Bureaucratic Deliberation and Performance: Evidence from a Field Experiment in Benin*

Shan Aman-Rana† Leonard Wantchekon‡ Lazare Kovo§

March 27, 2026

Abstract

Bureaucratic performance depends not only on individual incentives and ability, but also on shared expectations about how others within the organization behave. This paper examines whether structured collective deliberation among bureaucrats can shift these expectations and improve performance. We conduct a randomized controlled trial with municipal administrations in Benin, where bureaucrats in treated municipalities participate in a series of deliberation meetings based on audit reports, bureaucrat surveys, and citizen feedback, while control municipalities continue with business as usual. Using administrative audit data from 2014 to 2019, we find that deliberation increases municipal performance by 6.8 percentage points (8.7 percent relative to the control mean), with larger effects in municipalities with lower baseline performance. Survey evidence shows that deliberation leads to more pessimistic beliefs about rule enforcement and corruption, alongside a decline in interpersonal trust within the bureaucracy. These effects are observed across the distribution of beliefs and trust and extend to bureaucrats who did not directly participate in the meetings. We find no evidence that the effects operate through increased monitoring, learning, or coordination. Instead, deliberation shifts attention toward performance-related issues and reshapes how existing information is interpreted within the organization. The joint pattern of improved performance and declining trust is consistent with settings in which interpersonal trust sustains collusive arrangements that weaken accountability. These results highlight structured deliberation as a low-cost and scalable tool for improving bureaucratic performance by reshaping shared expectations and strengthening internal accountability.

JEL codes: D73, H11, O1

Keywords: State capacity, performance, monitoring, accountability, audit, corruption, deliberation, Benin

---

*We are grateful to Nick Bloom, Guido Friebel, Virginia Minni, John Van Reenen, Raffaella Sadun, Megha Patnaik, Silvia Vannutelli, Gerard Padro i Miquel, Paola Sapienza, Ivo Welch, Arthur Zeyang Yu, Sandip Sukhtankar, Shaoda Wang and participants at seminars and workshops for comments and suggestions. Sanjog Rajiv and Leyao Wang provided excellent research assistance. Financial support from J-PAL Governance Initiative is gratefully acknowledged. All mistakes are our own.

†Department of Economics, University of Virginia. sa8ey@virginia.edu.

‡Department of Politics, Princeton University and African School of Economics. lwantche@princeton.edu.

§Department of Politics, Emory University. lkovo@emory.edu.

1 Introduction

State capacity, or the ability of governments to design and implement policy, is widely recognized as a foundational determinant of economic development (Besley and Persson, 2009, 2010; Evans, 1995; Evans and Rauch, 1999; Kohli, 2004). A central pillar of state capacity is bureaucratic performance: the day-to-day decisions and effort of frontline officials that determine how policies are translated into outcomes on the ground.

While a large body of work emphasizes the role of individual incentives and ability in shaping bureaucratic behavior, a growing literature highlights that performance is also deeply embedded in social and organizational contexts (Ashraf and Bandiera, 2018). Bureaucrats form beliefs about how others in the system behave: whether rules are followed, whether corruption is pervasive, and whether effort is the norm or the exception. These beliefs, in turn, shape their own choices. If these shared expectations about performance can be shifted, they may offer a powerful and relatively low-cost lever for improving state effectiveness.

Changing shared beliefs within organizations, however, is not straightforward. Truthful information about performance can update an individual's private beliefs. However, updating shared expectations about performance requires that information be interpreted in a common way across the organization. Meaningful shifts in expectations therefore typically require public or coordinated communication.

Deliberation based on truthful information can serve as a powerful institution for updating such beliefs (Bicchieri, 2006, 2016). It creates opportunities for actors to jointly reflect on their work environment, share experiences, unearth inconsistencies in their beliefs and develop a clearer understanding of collective behavior, helping to generate common knowledge.

