journal of Public Administration Research And Theory*, 2018, 67–85.
- Madsen, Peter, Vinit Desai, Karlene Roberts, and Daniel Wong. (2006). “Mitigating Hazards Through Continuing Design: The Birth and Evolution of a Pediatric Intensive Care Unit.” *Organization Science* 17(2): 239-248.
- Martinez, Elizabeth A., Nancy Beaulieu, Robert Gibbons, Peter Pronovost, and Thomas Wang. 2015. Organizational Culture and Performance. *American Economic Review: Papers Proceedings* 2015, 105(5): 331–335
- McDonnell, Erin. (2017) “Patchwork Leviathan: How Pockets of Bureaucratic Governance Flourish within Institutionally Diverse Developing States.” *American Sociological Review* 82(3): 476-510.
- McDonnell, Erin. 2020. *Patchwork Leviathan: Pockets of Bureaucratic Effectiveness in Developing States*. Princeton: Princeton University Press.
- McKenzie, David. (2020). “Small Business Training to Improve Management Practices in Developing Countries: Reassessing the Evidence for “Training Doesn’t Work’ ”. World Bank Research Working Paper 9408.
- McKenzie, David, and Christopher Woodruff. (2014). “What are we learning from business training evaluations around the developing world?”, *World Bank Research Observer* 29(1): 48-82.
- 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.
- Perry, J. (1996). Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory* 1, 5-22.

- Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency." *Journal of Political Economy* 111(5): 929-958.
- Rasul, Imran, and Daniel Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *Economic Journal* 128(608): 413-446.
- Rasul, Imran, Daniel Rogger, and Martin J. Williams. 2019. "Management, Organizational Performance, and Task Clarity: Evidence from Ghana's Civil Service" . *Journal of Public Administration Research and Theory*.
- Riley, William, Stanley Davis, Kristi Miller, Helen Hansen, Francois Sainfort, and Robert Sweet. (2011). "Didactic and Simulation Nontechnical Skills Team Training to Improve Perinatal Patient Outcomes in a Community Hospital." *The Joint Commission Journal on Quality and Patient Safety* 37(8): 357-364.
- Roll, Michael (ed). 2014. *The Politics of Public Sector Performance: Pockets of Effectiveness in Developing Countries*. Abingdon: Routledge.
- Rousseau, D. M. 1995. *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements*. Sage, London.
- Schein, Edgar H. 1985. *Organizational Culture and Leadership*. San Francisco: Jossey-Bass.
- Sexton, John B., Robert L. Helmreich, Torsten B. Neilands, Kathy Rowan, Keryn Vella, James Boyden, Peter R. Roberts, and Eric J. Thomas. (2006). "The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research." *BMC Health Services Research* 6:44.
- Srivastava, Sameer B., Amir Goldberg, V. Govind Manian, and Christopher Potts. (2018). "Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations." *Management Science* 64(3):1348-1364.
- Tendler, Judith. 1997. *Good government in the tropics*. Baltimore: Johns Hopkins University Press.

- Thomas, Eric J., J. Bryan Sexton, Torsten B. Neilands, Allan Frankel, and Robert L. Helmreich. (2005). "The effect of executive walk rounds on nurse safety climate attitudes: A randomized trial of clinical units." *BMC Health Services Research* 5: 28.
- Van Maanen, J. and E. H. Schein (1979). "Toward of Theory of Organizational Socialization." *Research in Organizational Behavior*, 1: 209-264.
- Weaver, Sallie J., Lisa H. Lubomski, Renee F. Wilson, Elizabeth R. Pfoh, Kathryn A. Martinez, and Sydney M. Dy. (2013). "Promoting a Culture of Safety as a Patient Safety Strategy: A Systematic Review." *Annals of Internal Medicine* 158: 369-374.
- Wilson, James. 1989. *Bureaucracy*. New York: Basic Books.

# A Online Appendix: Training Details

## A.1 Training Delivery and Randomization

The TFP training content was initially drafted by an international consultant based on guidelines given by CSTC and the research team, and was then adapted and customized by CSTC trainers at a training-of-trainers workshop to improve its fit with the Ghanaian context. CSTC trainers then made minor adaptations to improve the flow of the training after a pilot training in February 2017, and again in June 2017. Slides and materials for the TFP training module used standard CSTC template and formatting, so were not visually distinguishable from the existing productivity training material developed by CSTC trainers.

The TFP T1 trainings at CSTC were delivered from February to December 2017. There were 17 ten-day Scheme of Service (SOS) trainings delivered during this period, out of which 11 randomized the content of the one day of productivity training; for six of the trainings, CSTC's classrooms were all fully occupied and so it was not possible to split the training cohort into two in order to deliver both the standard SOS and TFP T1 curricula. For these six trainings, the standard SOS productivity curriculum was delivered. Trainees attended all the ten training days together except for the productivity day. Trainees came from different organizations, but they attended the SOS training together with civil servants from the same grade (i.e., Deputy Director, Assistant Director I, Assistant Director 2B, and Assistant Director 2B, secretarial and sub-professional grades). On the day of the productivity module, the class was split up, and trainees were randomly chosen with 50 percent probability to attend either the new TFP productivity module or the standard SOS productivity module. The randomization was done at the start of the day using the attendance dataset, which collects information on trainees' daily attendance, so trainees were simply told to go to two different classrooms for the productivity module.

Table A1 shows the treatment distribution. In total, 527 civil servants from 47 organizations attended the SOS training at the CSTC training center in Accra. Among them, 31 officials were from junior grades or not civil servants (and thus outside of our study sample), and 74 came from organizations or units that were not part of the study sample (i.e. were excluded from

the baseline and endline surveys) for various reasons. Excluding these individuals leaves a total of 422 senior civil servants in our study sample who showed up to SOS trainings. Of these, 139 (from 90 divisions) were randomly assigned to the TFP productivity curriculum, and 283 (from 75 divisions) were assigned to the standard SOS productivity curriculum (with the imbalance arising from the six training sessions in which only the SOS productivity curriculum was delivered).

**Table A1: Treatment Distribution - Intervention Sample**

<i>Panel A: Treatment distribution among individuals who attended some training</i>					
		<u>Individual Level Training</u>			
		None	Status quo training (SOS)	T1 Individual	Total
T2 (team level) training?	Yes	522	110	62	694
	No	-	173	77	250
	Total	535	283	139	944

<i>Panel B: Treatment distribution across divisions having at least one official attending training</i>					
		<u>Individual Level Training</u>			
		None	Status quo training (SOS)	T1 Individual	Total
T2 (team level) training?	Yes	-	38	55	93
	No	-	37	35	72
	Total	-	75	90	165

Notes: Figures comprise only individuals from the core estimation sample who attended at least one training during the intervention period (panel A) or divisions that had at least one member attend a training (panel B). See text for details of treatment assignment and interpretation.

At the end of the ten-day SOS training (i.e. after the individual-level randomization and training), these individuals were randomized into receipt of the T2 training, which resulted in 172 trained individuals (representing 93 divisions) being randomized into receiving the follow-up T2 team-level training, with 250 (representing 72 divisions) not receiving the follow-up training. An additional 522 individuals who did not attend any individual training attended the T2 training,

because they were part of a division that had a member randomized into the T2 training. In this table we do not show the number of individuals or divisions who attended neither the T1 nor T2 trainings, as by definition these individuals did not appear in the sample of individuals we observed at the time of intervention.

Table A2 presents a balance check of individual characteristics of trainees attending the 10-days individual training at CSTC. The results show that this randomization was successful in balancing individual characteristics across nearly all observables.

In the first panel, the variable "sex" is a dummy variable taking 1 if the trainee is a female and 0 otherwise. "Share of training days attended before TFP session" indicate the proportion of days attended before being randomized in one of the productivity sessions; "Early team training" is a dummy variable taking 1 if the trainee attended the team training before attending the individual training, and 0 otherwise; "Took team training" is a dummy variable that takes 1 if the trainee attended the team training, and 0 otherwise. The variables "Division: FA", "Division: Human Resources", "Division: PPME", "Division: Research and statistics", "Division: other" are dummies variable taking 1 if the trainee work for one of the main organizational departments (FA, HR, RSIM, and PPME), and 0 otherwise. "Regional division" is a dummy variable that takes 1 if the trainees come from a regional division, and 0 otherwise. " Officials in first SOS session" is a variable that measures the number of officials attending the first SOS session attended by members of the official's division.

In the second and third panel, the variables represent a set of dummies variable that takes 1 if trainees took individual (team) training entry exam, individual (team) training exit exam, and individual (team) entry and exit exams, respectively.

In the last panel, the variables report the average test scores before attending the standard or the new productivity session in all the 7 topics.



**Table A2: Treatment Balance, TFP T1 vs SOS**

	SOS training (mean value)	TFP T1 training (difference from SOS)	N
<b>Sex</b>	0.550	-0.104** (0.049)	419
<b>Share of training days attended before TFP session</b>	0.839	0.021 (0.019)	419
<b>Early team training</b>	0.139	-0.039 (0.028)	419
<b>Team training</b>	0.729	-0.045 (0.049)	419
<b>Division: F&amp;A</b>	0.432	-0.087 (0.053)	419
<b>Division: Human Resources</b>	0.111	-0.003 (0.030)	419
<b>Division: PPME</b>	0.129	0.023 (0.033)	419
<b>Division: Research and statistics</b>	0.107	0.080** (0.040)	419
<b>Division: other</b>	0.007	0.022 (0.017)	419
<b># Officials in first SOS session</b>	1.836	0.143 (0.185)	419
<b>Took indiv. training entry exam</b>	0.746	0.059 (0.041)	419
<b>Took indiv. training exit exam</b>	0.750	0.092** (0.039)	419
<b>Took indiv. training entry and exit exam</b>	0.611	0.109** (0.045)	419
<b>Took team training entry exam</b>	0.343	0.053 (0.051)	419
<b>Took team training exit exam</b>	0.346	0.064 (0.052)	419
<b>Took team training entry and exit exam</b>	0.318	0.056 (0.051)	419
<b>Scores TFP Questions, 1st Entry exam</b>	0.378	-0.013 (0.029)	326
<b>Scores SOS Questions, 1st Entry exam</b>	0.290	-0.026 (0.035)	326
<b>Scores Other Questions, 1st Entry exam</b>	0.405	0.023 (0.028)	326
<b>Scores Management, 1st Entry exam</b>	0.546	-0.039 (0.036)	326
<b>Scores Incentives, 1st Entry exam</b>	0.601	0.005 (0.047)	326
<b>Scores Autonomy, 1st Entry exam</b>	0.491	-0.084** (0.042)	326
<b>Scores Management + TFP, 1st Entry exam</b>	0.460	-0.025 (0.027)	326

Note: See text for details of sample and variables.

Before the end of the ten-day SOS training, the research team randomly selected a subset of trainees whose divisions (their day-to-day work team in their home organization) would be invited for the follow-up TFP T2 training. Individuals were then randomly selected for T2 with 40 percent probability. This probability was calculated by simulation prior to the commencement of training to be the rate most likely to equalize sample sizes across treatment cells, with the aim of delivering approximately 100 T2 trainings over the course of the study period. Once a division had been selected for the T2 training it was removed from the randomization pool for future cohorts, so that the same division could not be selected twice for T2 training. In total, 91 divisions were selected for T2 training. These 91 divisions included 175 individuals who participated in the ten-day SOS training (in either the standard or TFP T1 productivity modules), as some divisions sent multiple members for training either in the same cohort or in different cohorts.

The T2 trainings were designed to be conducted approximately three to six weeks after the conclusion of the SOS training in which the individual from the division had participated, although in practice scheduling varied according to the availability of the division and trainers. Individuals whose divisions were selected were informed at the end of their ten-day service training, and OHCS subsequently wrote an official invitation letter to the division. The T2 trainings were delivered by the same set of CSTC trainers that delivered the TFP T1 trainings, but were held in a conference room at OHCS (rather than at CSTC classrooms) and ended slightly earlier in the afternoon than the individual-level productivity trainings, as it was impractical for an entire division to be away from the office for the whole day (although the content was substantively the same; see below). T2 trainings were capped at a maximum of 20 team members, as some divisions were very large; in these cases, facilitators worked with the division to identify the relevant sub-divisional unit with whom the individual trainee from the division worked most closely to determine which sub-group of division members should attend. This notwithstanding, CSTC and OHCS made strong efforts to ensure that all relevant staff from the division attended, up to and including the director of the division. Table A3 shows that this randomization into T2 training resulted in balance on observables across individuals whose divisions were and were not selected.