In this paper, we conduct a randomized controlled trial with municipal bureaucracies in Benin to examine whether structured collective deliberation based on performance-related information can improve bureaucratic performance. We find that deliberation increases bureaucratic performance, while reducing interpersonal trust and leading to a negative update in bureaucratic beliefs about the rule of law. These patterns are consistent with environments in which interpersonal trust can sustain collusive arrangements that weaken accountability (Tirole, 1986). Disrupting such arrangements lowers trust but increases individual accountability and effort. The findings suggest that structured deliberation can be a

low-cost and scalable tool for improving bureaucratic performance in low-capacity settings.

Municipal bureaucrats in Benin play a critical role in local governance. They manage annual budgets totaling approximately 23 billion FCFA (USD 33 million) and are responsible for delivering essential services, including civil registration, water and sanitation, and local infrastructure such as roads, schools, and health centers. Given their central role in service delivery, improvements in bureaucratic performance have direct implications for citizen welfare.

We conducted a field experiment in direct partnership with the Government of Benin in 20 randomly selected municipalities (out of 77), spanning 8 of the country's 12 departments. Implementing the experiment at this scale, and in collaboration with the government, helps address common concerns about external validity that often limit the generalizability of randomized evaluations (Muralidharan and Niehaus, 2017). Structured collective deliberation exercises were implemented in ten treatment municipalities, while the remaining ten municipalities served as controls and continued operations without any deliberation activities.

While collective deliberation can take many forms, our design focused on creating the conditions necessary to shift bureaucrats' beliefs and shared expectations about performance. Three elements were critical to this approach. First, each session followed a clear and consistent agenda, with meetings designed to last approximately 90 minutes and structured to include both presentation of information and open discussion. This helped ensure focused discussions and comparability across sites (Goold et al., 2012). Second, deliberation was grounded in credible, third-party information that participants could not easily dismiss or question.[1] Third, the deliberation environment was carefully designed to foster open communication. By excluding elected officials such as mayors and placing bureaucrats in charge of guiding the conversation, the intervention limited political interference and created space for candid reflection. This also helped ensure that discussions remained focused squarely on service delivery and on the role of bureaucrats in driving performance.

In the treated municipalities three collective deliberation meetings were held

[1]The credibility of the information was central to the intervention. If participants viewed the information as biased or selectively presented, they could dismiss it, limiting its ability to influence their beliefs or behavior. This concern is particularly relevant in bureaucratic settings, where information can be distorted or downplayed in ways that weaken accountability (Banerjee et al., 2008; Duflo et al., 2013; Dhaliwal and Hanna, 2017).

at monthly intervals, beginning in September 2016. All bureaucrats in the municipality were invited to the meetings, and 58% attended at least one of the three sessions. Each meeting focused on a distinct theme: the first centered on the central government auditors' report, the second on survey results from bureaucrats, and the third on survey results from citizens. The audit reports were already available within the administrative system but were not typically discussed collectively or used to coordinate responses across the bureaucracy. The two surveys, conducted by the research team one month prior to the intervention, complemented these reports by providing comparable information on bureaucrats' experiences and citizens' perceptions of service delivery.

Meetings typically lasted an average of 95 minutes and adhered to the same structure: they began with a presentation of empirical results, followed by an open discussion on the topic at hand. Approximately 40% of the time was dedicated to formal presentations, while 46% was allocated to discussing the presented material, broadly consistent with the intended structure of the intervention.

To guide the empirical analysis and interpretation, we develop a simple framework in which workers choose effort based on beliefs about peers' effort and the monitoring environment determined by managers' beliefs about workers' effort. In this setting, deliberation operates as a shift in the information environment, leading to coordinated updates in beliefs across the organization. These updates affect effort through two channels: a monitoring (penalty) channel, operating through managers' beliefs about aggregate effort and the resulting monitoring intensity, and a peer effects channel, operating through workers' beliefs about others' effort.