**Table A3: Treatment Balance, Individuals selected for T2 training**

	Individual training only (SOS or TFP T1; mean value)	T2 training (difference from SOS)	N
<b>Sex</b>	0.533	-0.025 (0.062)	419
<b>Share of days attended before TFP session</b>	0.864	-0.025 (0.022)	419
<b>Division: F&amp;A</b>	0.383	0.028 (0.103)	419
<b>Division: Human Resources</b>	0.050	0.084* (0.050)	419
<b>Division: PPME</b>	0.100	0.051 (0.049)	419
<b>Division: Research and statistics</b>	0.100	0.047 (0.058)	419
<b># Officials in first SOS session</b>	1.433	0.630* (0.377)	419
<b>Took indiv. training entry exam</b>	0.758	0.011 (0.049)	419
<b>Took indiv. training exit exam</b>	0.792	-0.016 (0.041)	419
<b>Took indiv. training entry and exit exam</b>	0.683	-0.051 (0.051)	419
<b>Scores TFP Questions, 1st Entry exam</b>	0.357	0.023 (0.043)	326
<b>Scores SOS Questions, 1st Entry exam</b>	0.260	0.029 (0.046)	326
<b>Scores Other Questions, 1st Entry exam</b>	0.388	0.034 (0.035)	326
<b>Scores Management, 1st Entry exam</b>	0.530	0.003 (0.044)	326
<b>Scores Incentives, 1st Entry exam</b>	0.599	0.005 (0.055)	326
<b>Scores Autonomy, 1st Entry exam</b>	0.462	0.000 (0.050)	326
<b>Scores Management + TFP, 1st Entry exam</b>	0.444	0.011 (0.038)	326

Note: See text for details of sample and variables.

While individual-level selection into TFP T1 and division-level selection into TFP T2 was randomized (conditional on being part of the ten-day SOS training), the selection of individuals to attend the ten-day SOS training itself was not randomized, nor was it controlled by CSTC or the research team. Each year CSTC publishes and circulates a calendar of trainings it will offer, and each organization then makes decisions about which individuals it will send to trainings on different dates. As organizations are not obliged to inform CSTC in advance of this, CSTC only finds out student numbers for each training based on who turns up at CSTC on the opening day of the training. When only a very small number of trainees arrive, trainings are sometimes cancelled and these trainees enrolled in the next training session for their grade. While individuals are supposed to attend the SOS training for their grade in the year preceding their becoming eligible to interview for promotion to the next grade, discussions with CSTC and investigation conducted by the research team into training timing with a sub-sample of four organizations revealed that this timing principle is not always followed, with other factors such as workload, budget availability, and convenience influencing when organizations choose to send their workers to the training. While we did not encounter evidence of organizations systematically basing their training timing decisions on factors that bias the SOS sample for our study period, we cannot rule out that selection into the ten-day SOS training was not quasi-random. Table A4 shows that while most observable characteristics of individuals who attended the ten-day, individual-level SOS training (with either the standard SOS or TFP T1 productivity module) were balanced with individuals who did not attend the ten-day SOS training, some variables do differ. Note that the sample for Table A4 is all individuals who attended a T2 training rather than the full short term learning questionnaire sample, as this allows us to measure productivity and management knowledge on a consistent scale.

**Table A4: Treatment Balance, Ten-day SOS vs. no individual-level training**

	Individuals not taking ten-day SOS training (mean value)	Individuals in ten-day SOS training (difference)	N
<b>Sex</b>	0.510	0.007 (0.047)	707
<b>Division: F&amp;A</b>	0.350	0.028 (0.061)	707
<b>Division: Human Resources</b>	0.116	0.076 (0.057)	707
<b>Division: PPME</b>	0.213	-0.050 (0.041)	707
<b>Division: Research and statistics</b>	0.135	0.028 (0.052)	707
<b># Officials in first SOS session</b>	1.507	0.540** (0.233)	707
<b>Official took team training entry exam</b>	0.892	-0.014 (0.031)	707
<b>Official took team training exit exam</b>	0.897	-0.002 (0.029)	707
<b>Official took team training entry and exit exam</b>	0.830	-0.010 (0.036)	707
<b>Scores TFP Questions, 1st Entry exam</b>	0.410	-0.032 (0.031)	612
<b>Scores SOS Questions, 1st Entry exam</b>	0.251	0.017 (0.034)	612
<b>Scores Other Questions, 1st Entry exam</b>	0.382	0.028 (0.044)	612
<b>Scores Management, 1st Entry exam</b>	0.520	-0.014 (0.032)	612
<b>Scores Incentives, 1st Entry exam</b>	0.652	-0.045 (0.059)	612
<b>Scores Autonomy, 1st Entry exam</b>	0.388	0.016 (0.046)	612
<b>Scores Management + TFP, 1st Entry exam</b>	0.454	-0.014 (0.023)	612

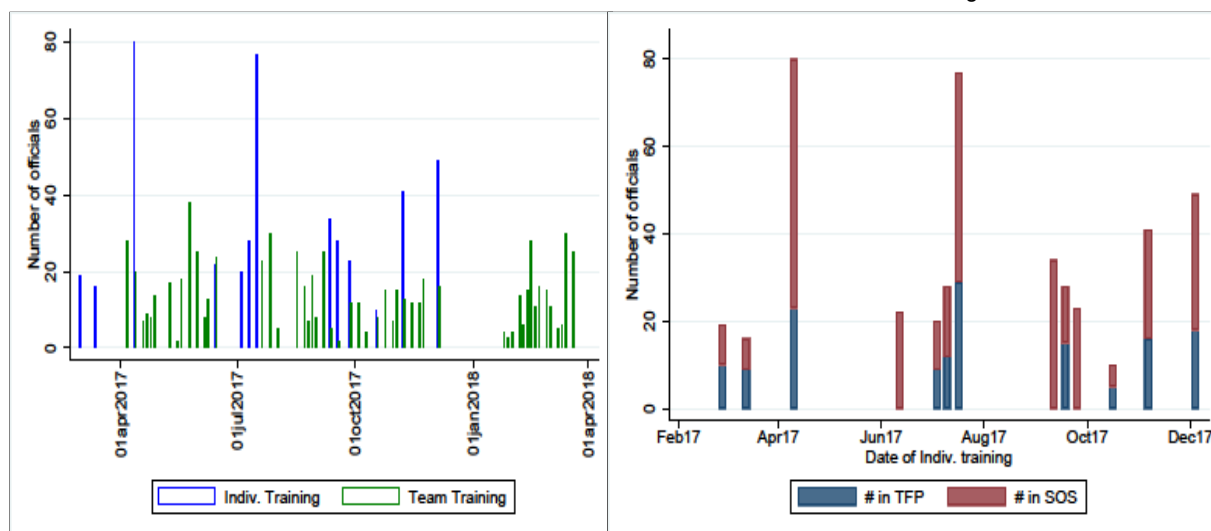
Note: See text for details of sample and variables.

Figure A1 plots the distribution of individual-level (SOS/TFP T1) and team-level (TFP T2) trainings across the study period (panel a) and the split between SOS and TFP T1 within each individual-level training (panel b). As panel (a) indicates, all individual-level trainings (SOS/TFP T1) occurred in the period February-December 2017, while team-level trainings occurred from April 2017 - March 2018 (due to the built-in lag). Panel (b) illustrates the roughly even split between standard SOS and TFP T1 productivity-trainings during the individual-level trainings in which randomization occurred. Figure A2 summarizes the timing of data collection and treatment periods.

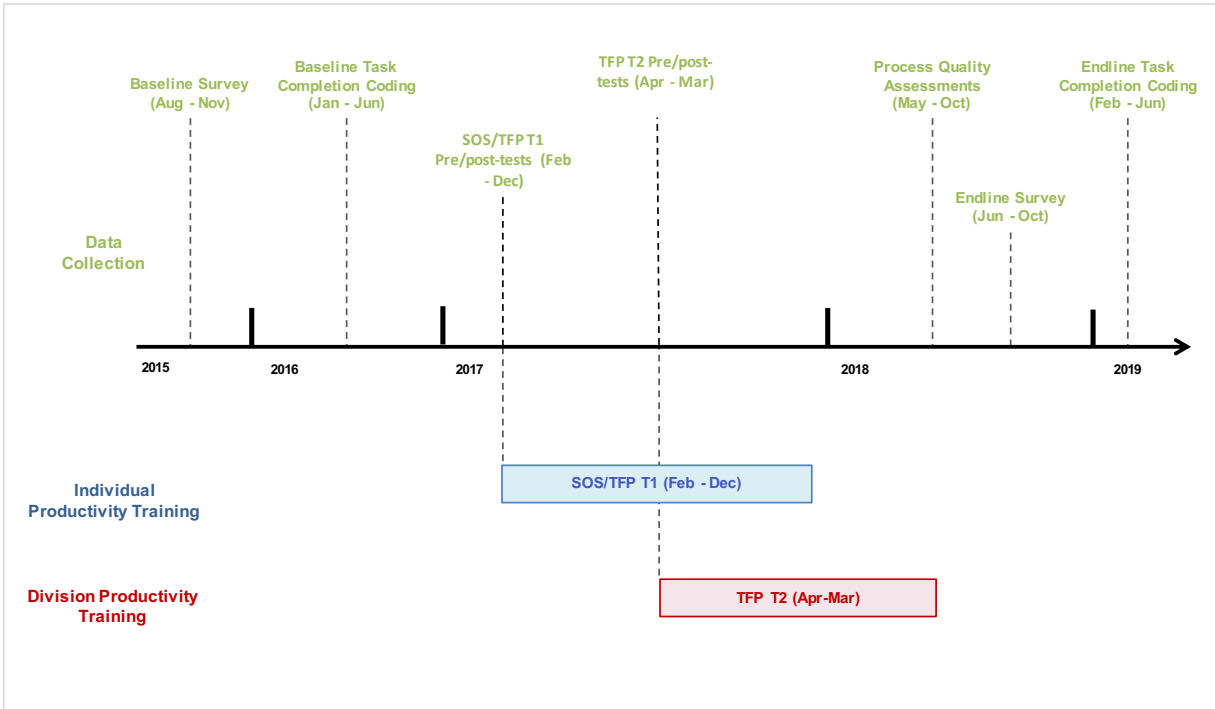
**Figure A1: Treatment timing and assignment**

(a) Dates of SOS/TFP T1 and TFP T2 trainings

(b) Split between SOS and TFP T1 trainings within individual training dates



**Figure A2: Treatment timing and assignment**



Note: 'Survey' refers to face-to-face enumerated interviews with public officials. 'Task completion coding' refers to coding of organization's annual and quarterly reports. 'Training Assessments' refer to pre- and post-training knowledge tests. 'Process Quality' Assessments refer to the quality of randomly chosen government files respectively.

## A.2 Training Content

The TFP training content revolved around three elements: 1) basic problem-solving skills related to identifying productivity constraints; 2) a motivational video about real Ghanaian civil servants; and 3) developing an innovation action plan and role playing suggesting it. These three elements were interwoven over the course of the day. We describe each element in turn.

The problem-solving skills section discussed the conceptualization of productivity in bureaucratic situations, presented some evidence gathered from the baseline survey that some organizations in Ghana’s Civil Service were more productive than others, discussed reasons for poor productivity, and prompted trainees to think of examples of productivity measurement and bottlenecks in their own and other government organizations. It also introduced four specific problem-solving skills: problem-tree analysis, force-field analysis, fishbone diagrams, and the “five whys”). The facilitator then helped trainees apply one of these methods in their own work context, and the class discussed this to solidify understanding.

The motivational video presented interviews with real Ghanaian civil servants from relatively high-performing divisions, identified by the research team in consultation with OHCS, as well as senior leaders (e.g. the Head of Civil Service). The use of a video was inspired by the evidence on the potential effectiveness of motivational videos in shifting norms (Bernard *et al* 2017) and the Head of Civil Service’s desire to promote wider take-up of effective practices already in use within the Civil Service, for which a video was a more scaleable intervention than study visits or other in-person approaches. Interviewees discussed the routines, skills, and norms their teams used to be productive on a daily basis, gave examples of small work process innovations they had come up with and introduced, and discussed the importance of productivity and innovation in the Civil Service. The video was approximately 20 minutes long, was produced by an Accra-based video producer with guidance from CSTC and the research team, and was shown in two parts to the trainees. Trainers facilitated discussion after each part, asked trainees to relate the video content back to their own experience, and referred to it throughout the latter part of the training. The video was extremely popular among trainees, with many of them citing it as their favorite part of the training and referring back to it in later follow-up discussions.