The framework implies that the effect of belief updating depends on the nature of peer interactions. When effort is a strategic substitute for others' effort, negative updates in beliefs about peers' effort increase individual effort—even as trust declines—leading to improvements in performance. In contrast, when effort is a strategic complement to others' effort, negative updates in beliefs about peers' effort reduce both effort and performance alongside declines in trust. While changes in managerial beliefs may increase monitoring and thus raise effort, the overall effect depends on the nature of the peer effects channel. This provides a direct empirical mapping between belief changes, trust, and performance, allowing us to use how these variables move in the data to interpret the mechanisms underlying the effects of deliberation.

To measure bureaucratic performance, we use municipality-level administra-

tive audit data from 2014 to 2019. Our primary outcome is a panel of audit scores assigned annually to each municipality, expressed as a percentage of the maximum attainable score. Municipal audits are conducted by the central government oversight body, the National Commission for Local Finance (CoNaFiL), and constitute the primary instrument through which municipal performance is evaluated within the administrative system. The audits assess core dimensions of bureaucratic functioning, including record keeping, transparency in public procurement, and overall administrative management.[2] As such, the audit-based measure provides a comprehensive and policy-relevant measure for bureaucratic performance, capturing administrative practices that directly affect service delivery. Consistent with this, we observe a strong positive correlation at baseline between audit scores and an index of citizen experiences from Afrobarometer data (0.62, $p = 0.1$).

We find that deliberation improves bureaucratic performance. Municipal audit scores increase by 6.8 percentage points (8.7 percent relative to the control mean; randomization inference $p < 0.01$) in treated municipalities. These improvements are not driven by outlier municipalities but are broadly distributed across treated units. Effects are heterogeneous by baseline performance, with larger gains among municipalities that performed worse at baseline.

We complement audit data with survey measures of bureaucrats' beliefs, trust, and workplace interactions to examine the mechanisms through which deliberation affects performance. We find that deliberation leads to a substantial negative update in beliefs about rule enforcement and corruption: bureaucrats in treated municipalities are 0.4 standard deviations (randomization inference $p < 0.01$) less likely than those in the control group to report that there is rule of law in the municipality, as measured by a combined index of perceived corruption and rule enforcement. At the same time, deliberation reduces interpersonal trust within the bureaucracy by 0.35 standard deviations (randomization inference $p = 0.06$). These effects are not confined to the mean but are observed across the distribution of both beliefs and trust and extend to bureaucrats who did not attend the meetings.

In our setting, declines in trust, together with negative updates in beliefs about corruption and rule enforcement, are associated with improvements in perfor-

[2]CoNaFiL audits local government funds and evaluates municipal performance. Its oversight role has been described as transparent and effective in prior work (Bedasso, 2021a; *Decentralisation and Communal Development Support Programme*, 2021; Riedl and Dickovick, 2014; Behanzin et al., 2018).

mance. While a large literature emphasizes the role of trust and social cohesion in enhancing organizational performance (Brown et al., 2015; Keefer and Vlaicu, 2024), higher trust can also facilitate collusive arrangements that undermine accountability (Tirole, 1986; Milgrom and Roberts, 1988; de Janvry et al., 2023). Consistent with this latter view, we find that, at baseline, the trust index is negatively associated with both municipal audit-based performance (-0.55, $p < 0.1$) and citizen experience (-0.53, $p = 0.14$). These findings align with a broader literature on the downsides of social capital, where strong within-group cohesion can sustain inefficient or collusive outcomes (Basu, 1986; Dasgupta, 2005; Bloch et al., 2007; Alesina and Giuliano, 2014; Banfield, 1958; Gambetta, 1993; Portes and Landolt, 1996). These patterns suggest that improving performance may initially require disrupting existing social arrangements, with the potential for more efficient forms of trust to emerge over time.

We next examine whether deliberation operated through the monitoring channel. While managers exhibit similar negative updates in beliefs, we find no corresponding increase in monitoring-related interactions. Instead, deliberation appears to have shifted the content of interactions: discussions of audit findings and corruption increase, while other topics remain unchanged.

We also consider alternative mechanisms. We do not find evidence that the effects operate through a learning or coordination mechanism, as there are no changes in workplace interactions or collaborative behaviors following the deliberation meetings. The persistence of the performance effects over multiple years (though imprecisely estimated) is inconsistent with a short-lived salience effect. A third possibility is that deliberation improved performance by increasing bureaucrats' sense of voice or autonomy. While the meetings may have provided a platform for expression, the data are not consistent with an empowerment-based mechanism: deliberation is associated with a decline in interpersonal trust and no increase in workplace interactions, patterns that run counter to improved morale.

Taken together, these patterns point to a mechanism operating through coordinated belief updating. Deliberation brought dispersed information into a common forum, enabling bureaucrats to update their beliefs about peers and about the functioning of the organization. A decline in trust alongside improvements in performance is consistent with a setting in which effort choices are strategic substitutes, where reduced trust disrupts low-effort equilibria and increases individual accountability.

Finally we examined downstream effects on citizens using Afrobarometer data from 2014 and 2017. We find positive, though imprecisely estimated, effects on an index of citizen experiences with government services. Consistent with the audit-based results, effects are concentrated among municipalities with below-average baseline performance. The direction of the estimates—particularly in these lower-performing municipalities—together with the strong baseline correlation between audit scores and citizen experiences, is consistent with improvements in administrative performance translating into better service delivery outcomes.

Our paper contributes to the literature on state capacity building, focusing on the political economy of bureaucracies. There is a growing realization that bureaucrats that deliver services are an important determinant of state capacity and that there is a need to open up the black box of internal organization of public sector bureaucracies (Finan et al., 2017).[3] To the best of our knowledge our paper is the first that shows how bureaucratic performance can be improved from within the organization by reshaping shared expectations and internal accountability, without changes to formal incentives or authority.

There has been a small but growing body of recent work that shows the importance of the autonomy of bureaucrats (Rasul and Rogger, 2018; Duflo et al., 2018; Bandiera et al., 2020; Aman-Rana, 2023). Our paper contributes to this literature by showing that deliberation among bureaucrats can serve as a way to endogenize bureaucratic autonomy and control over their own accountability. The paper therefore, complements the standard approaches to reducing corruption in bureaucracies, (Becker and Stigler, 1974; Besley and McLaren, 1993; La Porta et al., 1999; Di Tella and Schargrodsky, 2003; Burgess et al., 2012; Niehaus and Sukhtankar, 2013; Corbacho et al., 2016; Debnath et al., 2023) and adds collective deliberation forums organized by bureaucrats themselves as an additional tool available to governments in developing countries to improve accountability. Moreover, the study also highlights the social aspect of state capacity and how accountability mechanisms can lead to a negative effect on organizational cohesion

[3]There have been studies that focus on investigating the incentives of bureaucrats (Banerjee et al., 2012; Callen et al., 2023; Ashraf et al., 2014; Khan et al., 2016; Bertrand et al., 2019; Khan et al., 2018; Khan, 2023; Dipoppa and Gulzar, 2023); and the way to recruit or promote pro-social and talented bureaucrats (Dal Bó et al., 2013; Ashraf et al., 2020; Jia et al., 2015; Riaño, 2021). There are studies that investigate how politicians or central agencies monitor and control bureaucrats (Olken, 2007; Lichand et al., 2016; Gulzar and Pasquale, 2017), others investigate the interaction between the politicians and the bureaucrats and how political incentives and turnover affect bureaucratic performance (Iyer and Mani, 2012; Gulzar and Pasquale, 2017; Akhtari et al., 2022; Callen et al., 2023; Dasgupta and Kapur, 2020).

with a positive impact on internal working of the organization.