Over the course of the training, trainees completed a step-by-step action plan to help them develop a plan to improve work processes in their team. Trainees were prompted to focus on problems that were small and specific enough to be actionable within their team, rather than on large structural problems or on actions that would require significant resources or approvals from higher authorities. The completion of the action plan aligned with the training content through the day, so that trainees filled in the “problem identification” section of the action plan immediately after discussing how to identify productivity bottlenecks, and so on. Action plans were thus produced by trainees on a topic of their choosing, but with guidance, coaching, and feedback from the trainers. Appendix B presents the action plan templates. In addition to completing this form, facilitators helped trainees role play how to bring up their idea with colleagues and supervisors. Approximately halfway through the study period, the Head of Civil Service decided that promotion interview panels would henceforth ask interviewees about what steps they had taken to implement their action plans; from this point forward, facilitators informed trainees of this fact. Throughout the study period, however, trainees took a copy of their action plan away from training with them and were encouraged to discuss it with their colleagues and put their idea into action.

While the content of both the individual (T1) and team (T2) arms of the TFP training was substantively the same, the different formats necessitated some minor differences. First, in practice the T2 trainings ended slightly earlier in the afternoon than the T1 trainings, as it was not possible for many divisions to all be away from the office for an entire day. However, this was counter-balanced to an extent by the greater speed with which training groups could move through training components such as problem identification in the T2 trainings, since all trainees were from the same team and thus shared an operational context; this made it easier to apply concepts to practice, as facilitators did not have to discuss each trainee’s different context in turn. In addition, the 20-trainee cap on T2 trainings resulted in a slightly lower training cohort size in general than for T1 trainings. Thus while T2 trainings contained the same three major components as T1 trainings, some of the conceptual material and skills contained in the T1 training did not feature in the T2 trainings, but the application to trainees’ context was potentially stronger in the T2 training, so

the overall “strength” of the two treatment arms was as close to identical as possible. To ensure that these minor curricular differences did not drive our results, we ensured that the pre- and post-test questions and long-term learning questions (see Appendix C) focused only on material that each trainee would actually have learned. Second, whereas each individual trainee in T1 developed their own action plan, in T2 the entire team developed a single group action plan that they would implement together. While the challenge to be addressed and potential solutions were always identified during the training, many divisions opted to delegate finalization and follow-up of the action plan to a sub-team.

The existing SOS productivity curriculum had been developed by CSTC facilitators, in part under a donor-funded project some years prior. This training material focused much more heavily on the conceptual aspects of productivity, such as its proper definition and measurement and various terms stemming from the Japanese *kaizen* approach to continuous improvement. While trainers made an effort to give examples of how these concepts could be applied to Ghana’s Civil Service context and sought to make the sessions as interactive as possible, in general the existing SOS productivity sessions had fewer context-specific examples and lacked the concrete examples from the video.

While the standard SOS one-day productivity module did not include an action plan process, all trainees attending the ten-day SOS training were asked on the final day to complete an action plan (as was standard practice in CSTC’s SOS trainings prior to the study period) based on a different template than the TFP action plan template. Thus, while all trainees completed this SOS action plan, trainees who had been randomized into the TFP T1 productivity training on the previous day completed a second action plan on the final day. Trainees that had already drafted an action plan in the TFP T1 training were free to focus their SOS action plan on the same challenges and ideas that they had focused on in their TFP action plan. The templates were different in their details and structure, but both focused on getting trainees to identify a challenge in their work and develop a strategy to resolve it after the end of the training. Appendix B presents each template.

# B Online Appendix: Action Plan Templates

## B.1 TFP T1 Action Plan

### Handout 2

#### Improving Productivity: Action Planning Form

Name: ..... Date: .....

<b>1. INPUTS AND OUTPUTS</b>
<b>1.1 What are your inputs to your work?</b> <i>List the set of activities you undertake in your work day.</i>
<b>1.2 What are your outputs in your work?</b> <i>What do you produce from these activities? List the set of achievements from your work throughout the year.</i>
<b>2. PROBLEM STATEMENT</b>
<b>2.1 In what part of my division's work would I like to improve productivity?</b> <i>Think of one particular aspect of your area of work where you think there is scope to improve productivity – to do more for less, to increase the quantity or quality of outputs with the same or less resources.</i>
<b>3. IMPROVING PRODUCTIVITY</b>
<b>3.1 What needs to change to bring about this improvement?</b> <i>Identify changes in policy, work processes, skills, attitudes, resources or any other factors that may be relevant.</i>
<b>3.2 What are the main obstacles to making the necessary change(s)?</b> <i>Your thinking on this may change with further analysis of the problem – but record here your first thoughts.</i>



**3.3 What steps can I take to solve the problem? When?**

*Be specific about what you need to do to validate the nature of the problem and to identify and implement possible solutions, and when you intend to begin this.*

**3.4 What help or support do I need from others?**

*Think about what you need and from whom – including senior officers, colleagues, staff or stakeholders in other organisations – and how you will secure this.*

**3.5 What new skills or knowledge, if any, will I need to acquire to carry out this initiative?**

*Be specific about any new skills or knowledge you think you will need, and how and when you will seek to acquire these.*

**3.6 How will I know when I have succeeded?**

*As precisely as you can at this stage, say what your objective and your indicators of success for this initiative will be. Are your indicators SMART?*

**4. WHAT ARE YOU GOING TO DO ON MONDAY MORNING?**

**4.1 What are you going to do on Monday morning to start your productivity improvement?**

*As precisely as you can, say what you are going to do once you return to your office from the Scheme of Service trainings.*

## B.2 TFP T2 Action Plan

### Handout 2

#### Improving Productivity: Action Planning Form

Name: ..... Date: .....

<b>1. INPUTS AND OUTPUTS</b>
<b>1.1 What are your inputs to your work?</b> <i>List the set of activities you undertake in your work day.</i>
<b>1.2 What are your outputs in your work?</b> <i>What do you produce from these activities? List the set of achievements from your work throughout the year.</i>
<b>2. PROBLEM STATEMENT</b>
<b>2.1 In what part of the division's work would we like to improve productivity?</b> <i>Think of one particular aspect of your area of work where you think there is scope to improve productivity – to do more for less, to increase the quantity or quality of outputs with the same or less resources.</i>
<b>3. IMPROVING PRODUCTIVITY</b>
<b>3.1 What needs to change to bring about this improvement?</b> <i>Identify changes in policy, work processes, skills, attitudes, resources or any other factors that may be relevant.</i>
<b>3.2 What are the main obstacles to making the necessary change(s)?</b> <i>Your thinking on this may change with further analysis of the problem – but record here your first thoughts.</i>



**3.3 What steps can your division take to solve the problem? When?**

*Be specific about what you need to do to validate the nature of the problem and to identify and implement possible solutions, and when you intend to begin this.*

**3.4 What help or support do your division need from others?**

*Think about what you need and from whom – including senior officers, colleagues, staff or stakeholders in other organisations – and how you will secure this.*

**3.5 What new skills or knowledge, if any, will your division need to acquire to carry out this initiative?**

*Be specific about any new skills or knowledge you think you will need, and how and when you will seek to acquire these.*

**3.6 How will you know when the division have succeeded?**

*As precisely as you can at this stage, say what your objective and your indicators of success for this initiative will be. Are your indicators SMART?*

**4. WHAT ARE YOU GOING TO DO ON MONDAY MORNING?**

**4.1 What are you going to do on Monday morning to start your productivity improvement?**

*As precisely as you can, say what you are going to do once you return to your office after this in-plant training.*



### B.3 SOS Action Plan

#### ACTION PLAN FOR CIVIL SERVICE TRAINING CENTRE

Name of participant .....

MDA .....

Date of Scheme of Service Training.....

Current Grade .....

1. What are some of your job responsibilities in your organisation?

.....  
.....  
.....  
.....  
.....

2. What challenge(s) do you encounter when you want to meet these responsibilities?

.....  
.....  
.....  
.....  
.....

3. How can you resolve these challenges?

.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

4. After this scheme of service course, when can you implement the above mentioned changes?

.....

.....

.....

.....

.....

.....

# C Online Appendix: Data Collection Details

## C.1 Baseline and Endline Surveys

### C.1.1 Survey Sample and Procedure

The baseline and endline surveys were conducted in 2015 and 2018. The endline survey sought to replicate the methodology and structure of the baseline survey, with the exception of a handful of additional questions introduced specifically to measure the impacts of the training intervention (see below). Many members of the research team and even enumerators were the same across both surveys, which led to a high degree of consistency. Since we mainly use the endline survey in our analysis, we focus our discussion on the endline survey (noting significant differences in methodology where relevant).

The study sample of organizations for surveys was defined as all ministries and departments that form part of Ghana's core Civil Service (as opposed to the broader Public Service), with the exception of the Office of the President. This includes sectoral ministries responsible for overall policymaking and oversight (e.g. the Ministry of Education) but mostly excludes the frontline staff they oversee (e.g. teachers, who are part of the Ghana Education Service which is overseen by the Ministry of Education but organizationally separate). It also includes some central management organs of the state (e.g. Ministry of Finance, Controller and Accountant-General's Department) but excludes others (e.g. the Ghana Audit Service, Bank of Ghana), and also excludes independent agencies (e.g. the Driver and Vehicle Licensing Authority). Staff working in the national-level headquarters offices (including annex buildings) were included, but staff working in regional or district offices of the organization (including those in Accra localities) were excluded. This meant that even for departments that do fulfill frontline service delivery functions (e.g. Rent Control Department), the survey sample included the headquarters staff responsible for managing the organization but excluded most staff who are responsible for these service delivery functions.

To initiate the survey, OHCS sent a letter to each participating organization describing the survey and research partnership, and indicating the time window within which each survey team would visit the organization. A member of the survey team then made contact with the Director of

Human Resources Management in each organization to organize survey logistics and compile a list of all senior-grade staff who were currently operating within the organization, and which division they were part of; this step was necessary because the Civil Service's official personnel records are not always immediately updated when an officer moves organizations or moves between divisions in an organization.

A survey team comprising a research assistant and around five enumerators then visited each organization for a continuous period of anything from several days to several weeks (depending on the size of the organization), and conducted face-to-face, individual interviews with all staff on the staff list compiled by the organization. Enumerators explained to staff the goals of the study, that participation was voluntary and that individuals could refuse to participate without any consequences, and that individually identifiable data would be available only to the research team and would not be made available to the Civil Service. Interviews were conducted in private rooms. Since enumerators were also Civil Servants, the survey team leader ensured that enumerators did not know the names of the staff they were interviewing, no enumerator interviewed an officer they knew personally, and no enumerator conducted interviews in which they had previously worked. Survey teams made extensive efforts to interview every officer on the staff list, even if they were unavailable or were called to attend to work duties on short notice on the day initially planned for the interview, but officers who were unavailable (e.g. travelling, or on annual or maternity leave) throughout the survey team's entire period in the organization were not interviewed (representing the 14 percent of the total number of senior civil servants listed in the staff lists). In total, enumerators completed 3,302 interviews in 57 organizations during the endline survey, representing 84 percent of the total sample listed on the staff lists.

### **C.1.2 Survey Content**

To measure organizational culture, we adapt the Safety Attitudes Questionnaire developed by Sexton *et al* (2006) and used by Martinez *et al* (2015) to measure organizational cultures in hospitals that are oriented towards improving performance on patient safety. We take a subset of their scale items (given time constraints in the interview) and adapt their wording to focus on a

bureaucratic rather than medical context, which asks respondents to state their agreement on a 1-5 Likert scale with a set of statements about organizational culture, and administer the scale to the same subset of respondents to whom the WMS-style management quality questions were asked. In addition, we developed two additional sets of questions: one pertaining to the introduction of new ideas that we administer which we aggregate into an additional sub-index we term “fostering new ideas”; and one pertaining to individuals’ perceptions of how their division performs relative to others, which we aggregate into a “relative performance” sub-index. The questions on each score are averaged at the division level and then converted into normalized z-scores by taking unweighted means and standard deviation at the divisional level. The z-scores are aggregated into sub-indexes and then transformed in z-scores to obtain continuous variables with mean zero and variance one.

**Table A5: Comparison of Organizational Culture Scales**

<i>Sub-index</i>	<i>Scale items used in this paper</i>		<i>Sexton et al (2006) scale items</i>	
<i>Teamwork climate</i>	<p><b>It is easy for personnel here to ask questions when there is something that they do not understand</b></p> <p><b>The managers and other officers here work together as a well-coordinated team</b></p>	<p><b>Disagreements in this division are resolved appropriately (i.e., not who is right, but what is best for the service)</b></p>	<p><b>It is easy for personnel in this ICU to ask questions when there is something that they do not understand.</b></p> <p><b>The physicians and nurses here work together as a well-coordinated team.</b></p> <p>In this ICU, it is difficult to speak up if I perceive a problem with patient care.</p>	<p><b>Disagreements in this ICU are resolved appropriately (i.e., not who is right, but what is best for the patient)</b></p> <p>I have the support I need from other personnel to care for patients.</p> <p>Nurse input is well received in this ICU.</p>
<i>Performance climate (safety climate in Sexton et al)</i>	<p><b>The culture in this division makes it easy to learn from the errors of others</b></p> <p><b>You know the proper channels to direct questions regarding correct bureaucratic process in this division</b></p> <p><b>You receive appropriate feedback about your performance</b></p>	<p><b>Bureaucratic errors are handled appropriately in this division</b></p> <p><b>You are encouraged by your colleagues to report any work concerns you may have</b></p> <p><b>You would feel happy being served as a Ghanaian citizen by this division</b></p>	<p><b>The culture in this ICU makes it easy to learn from the errors of others.</b></p> <p><b>I know the proper channels to direct questions regarding patient safety in this ICU.</b></p> <p><b>I receive appropriate feedback about my performance.</b></p> <p>In this ICU, it is difficult to discuss errors.</p>	<p><b>Medical errors are handled appropriately in this ICU.</b></p> <p><b>I am encouraged by my colleagues to report any patient safety concerns I may have</b></p> <p><b>I would feel safe being treated here as a patient.</b></p>
<i>Job satisfaction</i>			<p>This hospital is a good place to work.</p> <p>Working in this hospital is like being part of a large family.</p> <p>I like my job.</p>	<p>I am proud to work at this hospital.</p> <p>Moral in this ICU area is high.</p>
<i>Stress recognition</i>	<p><b>Fatigue impairs your performance during high-pressure situations (e.g. when there are heavy demands on your division)</b></p>		<p><b>Fatigue impairs my performance during emergency situations (e.g., emergency resuscitation, seizure).</b></p> <p>I am more likely to make errors in tense or hostile situations.</p>	<p>When my workload becomes excessive, my performance is impaired.</p> <p>I am less effective at work when fatigued.</p>
<i>Perceptions of management</i>	<p><b>Staff (divisional management) doesn't knowingly compromise division services</b></p> <p><b>You get adequate, timely info about events that might affect your work from your division.</b></p>	<p><b>Staff (divisional management) supports your daily efforts</b></p>	<p><b>Hospital management does not knowingly compromise the safety of patients.</b></p> <p><b>I am provided with adequate, timely information about events in the hospital that might affect my work.</b></p>	<p><b>Hospital administration supports my daily efforts.</b></p> <p>The levels of staffing in this clinical area are sufficient to handle the number of patients</p>
<i>Working conditions</i>	<p><b>All the necessary information for diagnostic and effective decision making is routinely available to you</b></p>	<p><b>Trainees in your division are adequately supervised</b></p>	<p><b>All the necessary information for diagnostic and therapeutic decisions is routinely available to me.</b></p> <p>This hospital constructively deals with problem physicians and employees.</p>	<p><b>Trainees in my discipline are adequately supervised.</b></p> <p>This hospital does a good job of training new personnel</p>
<i>Fostering new ideas</i>	<p>Your suggestions about work place productivity would be acted upon if you expressed them to management</p> <p>You can see lots of ways to make your division work better.</p>	<p>Staff (divisional management) in this division are quick to adopt (are open to) new ways of doing things.</p>		
<i>Relative performance</i>	<p>Your division works better than others in this organization</p>	<p>Staff (divisional management) in this division are doing a good job</p>		

**Notes:** Items in bold are substantively identical in both this paper's and Sexton *et al*'s (2006) scales. Text in (parentheses) indicates item phrasing administered to non-management-level staff; all other items are identical for all respondents.

For four scale items, we administered slightly rephrased versions depending on whether the respondent had been previously identified in the survey as a manager (Directors, i.e. heads of divisions or equivalent rank) or non-managerial staff (the vast majority of respondents). This was due to the nature of the items, which referred to perceptions of subordinates/superiors. The phrasing used for managers is indicated in Table A5 in parentheses. All other questions were phrased identically for managers and non-managers. All scale items were identical at baseline and endline, with one exception: instead of non-managers receiving the non-manager-track item “Divisional management in this division are open to new ways of doing things” in the baseline survey, they received the manager-track item “Staff in this division are quick to adopt new ways of doing things.” This was due to a clerical error, which was corrected in the endline survey. Since the question still refers to the same underlying construct, we use non-managers’ responses to the manager-track item to construct the overall culture index in the baseline.

Since the Sexton *et al* (2006) organizational culture scale on which our scale is based was developed in a hospital context and shown to be related to patient safety performance in that context, we wanted to examine whether our adapted organizational culture measure was related to performance in our bureaucratic context. To do so, we examined whether organizations’ organizational culture score was correlated with their task completion, both measured at baseline (i.e. prior to the intervention). To do this we replicate the analysis of Rasul *et al* (2019), who regress task completion on an index of management quality index plus controls, but we substitute our organizational culture index for the management quality index. We therefore estimate the partial correlation of culture with task completion, after controlling for a wide range of other variables. While this association is not causal, and testing whether culture affects task completion is not the primary purpose of this paper, this analysis nevertheless provides a preliminary check that our adapted culture index is meaningful in the bureaucratic context.

More specifically, we follow Rasul *et al* (2019) in estimating a regression of the form:

$$y_{ijn} = \gamma_1 Culture_n + \beta_1 PC_{ijn} + \beta_2 OC_n + \lambda_j + \epsilon_{ijn} \quad (6)$$

where  $y_{ijn}$  is a binary measure of the completion of task  $i$  of type  $j$  in organization  $n$ ,  $Culture_n$

is our organizational culture index (or one of its sub-indices) expressed as a z-score,  $PC_{ijn}$  is a vector of task controls (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), and  $OC_n$  is a vector of noise controls (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). We also condition on a set of “noise” controls including 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. The coefficient of interest is  $\gamma_1$ , which represents the partial correlation of organizational culture with task completion.

**Table A6: Organizational Culture and Task Completion**

Dependent Variable:	Task Completion (binary)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Aggregate organizational culture	0.07** (0.03)							
Teamwork climate		0.08*** (0.02)						
Performance climate			0.05 (0.03)					
Stress recognition				0.10*** (0.03)				
Working conditions					0.01 (0.03)			
Perception of management						0.08*** (0.02)		
New ideas							0.01 (0.03)	
Relative Performance								0.07 (0.05)
<b>Task Controls</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<b>Organizational Controls</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<b>Noise Controls</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<b>Fixed Effects</b>	Task type, Sector	Task type, Sector	Task type, Sector	Task type, Sector	Task type, Sector	Task type, Sector	Task type, Sector	Task type, Sector
<b>Observations (clusters)</b>	3628 (31)	3628 (31)	3628 (31)	3628 (31)	3628 (31)	3628 (31)	3628 (31)	3628 (31)

Notes: \*\*\* denotes significance at 1%, \*\* at 5%, and \* at 10% level. Standard errors are in parentheses, and are clustered by organization throughout. All columns report OLS estimates. See text for details of controls and fixed effects.

Column 1 of Table A6 shows that the overall organizational culture index is significantly



positively related to task completion: a one standard deviation increase in organizational culture increases the likelihood of a task being fully completed by 7 percentage points. Since only 34 percent of tasks were fully completed in the baseline 2015 data, this is a substantial increase. Columns 2-8 then show that all the organizational culture sub-indices are also positively correlated with task completion, in some cases statistically significantly so. While these estimates cannot be conclusively interpreted as causal, they nonetheless provide some measure of validation that the organizational culture index is measuring something meaningful for understanding performance in a bureaucratic setting.

To measure management quality, we adapted the widely used World Management Survey (WMS) approach to measuring organizational management quality to the public sector context, following Rasul and Rogger (2018) and Rasul *et al* (2019). The method used in the endline survey was identical to that described in Rasul *et al* (2019), incorporating many of the techniques described by Bloom and Van Reenen (2007). Differently from Bloom and Van Reenen (2007) and other WMS applications, however, we focused the questions at the level of the division rather than the organization, in order to capture within-organization variation in the practices. For each of the practices listed in Table A7, the enumerator would ask the respondent an open-ended question about how that particular aspect of management worked in their division, and followed this up flexibly with probing follow-ups and requests for examples in order to establish as accurately possible the actual practice used in the division (as opposed to what is supposed to happen, or what the respondent thinks it should be). This gave the management practice module the feel of a semi-structured qualitative interview rather than the administration of a more standard closed-ended quantitative questionnaire. After the brief discussion of each practice, the enumerator would then benchmark that practice against the 1-5 scale defined in Table A7, with higher scores corresponding to commonly accepted notions of “better” management (although the relationship of these practices to actual performance is not necessarily straightforward, as Rasul *et al* [2019] demonstrate. To avoid social desirability bias, respondents were not told in advance that the enumerators were benchmarking their qualitative responses against a quantitative scale.

These individual practice scores were then aggregated by six management practices (roles,

flexibility, performance incentives, monitoring, staffing, and targeting) to create a division-level aggregate score for the overall measure, and for each of the three sub-indices (autonomy/discretion, incentives/monitoring, other) defined by Rasul and Rogger (2018) and Rasul *et al* (2019). In order to reduce the length of the survey, individuals were randomly allocated to receive either this management module or another question module, although the first interview in each division and the division's Director always received the management module in order to ensure that we obtained at least one management score per division.

**Table A7: 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 A7 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.
		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.
	Monitoring	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 A7 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.
	Targeting	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.
		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.

All respondents were asked a set of three questions that measured respondents' knowledge of productivity concepts and definitions that were delivered as part of the SOS and TFP productivity modules. Table A8 presents these questions. Each question and response set corresponds to a written definition presented to trainees during the respective training modules. These questions are a subset of the full training knowledge questions asked in the training pre- and post-tests (see section C.2 below).

**Table A8: Productivity Knowledge Questions in Endline Survey**

Question	Response Options (correct answer bolded)	Correspondance to Productivity Module Content
In your opinion, which of the following is NOT a key principle for conducting a successful brainstorm session?	<ul style="list-style-type: none"> <li>• Spontaneity</li> <li>• Suspension of judgment</li> <li>• <b>Serendipity</b></li> <li>• Speed</li> </ul>	TFP (T1 & T2)
In your opinion, which of the following is NOT a stage in the productivity movement?	<ul style="list-style-type: none"> <li>• Ownership Stage</li> <li>• Action Stage</li> <li>• <b>Improvement Stage</b></li> <li>• Awareness Stage</li> </ul>	SOS
What do you think is the best definition of productivity?	<ul style="list-style-type: none"> <li>• Increasing production at the expense of quality</li> <li>• Working harder and putting in overtime in order to achieve results</li> <li>• <b>Output divided by Input</b></li> <li>• Input divided by Output</li> </ul>	Both TFP & SOS

**Notes:** Questions and correct answers are taken from the SOS and TFP (T1 & T2) productivity training modules.

The module measuring follow-up on action plan implementation was administered to all respondents, but some questions (e.g. on action plan follow-up) were only asked to officers who indicated either that: they had attended a Scheme of Service training at CSTC; or they had attended a training with their division at OHCS since March 2017. Table A9 presents the full set of questions asked to individuals about their learning and follow-up on the training (for those who indicated they had taken training in the past year). For brevity we present the questions that correspond to the individual-level trainings (SOS and TFP T1) delivered at CSTC; the same questions (but phrased to refer to the individual’s division, and OHCS as the training location) were asked to those who had attended a T2 training at OHCS. The second column of Table A9 indicates which responses for each question corresponded to an intermediate outcome that the TFP productivity trainings aimed to affect. We test whether each of these outcomes were affected separately, but for power we also create several indices of related outcomes following Anderson (2008) as well as an overall index that captures all 18 potential outcomes.