Another strand of the literature that the paper contributes to is the rapidly expanding Economics, Political Science and Public Administration literature on public deliberation. Public reason and deliberation have been considered key for establishing the legitimacy of governments (Mendelberg, 2002; Parkinson and Mansbridge, 2012; Hadfield and Macedo, 2012).[4] While the literature on public deliberation has focused on how increased agency of the voters can strengthen the political process, we complement the literature by presenting empirical evidence showing that deliberation between bureaucrats can be an important determinant of public service delivery.

Finally, our paper contributes to the management literature. Previous studies have shown the importance of management skills in improving firm productivity (Bloom and Van Reenen, 2007; Bloom et al., 2013) and the existence of passive waste in public sector organizations (Bandiera et al., 2009). This paper proposes collective deliberation among bureaucrats as a management practice and a way to overcome passive waste.

2 Conceptual Framework

This section presents a simple framework illustrating how beliefs about effort shape behavior through two channels: monitoring by the principal (manager) and peer expectations among agents (workers). We define the environment as a standard principal-agent setting with imperfect observability of effort.

This conceptual framework serves to provide a lens through which the empirical results of the deliberation can be interpreted. As the following sections will show, these structured meetings serve as an information shock that alters both bureaucrats' beliefs about their peers' effort and managers' beliefs about collective diligence and compliance with rules. This coordinated communication

[4]Studies have shown that bringing voters and politicians together in town hall style meetings can help strengthen the political process, reduce the prevalence of clientelism, and increase parties' vote shares in Benin and Philippines (Fujiwara and Wantchekon, 2013; López-Moctezuma et al., 2022). Wantchekon and Guardado (Forthcoming) show that such meetings can lead to a more informed citizenry, higher electoral participation, lower policy polarization, and more importantly "ethical" voters who respond negatively to vote-buying. In Sierra Leone, Bidwell et al. (2020) find a positive impact of exposure to political candidates' debates on voting behavior, campaign spending, and politicians' performance. Similar positive effects have been observed in other contexts (Collier and Vicente, 2014).

changes the way monitoring and trust interact within the workplace, moving the organization away from its baseline equilibrium.

2.1 Environment

There is a continuum of workers indexed by $i \in [0, 1]$. Each worker chooses an effort level $e_i \geq 0$ after forming expectations about both the behavior of her peers and the monitoring intensity of the principal. The principal cannot observe individual effort directly but determines monitoring based on her belief about the average level of effort in the organization, $\hat{\mu}_P$.

Principal. The principal benefits from higher aggregate effort but incurs a cost of monitoring. When the principal holds more pessimistic beliefs about aggregate performance, she monitors more strictly, effectively increasing the scrutiny on all agents. For the purposes of the comparative statics, we treat $\hat{\mu}_P$ as an exogenous belief shift—arising from the deliberation intervention—that temporarily changes the principal's expectations before effort adjusts.

Workers. Each worker receives a fixed wage w, which does not depend on effort, and experiences an intrinsic disutility of effort represented by a twice-differentiable cost function $\tilde{C}(e_i)$, with $\tilde{C}'(e_i) > 0$ and $\tilde{C}''(e_i) > 0$. The worker faces an expected penalty $p(e_i, \hat{\mu}_P)$ that depends on both her own effort and the principal's belief about average effort, $\hat{\mu}_P$. This term summarizes the expected probability of detection and sanction conditional on monitoring intensity. We assume $p_{\hat{\mu}_P}(e_i, \hat{\mu}_P) < 0$, meaning the perceived risk of shirking rises as the principal's belief about aggregate effort falls.

Workers may also derive utility from how their behavior relates to that of their peers.[5] Let $\hat{\mu}_i$ denote worker i's belief about the average effort and rule compliance of her co-workers, and let $\Psi(e_i, \hat{\mu}_i)$ capture peer-related payoffs—the way these beliefs shape incentives through peer interactions and the perceived scope for mutual protection or exposure.

[5] A large literature shows that workers' effort responds to the behavior of their peers. See Kandel and Lazear (1992), Mas and Moretti (2009), and Bandiera et al. (2010, 2005) for theoretical and empirical evidence, and Ashraf and Bandiera (2018) for a review. These studies highlight how social interactions within organizations shape individual effort even in settings with weak formal incentives.

The utility of worker i is therefore given by

$$U_i(e_i, \hat{\mu}_i, \hat{\mu}_P) = w - \tilde{C}(e_i) - p(e_i, \hat{\mu}_P) + \Psi(e_i, \hat{\mu}_i). \tag{1}$$

The worker chooses e_i to maximize U_i, taking beliefs $\hat{\mu}_i$ and $\hat{\mu}_P$ as given. This leads to a first-order condition

$$\frac{\partial U_i}{\partial e_i} = -\tilde{C}'(e_i) - p_e(e_i, \hat{\mu}_P) + \Psi_e(e_i, \hat{\mu}_i) = 0,$$

which defines the worker's best-response effort function $e_i = e(\hat{\mu}_i, \hat{\mu}_P)$.

To connect the experimental intervention to the model, we define the deliberation meetings as a shift in the information environment. Structured deliberation grounded in third-party information serves as a coordination device that facilitates a joint update of the belief vector $(\hat{\mu}_i, \hat{\mu}_P)$, thereby shifting the organizational equilibrium.

The worker's best response effort, $e_i = e(\hat{\mu}_i, \hat{\mu}_P)$, responds to these shifts through two primary channels:

- **Monitoring (Penalty) Channel:** When beliefs about aggregate effort fall ($\hat{\mu}_P$ decreases), the principal may increase monitoring intensity. This raises the expected penalty for individual shirking, thereby increasing effort through pure deterrence: $\frac{\partial e_i}{\partial \hat{\mu}_P} < 0$.
- **Peer Effect Channel:** The impact of peer beliefs ($\hat{\mu}_i$) on individual performance depends on the strategic nature of the workplace.
  - **Strategic Complements ($\Psi_{e\hat{\mu}_i} > 0$):** In high-accountability settings, a belief that peers are working hard increases one's own incentive to perform, while a belief that others are shirking reduces it. In this regime, deliberation that reveals low compliance would lead to a decline in individual performance.
  - **Strategic Substitutes ($\Psi_{e\hat{\mu}_i} < 0$):** In lower-accountability environments, as a worker's belief in peer compliance ($\hat{\mu}_i$) rises, their own effort falls, as the risk of being singled out for shirking is low. Conversely, a negative update in perceived peer effort compels the worker to increase effort to avoid personal liability in a newly transparent environment.

These forces imply that a negative update in peer beliefs can generate different organizational responses.

Lemma 1. *In environments with strategic complements, negative updating of beliefs about peers' effort reduces both individual effort and trust in others. In contrast, when peer efforts are strategic substitutes, negative updating of beliefs about peers' effort increases individual effort even as trust in others declines. In this case, reductions in trust are associated with improved performance rather than deterioration.*

3 Context and research design

3.1 Context

Local administration. Benin's decentralization reform in 1999 established 77 municipalities with political and financial autonomy, placing local governments at the center of service delivery and public investment. Each municipality is governed by an elected municipal council, which in turn elects the mayor, who provides political leadership and oversees the organization of municipal administration. Day-to-day administration is carried out by a municipal bureaucracy responsible for implementing development plans, managing budgets, and delivering core services such as civil registration, infrastructure, water, and sanitation.

The administrative apparatus operates within a hierarchical structure led by the General Secretary (GS), who coordinates functions across departments. Below the GS are departmental heads (e.g., finance, planning, technical services), followed by "chefs," who manage administrative units and supervise lower-tier staff responsible for implementation. These staff include agents involved in civil registry, tax collection, and land-related services, many of whom interact directly with citizens in the course of service delivery. While the mayor retains authority over the organization of municipal administration—including task assignments, the allocation of personnel across departments, and the recruitment of certain contractual and support staff—formal hiring, promotion, and inter-municipality transfers for many bureaucrats remain governed by the central civil service. Municipal administrations thus operate under decentralized political authority but remain embedded in a centrally governed civil service system, where personnel decisions are only partially controlled at the local level (Riedl and Dickovick, 2014; Bedasso, 2021b).