**Table A9: Intermediate Outcomes Used in Anderson Indices**

Question	Response Options Included in Anderson Index	Response Options Not Included in Anderson Index	Included In Indices
In which components of the Scheme of Service training did you learn the most?	<ul style="list-style-type: none"> <li>• Productivity</li> </ul>	<ul style="list-style-type: none"> <li>• Managing people/HRM</li> <li>• Ethics</li> <li>• Monitoring and evaluation</li> <li>• Administrative writing/skills</li> <li>• General Civil Service issues</li> <li>• Other</li> </ul>	<ul style="list-style-type: none"> <li>• Overall</li> </ul>
Which components of the Scheme of Service training were most useful once you went back to your workplace?	<ul style="list-style-type: none"> <li>• Productivity</li> </ul>	<ul style="list-style-type: none"> <li>• Managing people/HRM</li> <li>• Ethics</li> <li>• Monitoring and evaluation</li> <li>• Administrative writing/skills</li> <li>• General Civil Service issues</li> <li>• Other</li> </ul>	<ul style="list-style-type: none"> <li>• Overall</li> </ul>
What were the biggest obstacles towards putting the skills you learned in Scheme of Service training to use in your workplace? (reverse-scaled)	<ul style="list-style-type: none"> <li>• Training material was not relevant to my work</li> <li>• Manager not supportive of implementing new ideas/practices</li> <li>• Division as a whole not supportive of implementing new ideas/practices</li> </ul>	<ul style="list-style-type: none"> <li>• Lack of resources/logistical constraints</li> <li>• Putting training material into practice was too difficult</li> <li>• Other</li> </ul>	<ul style="list-style-type: none"> <li>• Obstacles to Applying Learning</li> <li>• Overall</li> </ul>
Did the process of formulating an Action Plan help you think of new ideas to improve productivity?	<ul style="list-style-type: none"> <li>• Yes</li> </ul>	<ul style="list-style-type: none"> <li>• No</li> </ul>	<ul style="list-style-type: none"> <li>• Overall</li> </ul>
After you finished training and returned to your organization, did you take to try to implement your Action Plan?	<ul style="list-style-type: none"> <li>• Yes</li> </ul>	<ul style="list-style-type: none"> <li>• No</li> </ul>	<ul style="list-style-type: none"> <li>• Overall</li> </ul>
Which of the following steps did you take to try to implement your Action Plan?	<ul style="list-style-type: none"> <li>• Discussed with superiors (e.g. Director, Chief Director)</li> <li>• Discussed with other colleagues</li> <li>• Set up a team/committee</li> <li>• Undertook additional feasibility research</li> </ul>	<ul style="list-style-type: none"> <li>• Worked as an individual to implement Action Plan</li> <li>• Other</li> <li>• None of the above</li> </ul>	<ul style="list-style-type: none"> <li>• Steps Taken</li> <li>• Overall</li> </ul>
Were you able to implement your Action Plan?	<ul style="list-style-type: none"> <li>• Yes, fully</li> <li>• Partially</li> </ul>	<ul style="list-style-type: none"> <li>Not at all</li> </ul>	<ul style="list-style-type: none"> <li>• Overall</li> </ul>
What were the main obstacles to implementing your Action Plan? (reverse-scaled)	<ul style="list-style-type: none"> <li>• Idea was not a good one</li> <li>• Lack of resources/logistical constraints</li> <li>• Manager not supportive of implementing new ideas/practices</li> <li>• Division as a whole not supportive of implementing new ideas/practices</li> </ul>	<ul style="list-style-type: none"> <li>• Putting idea into practice was too difficult</li> <li>• No incentives/motivation for implementing Action Plan</li> <li>• Other</li> </ul>	<ul style="list-style-type: none"> <li>• Obstacles to Implementing Action Plan</li> <li>• Overall</li> </ul>
On a scale of 1-5, where 1 is not at all and 5 is very significantly, how much do you think the implementation of your Action Plan improved your division's productivity?	<ul style="list-style-type: none"> <li>• Significantly</li> <li>• Very significantly</li> </ul>	<ul style="list-style-type: none"> <li>• Not at all</li> <li>• A little bit</li> <li>• Somewhat</li> </ul>	<ul style="list-style-type: none"> <li>• Overall</li> </ul>

**Notes:** First column indicates question asked to respondents; second column indicates response options that were included as binary outcomes (i.e. selected/not selected) in the intermediate outcomes Anderson indices; third column indicates response options that were not included don't know/refused not included in any indices; fourth column indicates which indices each question set formed part of. Responses of "Don't know" or refusals to answer not included in any indices.

## C.2 Pre- and Post-Training Tests

The pre- and post-training tests were administered on the first and final days of the ten-day SOS productivity training, and an abbreviated version first thing in the morning and at the end of the day of the division-level TFP T2 training. The percentage of training participants completing both tests was 61, 73, and 83 percent in SOS, TFP T1, and TFP T2 treatment groups, respectively.

Two question sets were developed. Within each ten-day SOS training and TFP T2 training, one set was administered as the pre-test and the other was administered as the post-test, and for the subsequent session this was reversed so that the set previously used as the post-test was used as the pre-test as vice versa. At any given sitting, however, all trainees were responding to the same question set, to avoid the risk of participants learning the contents of the other question set prior to taking it as a post-test, and all trainees took different question sets for their pre- and post-tests. The full set of questions is presented in Table A10.

**Table A10: Productivity and Management Questions in Training Pre- and Post-Tests**

Question	Response Options (correct answer bolded)	Module	Question Sets
<i>Questions related to productivity training modules</i>			
Which of the following is NOT a stage in the productivity movement?	<ul style="list-style-type: none"> <li>• Ownership Stage</li> <li>• Action Stage</li> <li>• <b>Improvement Stage</b></li> <li>• Awareness Stage</li> </ul>	SOS	SOS/TFP T1 (A) TFP T2 (A)
When conducting a problem-tree analysis, what should your starting point be?	<ul style="list-style-type: none"> <li>• <b>The problem</b></li> <li>• The desired solution</li> <li>• The effects of the problem</li> <li>• The causes of the problem</li> </ul>	TFP	SOS/TFP T1 (A)
What is the “Kaizen” approach to management?	<ul style="list-style-type: none"> <li>• Improving adherence to protocols</li> <li>• Encouraging staff to work harder with incentives</li> <li>• Implementing best practices across all aspects of the organization</li> <li>• <b>Making gradual, continuous changes in workplace practices</b></li> </ul>	SOS	SOS/TFP T1 (A)
Which of the following is NOT a characteristic of an effective team?	<ul style="list-style-type: none"> <li>• There is a focus on both the tasks (what do we need to do?) and the process (how do we achieve this?)</li> <li>• There is a lot of discussion in which everyone participates – listens, speaks, and is heard</li> <li>• There is a range of individuals who contribute in different ways</li> <li>• <b>There is no space for disagreements, since these could cause disharmony</b></li> </ul>	TFP	SOS/TFP T1 (A) TFP T2 (A)
Which one of the following is NOT a dimension of productivity?	<ul style="list-style-type: none"> <li>• <b>Professional</b></li> <li>• Organizational</li> <li>• Individual</li> <li>• National</li> </ul>	SOS	SOS/TFP T1 (A) TFP T2 (A)
Which of the following is NOT a key principle for conducting a successful brainstorm session?	<ul style="list-style-type: none"> <li>• Spontaneity</li> <li>• Suspension of judgment</li> <li>• <b>Serendipity</b></li> <li>• Speed</li> </ul>	TFP	SOS/TFP T1 (A) TFP T2 (A)
Which of the following is NOT part of the Awareness Stage of the productivity movement?	<ul style="list-style-type: none"> <li>• Soliciting buy-in for the new direction</li> <li>• Creating a mirror picture of success</li> <li>• Setting the objective for change</li> <li>• <b>Mobilizing funding</b></li> </ul>	SOS	SOS/TFP T1 (B) TFP T2 (B)
How does a force-field analysis try to solve problems?	<ul style="list-style-type: none"> <li>• <b>By isolating the resisting forces and the driving forces preventing you from moving from an undesired to a desired state</b></li> <li>• By focusing on immovable barriers to moving from an undesired to a desired state</li> <li>• By considering problems in a static context</li> <li>• By creating a problem tree diagram to help solve the problem</li> </ul>	TFP	SOS/TFP T1 (B)
Which of the following is NOT a principle of Total Quality Management?	<ul style="list-style-type: none"> <li>• <b>Hierarchy</b></li> <li>• System approach to management</li> <li>• Continuous improvement</li> <li>• Mutually beneficial stakeholder relationships</li> </ul>	SOS	SOS/TFP T1 (B)
Which of the following is NOT a component of competency?	<ul style="list-style-type: none"> <li>• Intellect</li> <li>• Drive/motivation</li> <li>• Personality/style</li> <li>• <b>Rank</b></li> </ul>	TFP	SOS/TFP T1 (B) TFP T2 (B)
What is the definition of productivity?	<ul style="list-style-type: none"> <li>• Increasing production at the expense of quality</li> <li>• Working harder and putting in overtime in order to achieve results</li> <li>• <b>Output/Input</b></li> <li>• Input/Output</li> </ul>	SOS	SOS/TFP T1 (B) TFP T2 (B)
Which of the following is NOT an important part of a brainstorming session?	<ul style="list-style-type: none"> <li>• <b>Having a solution in mind when you start</b></li> <li>• Having a facilitator to encourage and prompt thinking</li> <li>• Having a scribe keep all ideas visible by writing them on a flipchart or on papers</li> <li>• Encouraging silent participants to come out and speak</li> </ul>	TFP	SOS/TFP T1 (B) TFP T2 (B)

**Notes:** Questions and correct answers are taken from the SOS and TFP (T1 & T2) productivity training modules. For management practice survey modules, “correct” option indicates the response that corresponds most closely to a score of 5 on the survey-based management practice module (see Appendix C.1).



**Table A10 Continued: Productivity and Management Questions in Training Pre- and Post-Tests**

Question	Response Options (correct answer bolded)	Module	Question Sets
<i>Questions related to non-productivity training modules</i>			
Suppose that your division has been tasked with creating a new policy on some topic within your area of competence. What should be your first step?	<ul style="list-style-type: none"> <li>• Draft a policy document specifying the details of the policy</li> <li>• Create a monitoring plan for the policy</li> <li>• <b>Conduct a review of existing evidence on the effectiveness of similar policies on this topic</b></li> <li>• Conduct an impact evaluation of your policy</li> </ul>	Evidence-based policymaking	SOS/TFP T1 (A)
Which of the following is NOT an example of an official security classification?	<ul style="list-style-type: none"> <li>• SECRET</li> <li>• RESTRICTED</li> <li>• <b>INTERNAL USE ONLY</b></li> <li>• CONFIDENTIAL</li> </ul>	Organizational safety and security	SOS/TFP T1 (A) TFP T2 (A)
Which one of these should always be included in handing-over notes?	<ul style="list-style-type: none"> <li>• Advice for the relieving officer</li> <li>• <b>An itemized list of office assets</b></li> <li>• The writer's opinion on the effectiveness of policies in this area</li> <li>• A monitoring plan</li> </ul>	Administrative writing	SOS/TFP T1 (A)
Which of the following is NOT necessarily a member of a Disciplinary Committee constituted under the Civil Service Code?	<ul style="list-style-type: none"> <li>• A Senior Officer nominated by the Disciplinary Authority as Chairman</li> <li>• <b>An officer from the Office of the Head of Civil Service</b></li> <li>• One representative of the Departmental Local Labour Union</li> <li>• The Personnel Officer or an officer acting in that capacity</li> </ul>	Civil service administration principles	SOS/TFP T1 (A)
What are the three essential characteristics of a written work standard for a task?	<ul style="list-style-type: none"> <li>• History of the organization; Procedure for carrying out the task; Standard of performance which must be achieved</li> <li>• Procedure for carrying out the task; Standard of performance which must be achieved; Monitoring plan for the task</li> <li>• Responsibility for the task; Standard of performance which must be achieved; Monitoring plan for the task</li> <li>• <b>Responsibility for the task; Procedure for carrying out the task; Standard of performance which must be achieved</b></li> </ul>	Work ethic and work standards	SOS/TFP T1 (A)
Which of the following is NOT an example of an individual working style?	<ul style="list-style-type: none"> <li>• The Analytical</li> <li>• The Amiable</li> <li>• <b>The Evaluative</b></li> <li>• The Expressive</li> </ul>	Human relations	SOS/TFP T1 (A)
Which is NOT a step in the policy development cycle?	<ul style="list-style-type: none"> <li>• Developing policy options</li> <li>• Evaluating the outcome</li> <li>• Selecting and recommending a policy option</li> <li>• <b>Hiring new personnel</b></li> </ul>	Evidence-based policymaking	SOS/TFP T1 (B)
Which of the following is NOT part of the record life cycle?	<ul style="list-style-type: none"> <li>• Semi-current stage</li> <li>• Non-current stage</li> <li>• Current stage</li> <li>• <b>Partially current stage</b></li> </ul>	Organizational safety and security	SOS/TFP T1 (B) TFP T2 (B)
Which of the following is a type of administrative writing?	<ul style="list-style-type: none"> <li>• Acts of Parliament</li> <li>• Press releases</li> <li>• <b>Circulars</b></li> <li>• Legislative Instruments</li> </ul>	Administrative writing	SOS/TFP T1 (B)
Who has the responsibility for initiating recruitment in the Civil Service?	<ul style="list-style-type: none"> <li>• The Ministry of Finance</li> <li>• The Head of the Civil Service</li> <li>• <b>The Heads of Ministries, Departments, and Agencies (MDA)</b></li> <li>• The Public Services Commission</li> </ul>	Civil service administration principles	SOS/TFP T1 (B)
Which of these is NOT a reason why written work standards are important?	<ul style="list-style-type: none"> <li>• <b>Officers should only do things that are specified in writing.</b></li> <li>• Written work standards help avoiding management by generalization and personality.</li> <li>• To assist in recruitment and hiring by clearly defining what a good job looks like.</li> <li>• To provide the baseline references that are necessary for learning, and providing a stable platform for collecting performance measurements.</li> </ul>	Work ethic and work standards	SOS/TFP T1 (B)
Which of the following is an example of a type of bias in perceptions we can have about other people?	<ul style="list-style-type: none"> <li>• Recognizing</li> <li>• <b>Stereotyping</b></li> <li>• Discussing</li> <li>• Expressing</li> </ul>	Human relations	SOS/TFP T1 (B)

**Notes:** Questions and correct answers are taken from the SOS and TFP (T1 & T2) productivity training modules. For management practice survey modules, "correct" option indicates the response that corresponds most closely to a score of 5 on the survey-based management practice module (see Appendix C.1).

**Table A10 Continued: Productivity and Management Questions in Training Pre- and Post-Tests**

Question	Response Options (correct answer bolded)	Module	Question Sets
<i>Questions related to management practice survey modules</i>			
How should each division/directorate in an organization track how well it is delivering services?	<ul style="list-style-type: none"> <li>• A division should have as many performance indicators specified as possible, and should rely exclusively on formal indicators of performance.</li> <li>• <b>A division should track performance continuously with selected indicators that are directly relevant to the division's objectives, using both formal indicators and other less formal assessments of performance.</b></li> <li>• A division should rely mainly on less formal ways of tracking performance, based on the knowledge of the division's Director.</li> <li>• I'm not sure</li> </ul>	Incentives/ monitoring	SOS/TFP T1 (A) TFP T2 (A)
Within a public sector organization, how much discretion should senior officers be given to carry out their assignments?	<ul style="list-style-type: none"> <li>• Officers should have some independence about how they go about their work.</li> <li>• Officers should do what their Director tells them to do and not deviate from those instructions.</li> <li>• <b>Officers should have a lot of independence to decide how best to complete their tasks.</b></li> <li>• I'm not sure</li> </ul>	Autonomy/ discretion	SOS/TFP T1 (A) TFP T2 (A)
How much should civil servants make efforts to adjust to the specific needs and peculiarities of different communities, clients, or other stakeholders?	<ul style="list-style-type: none"> <li>• Civil servants should use the same procedures in all cases, regardless of the particularities of who they are serving.</li> <li>• <b>Civil servants should adjust their procedures as much as possible, within regulations, to respond to the needs of particular communities/ clients.</b></li> <li>• Civil servants should only make minor adjustments to procedures based on who they are serving.</li> <li>• I'm not sure</li> </ul>	Autonomy/ discretion	SOS/TFP T1 (A)
Should public sector organizations use performance targets, or other indicators for tracking and rewarding (financially or non-financially) the performance of staff?	<ul style="list-style-type: none"> <li>• <b>Performance should be closely monitored, with significant rewards or punishments for those who overperform or underperform based on those indicators.</b></li> <li>• Officers' performance should not be compared.</li> <li>• Good or bad performance should be rewarded or punished, but the criteria for this should be up to management rather than based on formal indicators.</li> </ul>	Incentives/ monitoring	SOS/TFP T1 (A)
How should officers be disciplined for breaking the rule of the Civil Service?	<ul style="list-style-type: none"> <li>• Occasional violations do not need to be punished every time, but officers that break the rules regularly should be disciplined.</li> <li>• In general, breaking the rules should not be punished.</li> <li>• <b>Any violation of the rules should always be punished.</b></li> <li>• I'm not sure</li> </ul>	Incentives/ monitoring	SOS/TFP T1 (B) TFP T2 (B)
What kind of contributions should staff be able to make to the process of policy formulation and implementation?	<ul style="list-style-type: none"> <li>• Only top management should make substantive contributions to organisational policies and their implementation.</li> <li>• <b>All officers should make substantive contributions to organisational policies and their implementation.</b></li> <li>• All officers should make substantive contributions in staff meetings, but otherwise it should be left for top management.</li> <li>• I'm not sure</li> </ul>	Autonomy/ discretion	SOS/TFP T1 (B) TFP T2 (B)
How flexible should divisions or directorates be in terms of responding to new and improved work practices?	<ul style="list-style-type: none"> <li>• There is no need to make a deliberate effort to identify new ideas or practices into daily work.</li> <li>• <b>Seeking out and adopting improved work practices is an integral part of every division's work.</b></li> <li>• New ideas or practices can be adopted when they arise, but it's not necessary to seek them out.</li> <li>• I'm not sure</li> </ul>	Autonomy/ discretion	SOS/TFP T1 (B)
How should under-performance be dealt with in the Civil Service?	<ul style="list-style-type: none"> <li>• <b>Repeated poor performance should be systematically addressed, starting with targeted intermediate interventions.</b></li> <li>• Poor performance should be addressed on an ad hoc basis, depending on the individual's superior.</li> <li>• There is no need to address poor performance or impose consequences on under-performing staff.</li> <li>• I'm not sure</li> </ul>	Incentives/ monitoring	SOS/TFP T1 (B)

**Notes:** Questions and correct answers are taken from the SOS and TFP (T1 & T2) productivity training modules. For management practice survey modules, "correct" option indicates the response that corresponds most closely to a score of 5 on the survey-based management practice module (see Appendix C.1).

### C.3 Task Completion Data

We followed Rasul *et al* (2019) in collecting, standardizing, and coding the completion of bureaucratic tasks listed in organizations’ annual and quarterly progress reports. OHCS requires each organization to file these reports regularly, and prescribes a template for a table containing a list of all tasks, projects, outputs, and processes the organization has planned to undertake during the specified time period together with the actual actions taken during the reporting period. Sub-sections C.3.1 and C.3.2 describe this procedure in full detail. We replicate Rasul *et al*’s (2019) procedure with only two exceptions: task completion was single-coded in 2018 rather than double-coded, due to time and resource constraints; and one variable (Coordination Required) was coded as “choose all that apply” in 2018 as opposed to “choose one” in 2015. Sub-sections C.3.1 and C.3.2 are quoted verbatim from Appendices A.3 and A.4 of Rasul *et al* (2019), with the two exceptions described above.

#### C.3.1 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.”

### C.3.2 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*).” “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 any of the following that applied: “(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.”

## **C.4 Process Quality Data**

Working with OHCS and PRAAD, from May-October 2018 the research team randomly sampled an average of two open and two closed files from file indexes available at each record unit of organizations in the study sample. Files in our sample were indicatively opened starting in 2015 or later, and were not confidential nor related to personal issues. In total, 763 files were assessed, coming from 55 organizations. Randomly sampling across the four main line directorates contained in most organizations (Finance and Administration; Research, Statics, and Information Management; Policy Planning, Monitoring, and Evaluation; and Human Resource Management) and technical units, the research team audited 763 files from a total of 256 divisions across 55 organizations. For the analysis in this paper, we restricted the sample of files to those files from divisions that had at least one member participate in one of the training treatments and which were open (i.e. in use) during or after the training period. This represented a total of 286 files from 106 divisions in 37 organizations.

We worked closely with a retired civil servant and a PRAAD official to shape the sampling and coding procedure. During a piloting period, the tool was adjusted and improved to reflect the records management practices within Ghana's Civil Service. Sampled files were assessed by three Assistant Management Analysts from the Management Services Department (MSD) on two main areas: (i) quality of procedure, and (ii) quality of content.

### **C.4.1 Quality of Procedure**

Procedural quality is the level to which principal components of a file adhere to the general filing rules. The assessment tool collects information on whether the file ladder, folios, memos, minutes, letters and related documents are compiled correctly, following the public service rules, as Table A11 lays out.

The file ladder is an important element of a file, summarizing the file circulation within an organisation and expressing how valuable a file is. According to the general procedure, the file ladder should document each file circulation, specifying the date and the documents involved.

To guarantee a better accessibility to a file, all documents should be numbered consecutively, starting with folios from the opening of the file to the most recent ones. If actions are required, documents/letters should be minuted, dated and signed, clearly stating from whom the letters are coming and to whom they are directed to work on. The same procedure is applicable to memos and other relevant records in the file.

In addition to date and signature, incoming and outgoing correspondence requires specific stamps: the organizational (incoming) and dispatch (outgoing) stamp. Once a file has been passed on to other record officers or stored in the record office, it should not contain either duplicated and draft documents or misfiling, and miscellaneous items. The tool scored the procedure in handling government files by assessing: (i) the completeness of the file ladder, (ii) the consecutive organisation of folios within a file, (iii) the availability of minutes, memos and other relevant document, and (iv) the proportion of incoming/outgoing correspondence with dates, stamps and signatures. The level of procedural adherence of each component is measured on a 5-point scale from 1 (i.e. lower level of procedural quality) to 5 (i.e. higher level of procedural quality). Consequently, a score of 3 means that files present a fair level of procedural adherence. The tool also records whether a file presents discrepancies relate to duplicates, drafts, irrelevant materials and miscellaneous items or not.

**Table A11: Process Productivity - Quality of Procedure**

Component	Questions for Assessment	Score 1	Score 3	Score 5
File Ladder	How complete is the file ladder? (Each transfer should be documented.)	0-19%	40-59%	80-100%
	Does each step in the file ladder have dates (each transfer is associated with a date)	0-19%	40-59%	80-100%
Folios	Are folios within the file organised and numbered consecutively?	0-19%	40-59%	80-100%
Memo and Minutes	Where applicable, are minutes, memos and other necessary records present and complete (including from whom, to whom and signature)?	0-19%	40-59%	80-100%
Correspondence	What proportion of incoming correspondence has an organisational stamp/date/signature?	0-19%	40-59%	80-100%
	What proportion of outgoing correspondence has a despatch stamp/date/signature?	0-19%	40-59%	80-100%

**Notes:** See text for further details of assessment.

#### C.4.2 Quality of Content

As presented in Table A12, the quality of content measures the overall clarity of the file subject, assessing the file along six margins: (i) clarity of background to issues, (ii) clear courses of action available or taken, (iii) file organized in a logical flow, (iv) choices based on evidence in the file, (v) clarity on who should take action and (vi) proportion of relevant materials with clear deadlines. Likewise, the quality of content is measured on a 5-point scale from 1 to 5, where 1 indicates the lower content quality and 5 the higher content quality. Consequently, a score of 3 means that files present a fair level of content adherence.



**Table A12: Process Productivity - Quality of Content**

Component	Questions for Assessment	Score 1	Score 3	Score 5
Background to Issue	Background to issues	Very poor	Neither poor or good	Very good
Course of Action	Clearly outlining what courses of action are available or taken	Very poor	Neither poor or good	Very good
Logical Flow	The file is organised in a logical flow (where applicable, with an issue arising, being treated consecutively, and then resolved)?	Very poor	Neither poor or good	Very good
Choices	Choices are based on evidence in file	Very poor	Neither poor or good	Very good
Action Taken	Clarity on who should take actions at each stage	Very poor	Neither poor or good	Very good
Clear Deadline	What proportion of relevant materials have a clear deadline	Very poor	Neither poor or good	Very good

**Notes:** See text for further details of assessment.

## D Online Appendix: Sample Task Completion Data

**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
<b>MANAGEMENT AND ADMINISTRATION</b>				
<b>ADMINISTRATION</b>				
<b>LEGAL DIVISION</b>				
<b>Output 1: Competition Policy Developed and Approved</b>				
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	
<b>Output 2: Competition Bill prepared and approved</b>				
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

*Annotations in the image: 'Division name' points to 'LEGAL DIVISION'; 'Expected output' points to '1. Develop Draft Competition Policy'; 'Actual output' points to the 'ACTUAL OUTPUT' column for the first row.*

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

## E Online Appendix: Matching Observations Across Datasets

Linking these datasets required two exercises in observation matching: at the individual level, matching trained individuals to individuals surveyed in the endline; and at the divisional level, matching trained divisions to divisions observed in the endline survey and the task completion data. Since neither individuals nor divisions in Ghana’s Civil Service have unique identifiers that were widely in use at the time of the study, this posed a significant challenge.

To match individual trainees to endline survey observations, we compared CSTC’s training rosters to the organizational staff lists provided by each organization’s HR director. We first did this by matching individuals on their names in these two datasets, but since individual name spellings and name orderings vary across different datasets, this left most individuals unmatched. We then used a combination of algorithmic and manual matching methods using their primary demographic (i.e., gender, age, and education background) and work characteristics (i.e., organization, division, position, and tenure). In total we were able to match: 46 percent of individuals in the training data to the baseline survey data; 62 percent of individuals in the training data to the endline survey data; and 39 percent of individuals in the endline survey to the baseline survey. These rates are a combination of natural attrition (e.g. through individuals leaving the service) and matching failure, although we cannot disaggregate these two sources of attrition.

To match divisions across data sources, we used a similar mix of algorithmic and manual matching methods. However, divisional names are also written differently in different data sources and some have colloquial names that are different than their official ones, which complicated this task, even with access to organization charts and support from the Civil Service. In addition, division structures change relatively frequently, introducing natural attrition into our matching. In total we were able to match: 72 percent of divisions in the training data to the baseline survey data; 92 percent of divisions in the training data to the endline survey data; and 82 percent of divisions in the endline survey to the baseline survey.

In addition, we were unable to use one of the data sources we planned (records of individuals’ scores at promotion interviews subsequent to their training) as the absence of division-level identifiers in the promotion interview data resulted in a very low match rate which would have left

this analysis underpowered.

To evaluate how much these matching issues affect our main results, we produce balance tables to check whether there is differential attrition according to treatment status. We do so by running a regression of each outcome on a dummy for whether the division was exposed to T1, a dummy for whether the division was exposed to T2, and a series of dummies indicating the number of officials in the division attending the status quo training. The results are presented in Table A13.

The results suggest that divisions in T1 do not face differential attrition relative to divisions only exposed to the status quo training. This is true both for the endline survey attrition (panel A), for attrition regarding the output completion data (panel B), and for attrition regarding the process efficiency data (panel C). Moreover, within the endline samples, T1 is mostly balanced in terms of division characteristics that are expected to be exogenous to treatment - such as the share of women, average age, experience in the division, project composition, and so on. This suggests that selection bias due to attrition is unlikely to explain the results for individual training.

The results show that divisions exposed to T2 are more likely to be represented in the endline survey, task completion data, and administrative process quality data. At the same time, within each of these datasets, we do not observe significant imbalances between divisions exposed to T2 relative to non-T2 divisions in terms of the division characteristics in the survey data (panel A), in the task completion data (panel B), and in the process quality data (panel C). This suggests that the lower attrition rate between T2 and non-T2 divisions arises due to ease of reaching T2 divisions once they were trained, but that this attrition is uncorrelated with our dependent variables and thus is unlikely to be driving our results.

**Table A13: Attrition by treatment status**

	Mean dep. variable	T1 (individual)	T2 (team)	N
<b>Panel A: Endline survey attrition</b>				
Division in endline survey	0.854	-0.077 (0.052)	0.078* (0.046)	165
Division with non-missing management/culture scores	0.831	-0.023 (0.059)	0.050 (0.053)	165
Division with non-missing management scores	0.674	-0.108 (0.079)	0.147* (0.079)	165
Respondent knowledgeable about workplace and organization	0.619	0.116** (0.052)	-0.003 (0.056)	152
Interview morning	0.351	-0.001 (0.061)	-0.003 (0.062)	152
Interview duration	0.263	-0.058 (0.044)	-0.003 (0.046)	152
Share senior officials answering management/culture module	0.684	0.022 (0.093)	0.019 (0.085)	152
Share non-senior officials answering management/culture module	0.882	0.021 (0.067)	0.024 (0.064)	152
Share officials answering management/culture module	0.974	0.058* (0.030)	-0.027 (0.029)	152
Share female	0.516	0.047 (0.053)	0.108** (0.053)	152
Age	41.832	0.845 (1.041)	0.151 (0.999)	152
Share with undergraduate degree	0.823	-0.091** (0.044)	-0.079** (0.040)	152
Share with postgraduate degree	0.367	-0.072 (0.050)	-0.062 (0.054)	152
Mean years in civil service	13.414	-0.142 (0.992)	1.340 (1.105)	152
Mean years in organization	6.246	-0.689 (0.995)	0.090 (0.898)	152
Mean years in division	4.525	-0.043 (0.573)	-0.182 (0.535)	152
Number non-senior officials interviewed	7.796	-0.934 (1.311)	0.665 (1.251)	152
<b>Panel B: Output completion attrition</b>				
Division has output completion data	0.472	-0.050 (0.084)	0.170** (0.086)	165
Share of projects with irregular periodicity	0.683	0.080 (0.098)	0.003 (0.101)	84
Share of projects with multiple outputs	0.272	-0.175** (0.087)	0.015 (0.086)	84
Share of projects with external coordination	0.574	0.040 (0.105)	-0.047 (0.092)	84
Number of projects	7.464	-1.020 (1.962)	0.356 (1.997)	84
<b>Panel C: process efficiency attrition</b>				
Division with process productivity data	0.596	0.024 (0.085)	0.257*** (0.085)	165
Day info collected in field	67.782	-6.780 (5.635)	-3.529 (6.193)	106
Mean complete information on file	0.950	-0.019 (0.043)	0.066 (0.046)	106
# of endline files	2.698	0.185 (0.254)	0.343 (0.256)	106
# files closed at or before 2017	0.755	-0.228 (0.179)	-0.033 (0.241)	106
# files open in 2018-19	0.604	0.097 (0.197)	-0.013 (0.226)	106
Number of total files	3.453	-0.043 (0.285)	0.309 (0.348)	106

# F Online Appendix: Pre-Analysis Plan

## TRAINING FOR PRODUCTIVITY? EXPERIMENTAL EVIDENCE FROM GHANA'S CIVIL SERVICE \*

IMRAN RASUL      DANIEL ROGGER      MARTIN J. WILLIAMS<sup>†</sup>

PRE-ANALYSIS PLAN  
MAY 2019

### Abstract

In-service training is a potential channel through which civil services can endogenously improve performance and productivity. However, despite the large amount of resources governments devote to training bureaucrats, there is little rigorous evidence on its effectiveness. We partnered with the Government of Ghana to design a new productivity training module and deliver it on a randomized basis to civil servants in central government. The same training content was delivered through two treatment arms: an individual-level treatment in which the training cohort was comprised of officers of the same rank from different organizations, and a team-level treatment in which the training cohort comprised an entire team of people who work together. This pre-analysis plan details the context, intervention, and our core hypotheses and analysis.

---

\*Rasul: University College London and the Institute for Fiscal Studies [i.rasul@ucl.ac.uk]; Rogger: World Bank Research Department [drogger@worldbank.org]; Williams: Blavatnik School of Government, University of Oxford [martin.williams@bsg.ox.ac.uk]. We gratefully acknowledge funding from the International Growth Centre, Economic Development and Institutions, and World Bank i2i. We thank 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 Simprii-Duncan, and Liah Yecaló-Teclé for excellent research assistance, and Head of Civil Service Nana Agyekum-Dwamena, Mrs. Dora Dei-Tumi, members of our project steering committee, Andrew Wyatt, and Stefan Dercon for invaluable advice and guidance. Special thanks are due to over 80 civil servants from across the service who generously and energetically led trainings, conducted interviews, collected data, and shared their insights. This project received ethics approval from the Blavatnik School of Government Departmental Research Ethics Committee (SSD-CUREC1A-BSG-C1A-17-001). All errors are our own.

<sup>†</sup>Rasul: University College London and the Institute for Fiscal Studies [i.rasul@ucl.ac.uk]; Rogger: World Bank Research Department [drogger@worldbank.org]; Williams: Blavatnik School of Government, University of Oxford [martin.williams@bsg.ox.ac.uk].

# 1 Introduction

The effective functioning of government bureaucracies matters for growth and the supply of public goods. The human capital embodied in the civil service is central to this effective functioning. Governments and donors frequently use civil service training to improve the capacity of the public sector by increasing the human capital of civil service staff. While full figures are not available, a recent review of outstanding commitments by the World Bank indicated it had open investments of USD7.5 billion in these areas. However, despite such investments, there is little evidence that training programs maximize their potential impact on staff capacity (World Bank, 2016). This impact evaluation aims to utilize an internationally recognized training module that aims to maximize the productivity impact of standard civil service training curricula (that we will call ‘applied productivity training’ for the remainder of the document).

We partner with Ghana’s Civil Service Training Centre (CSTC) and Office of the Head of Civil Service (OHCS), as well as an international consultant, to design a new productivity training program for bureaucrats: the Training for Productivity (TFP) program. This training was intended to be more applied and action-oriented than CSTC’s existing productivity curriculum. We integrated the applied productivity training developed by TFP into Ghana’s Civil Service training routine, and delivered it through two treatment arms at full scale. Our evaluation aims to provide new evidence about the impacts of Civil Service training programs on public sector management and productivity.

This pre-analysis plan (PAP) describes the intervention and presents our core hypotheses and empirical analysis for our key outcomes. While we pre-specify our core analysis for the purpose of transparency, we choose not to pre-specify our full analysis (e.g. heterogeneous effects) or pre-specify every aspect of our treatment of the data (e.g. handling of attrition).

We believe this approach is the most appropriate for this study for two reasons. First, the project has presented a range of uniquely challenging data and measurement issues. We do not feel that we can make sensible decisions about how best to handle these issues without examining the data, and thus we do not pre-specify our handling of them. Second, to our knowledge this is the first large-scale RCT of a civil service training program in a low- or middle-income country. In this context there are fewer established theoretical predictions, descriptives, or stylized facts. This makes pre-specifying our extended hypotheses more speculative, and also increases the returns to exploratory empirical analysis. We therefore feel that pre-specifying our core hypotheses and analysis, but not every aspect of our handling of the data or our extended analysis, strikes the right balance between transparency and learning from the data. In our write-up, we plan to explain this for readers and delineate which aspects of our analysis were pre-specified and which were not.

The next section presents details of our research design, including the intervention, data, key outcome measures, and hypotheses. Section 3 presents our core empirical analysis. Section 4 concludes with a brief discussion of further issues and considerations.

## 2 Research Design

### 2.1 Context

Ghana’s Civil Service operates a ‘Scheme of Service’ (SOS) training schedule as the core of its in-service training. Officers undergo compulsory training before becoming eligible for promotion to the next grade. Since promotion is largely based on tenure, most officers go through an SOS training approximately once every three years, although the exact timing is not fixed. Trainings are conducted by the Civil Service Training Centre (CSTC), which operates under the supervision of OHCS.

SOS trainings are usually delivered at CSTC to a group of 10-40 civil servants of equivalent grade (rank), but from different organizations. Each SOS training lasts 10 days, and comprises three training sessions per day. The topics are diverse, but always include one day focused on productivity topics. Under the *status quo*, all trainings were designed and delivered by CSTC’s own trainers. SOS trainings were conducted as normal during the RCT (that our control are subject to); the only components that were varied or controlled are the content of the day of productivity training (treatment arm T1), and whether a follow-up training is delivered to trainees together with their work teams from their organization (treatment arm T2). We explain this in detail in the next sub-section.

### 2.2 Intervention

Table 1 summarizes our randomization strategy. The first arm (T1) generates a control group that receives the *status quo* SOS productivity training and a treatment group of individual officers that instead receives the individual-level TFP training module. Within each SOS training, for the day of training that is focused on productivity, trainees are randomized into the “old” *status quo* SOS module or the “new” TFP module. The TFP productivity trainings are delivered by CSTC’s existing pool of trainers. This will allow us to assess the impact of the TFP training content and style relative to the *status quo*.

At the end of the two-week SOS training, approximately 40% of trainees were randomly selected and informed that a day of productivity training would be organized for them and their entire work team of 5-20 people in the same division, including officers of *all* ranks – including, crucially, their director (boss). These team-based productivity trainings (T2) were conducted 3-6 weeks after the end of the SOS training, at the offices of OHCS. The training content and style was exactly the same as the TFP T1 training. The major difference between T1 and T2 is whether the TFP training is conducted with other officers of equivalent rank but from different organizations (T1) or whether the TFP training is conducted with officers of different ranks but from the same organization and division (T2). By construction, a more minor difference between the trainings is that in the T2 trainings, at least one participant had undergone the SOS training previously

(with either the “old” SOS productivity module or the “new” TFP T1 productivity module).

The T2 arm generates treatment and control groups at the division level. In combination with the T1 arm, this allows us to evaluate the effect of the applied productivity training when only individual trainings are undertaken, when the whole division is subject to treatment, and when both individuals and their divisions are treated.

The “pure” control are individuals who did not receive scheme of service trainings this year. The pure control can also contain officials who have recently joined the service, and so never participated in any training yet. Since officers attend scheme of service trainings every three years, many officials will have experienced the *status quo* training in past years, so the pure control is actually the group who had undergone training in previous (but not the current) year. While a pure control group who had never had any training would be desirable, such a group does not exist in Ghana’s civil service. Furthermore, the timing of selection into SOS trainings is not randomized, so that we cleanly identify the additional impact of being involved in TFP training relative to the *status quo* civil service training.

## 2.3 Data and Outcomes

We collected data both on the sub-sample of civil servants who participated in one of the trainings, and on the universe of senior-grade civil servants and their divisions. We discuss each in turn.

Our intervention sample comprises all civil servants who attended SOS trainings from March 2017 to March 2018 and whose training schedule includes one day of productivity training. This comprises professional-grade, university-educated officers between the ranks of Assistant Director IIB (and analogous grades) and Deputy Director (and analogous grades). The TFP T1 training was conducted at all SOS trainings where it was logistically feasible to do so. T1 could not be delivered at a small number of SOS trainings due to limited training space; this was driven purely by logistical course scheduling and room booking issues, not by anything that might feasibly be correlated with trainee characteristics. All individuals eligible for randomization into T1 were also eligible for randomization into T2.

During each TFP and SOS training session, we collected basic information about each session and attendance. We also gave pre- and post-tests to all trainees on their first and last days of the two week training, to test measures of learning gain from both the productivity and non-productivity training components.

Before and after the intervention, we conducted a baseline (2015) and endline (2018) survey of the universe of professional-grade civil servants in Ghana’s central government. This comprised approximately 3000 officers in each year, from 45 and 56 civil service organizations in 2015 and 2018, respectively. Ghana increased the number of ministries in 2017, hence the increase in the number of organizations. Full details of our survey methodology are discussed for the baseline in Rasul *et al.* (2019); the method for the endline was identical. The population was “senior” civil



servants (used by Ghana’s civil service to denote the professional grades of the service, not to indicate age or tenure in the service) working in the headquarters of ministries and departments. Before starting the survey in each organization, our survey team worked with the human resources department of the organization to create a roster of eligible individuals.

In addition, we digitized and coded a range of administrative data held by OHCS. The most important of these are quarterly and annual progress reports generated by each organization that detail their achievement of outputs against their plans for each period. We use these as a measure of output delivery, following Rasul *et al.* (2019). We also digitized records of the performance of officers in promotion interviews, for the sub-set of officers who underwent promotion interviews during our study period. These record the ratings given by the three-person interview panel to each officer on a structured scorecard.

During the baseline data collection, the team worked closely with the HR departments of individual organizations to create a roster of currently employed civil servants. However, due to privacy issues, the team only partially collected respondents’ names, and therefore a subset of officials could only be tracked across data collection rounds based on their primary demographic (gender, age, and educational background) and work characteristics (organization, division, position, and tenure). Between data collection rounds, there was a natural process of entry and exit into the civil service (sometimes into the public service, and other times into the private sector). We were therefore tasked with matching a distinct population of civil servants back to baseline. At endline, we asked officials about their relocations within the service and broader working history to try to support this matching effort. Overall, we were able to match roughly half of officials observed in the baseline with officials at endline. This figure represents a combination of natural churn within the service, as well as some issues with our ability to match officials across time.

At the division level, the Ghanaian Civil Service has changed its structure between 2015 and 2018, with new organizations being created, and existing organizations changing their internal structures. Overall, the internal structures have become more concentrated. Documenting which officials belong to which division is also complicated by the fact that informal nomenclatures do not always map directly on to *de jure* division names. Civil servants’ perceptions of what makes a division is not always a real division, but instead work teams that are functionally important within an organization’s activities. Therefore, we disaggregated the formal divisions in 2018 and 2015 to ensure a coherent match of their functions over time.

Table 2 summarizes the key outcome measures which we hypothesize the T1 and T2 interventions are likely to affect. Our outcomes are measured at three levels: individual; divisional (i.e. work teams, the largest sub-organizational unit); and the project or task, which can be uniquely assigned to divisions. Our analysis in the section below will take these three levels as our units of analysis. The outcomes are also sequenced between the short-run and longer-run. The former set of outcomes can be considered as intermediate outcomes that the treatments should impact: such as short-term learning post productivity training, perceptions of good management, and the

formulation of action plans post training.

We hypothesize that T1 and T2 will have a positive effect on these outcomes, relative to SOS training alone. We also hypothesize that the effect of T2 will be stronger than that of T1 for outcomes at division- or project-level, but not necessarily for individual-level outcomes. We might find positive complementarities between T1 and T2, or there could be substitution between the two. For the management practice outcomes, we expect to find positive effects on the autonomy/discretion indices, and null or weak effects on the incentives/monitoring indices.

These are our primary outcome measures. We will also explore the effects of the interventions on other variables collected from the management survey, administrative data coding, and data collection during the intervention, although we do not pre-specify these.

### 3 Empirical Analysis

Our first specifications will examine impacts on the outcomes that are measured at the individual level. To estimate individual-level impacts based on the factorial design described in Table 1, we will estimate regression models of the form:

$$y_{i,j} = \alpha_0 + \alpha_1 SOS_i + \beta_1 T1_i + \beta_2 T2_j + \beta_3 (SOS_i * T2_j) + \beta_4 (T1_i * T2_j) + \gamma \mathbf{X}_i + \delta \mathbf{Z}_j + \epsilon_{i,j} \quad (1)$$

where  $y_{i,j}$  is an outcome measure (learning gain from pre/post-training tests, or promotion interview score) for bureaucrat  $i$  in division  $j$ ,  $SOS_i$  is an indicator for individual participation in the standard SOS training (with the “old” productivity module),  $T1_i$  is an indicator for individual participation in the SOS training with the new TFP productivity module,  $T2_j$  is an indicator of division participation in the follow-up training,  $\mathbf{X}_i$  and  $\mathbf{Z}_j$  are vectors of individual and division characteristics, and  $\epsilon_{i,j}$  is an error term.

Given that the intervention will be organized as a randomized control trial,  $\beta_1$  through  $\beta_4$  provide ITT estimates of the main experimentally identified impacts:  $\beta_1$  is the ITT estimate of the impact of the new training module relative to the status quo training;  $\beta_2$  is the ITT estimate of division members undergoing the follow-up training;  $\beta_3$  is the ITT estimate of the divisional follow-up on individuals who participated in the normal SoS training; and  $\beta_4$  is the ITT estimate of the divisional follow-up on individuals who participated in our new modules. The quantities  $\beta_1$  and  $\beta_2$  are informative about the optimal content and delivery of trainings; we expect that both will be positive, but are agnostic about their relative magnitudes (for the individual-level analysis).  $\beta_3$  and  $\beta_4$  capture any complementarities or substitution between individual and group (division) level trainings; we are agnostic as to their relative magnitudes. As discussed previously, the interpretation of  $\alpha_1$  is limited given that much of the “pure” control group will have actually have undergone trainings in previous years and that the timing of selection into SOS trainings is

non-random.

There may be within-division spillovers, and we are directly exploiting this possibility through our second treatment. However, we will also assess the impact of trainings at the division-level, making the interpretation of our parameters distinct from the above conditional on our division-level analysis. That we are collecting detailed division-level information ensures that, allows us to explicitly test for such spillovers.

To estimate division-level impacts (on survey-based management quality and process productivity measures), we will estimate regression models of the form:

$$y_j = \alpha_0 + \alpha_1 SOS_j + \beta_1 T1_j + \beta_2 T2_j + \beta_3(SOS_j * T2_j) + \beta_4(T1_j * T2_j) + \delta \mathbf{Z}_j + \epsilon_j \quad (2)$$

where all variables and parameters are as in (1), except that the variables  $SOS_j$  and  $T1_j$  are substituted for  $SOS_i$  and  $T1_i$ , respectively.  $SOS_j$  and  $T1_j$  are defined as the percentage of officers in a division who attended the SOS training with “old” productivity module and SOS training with new TFP T1 productivity module, respectively. This is necessary since a division can have more than one officer attend each type of training. Defining these variables as fractions also allows us to adjust our parameter sizes to correct for differences in intensity of treatment: both  $\beta_1$  and  $\beta_2$  can be interpreted as the impact of treating an entire division with T1 or T2 (respectively). We also plan to examine non-linear effects of treatment saturation within a division, but do not pre-specify the form of this analysis since it will depend on the results of our core analysis. For these divisional outcomes, we expect that  $\beta_2 > \beta_1 > 0$ .

Finally, we estimate project-level impacts on output completion by estimating a regression of the form:

$$y_{n,j} = \alpha_0 + \alpha_1 SOS_j + \beta_1 T1_j + \beta_2 T2_j + \beta_3(SOS_j * T2_j) + \beta_4(T1_j * T2_j) + \delta \mathbf{Z}_j + \nu \mathbf{P}_n + \epsilon_{n,j} \quad (3)$$

where all variables are as in equation 2, except that  $y_{n,j}$  is a measure of output completion (coded on a scale of 1-5, then re-scaled to the interval [0,1]),  $\mathbf{P}_n$  is a vector of output characteristics, and the error term is  $\epsilon_{n,j}$ , reflecting that the unit of analysis here is the project (which are nested within divisions). As with the division-level analysis, we expect that  $\beta_2 > \beta_1 > 0$ .

We will cluster standard errors at the division level when undertaking individual-level regressions, and use robust standard errors when undertaking analysis at the division-level.

The extended analysis we will undertake will depend on what we learn from the results of this core analysis and from working with the data more generally. We will clearly note in our write-up which analysis was not pre-specified so that readers can interpret those results accordingly.

## References

- [1] RASUL,I, D.ROGGER AND M.J.WILLIAMS (2019) Management and Bureaucratic Effectiveness: Evidence from the Ghanaian Civil Service, mimeo UCL.
- [2] WORLD BANK (2016) Towards a New Approach to Capacity Building in Africa, mimeo, World Bank.

**Table 1: Treatment cross-randomization, at individual level**

<b>Division level treatment? (T2)</b>	<b>Individual level treatment? (T1)</b>		
	No, not selected for SoS	No, "old" SOS	Yes, T1
No	Pure control	SOS control	T1 only
Yes	T2 only	SOS + T2	T1 + T2

**Table 2: Main Outcome Variables**

Outcome	Description	Level
<i>Test-based assessment of short-term learning from Scheme of Service training sessions</i>	<i>Score on components of training learning evaluations conducted as short-term assessments at start and end of SOS and T2 trainings. T2 pre/post-tests are an abbreviated version of the SOS pre/post-tests, but with the same topical coverage</i>	<i>Individual</i>
SOS/TFP productivity content		
Perceptions of good management (autonomy/discretion-related)		
Perceptions of good management (incentives/monitoring-related)		
<i>Promotion interview performance</i>	<i>From report scoring promotion interview after SOS (if applicable)</i>	<i>Individual</i>
Overall score		
<i>Survey-based intermediate outcomes</i>	<i>From endline survey</i>	<i>Individual/divisional</i>
SOS/TFP productivity content (medium-term)		Individual
Formulating Action Plan helped officer think of new ideas (individual and division)		Individual/divisional
Tried to implement Action Plan		Individual/divisional
Implemented Action Plan		Individual/divisional
Impact of implementing Action Plan		Individual/divisional
Discussion and adoption of work-process improvements		Individual
<i>Survey-based measures of management quality</i>	<i>Adapted World Management Survey in endline survey, following Rasul et al. [2019]</i>	<i>Division</i>
Autonomy/discretion index z-score		
Monitoring/incentives index z-score		
Other practices index z-score		
<i>Process productivity</i>	<i>Measures of quality and efficiency of internal processes from file audit and letter tracing</i>	<i>Division</i>
Quality of record-keeping		
Speed and completeness of replies to internal letters		
<i>Output completion</i>	<i>For each output, its coded completion rate (1-5 scale, following Rasul et al. [2015])</i>	<i>Project/ task</i>