Monitoring within municipalities follows this hierarchical structure, with supervisory responsibilities distributed across levels of the administration. At the same time, bureaucrats operate under fixed wages and limited performance-based incentives, implying that effort and performance depend not only on formal oversight but also on organizational norms and expectations.

Sources of municipal funds and their accountability. To support municipalities in meeting their budgets and development needs, the central government introduced the Support Fund for the Development of Municipalities (FADeC) in 2008 (Decree No. 2008-276). Transfers from FADeC are categorized into two types: general-purpose grants and investment grants. The general-purpose grant is allocated to cover recurrent expenditures, while the investment grant is designated for funding local public goods and services.

The investment grant consists of two components: an earmarked grant targeting specific policy areas (e.g., health and education), and a formula-based grant allocated based on socio-economic and performance indicators, including population, poverty levels, and prior administrative and financial performance. Administrative and financial performance is assessed by the National Commission for Local Finance (CoNaFiL), while socio-economic indicators are provided by the National Institute of Statistics (INSAE). FADeC investment grants constitute a major source of financing for local public goods, accounting for nearly 90% of municipal funding.

The National Commission for Local Finance (CoNaFiL), in collaboration with the Office of the Auditor General, conducts yearly audits across all municipalities. Multiple studies highlight the importance of CoNaFiL as a central oversight institution with substantial technical capacity and authority (Riedl and Dickovick, 2014; Behanzin et al., 2018; Bedasso, 2021a). According to the IMF's Public Investment Management Assessment (PIMA), Benin's regulatory framework for public investment is of relatively high quality and compares favorably to regional peers (Imbert et al., 2018).

These audits are based on a municipality's utilization of FADeC funds disbursed annually and are therefore conducted at the end of each administrative year (January–December). During these audits, auditors assess the management of funds received by municipalities, focusing on record-keeping, transparency in public procurement, compliance with procurement regulations, and the quality of

bureaucratic management. They also evaluate broader administrative processes, including adherence to budget adoption rules and financial reporting requirements.

Following the audits, auditors prepare detailed reports summarizing findings and assigning scores across multiple sub-components of performance. These reports are shared with municipal administrations and form the basis for performance-based allocations in subsequent FADeC transfers. Municipalities that perform well receive additional transfers in subsequent years, as outlined in the FADeC Procedural Manual. Consistent with this, evidence from our sample shows that a one percentage point increase in audit scores is associated with a 2 million XOF increase in investment transfers in the following year, equivalent to approximately 2 percentage points of total FADeC investment transfers (Table A.1). This indicates that municipalities face meaningful financial incentives to perform well on the evaluated criteria.

Audit institutions generate detailed information and recommendations on municipal performance and are intended to guide performance improvements. In some cases municipalities establish ad hoc committees or internal processes to follow up on audit recommendations (*Decentralisation and Communal Development Support Programme*, 2021; Bedasso, 2021a). However, there is no standardized or institutionalized forum for collectively interpreting audit findings or coordinating responses across different levels of the bureaucracy, limiting the extent to which audit information translates into coordinated administrative action. In this setting, mechanisms that facilitate collective deliberation may play an important role in enabling the effective use of performance information.

Despite this system of hierarchical monitoring and performance-based transfers, municipal performance remains uneven, and accountability remains a significant challenge. According to the World Economic Forum's Global Competitiveness Report (2015), several provisions in the law regarding citizens' rights to access public information and the Transparency Act's provisions concerning public procurement are not consistently enforced. This lack of adherence contributes to limited documentation and transparency in public procurement contracts, creating an environment conducive to corruption.

Various studies have indicated that local government officials in Benin are often co-opted by private vendors, further exacerbating the issue of accountability. In fact, research has shown that 63% of private businesses prefer utilizing public

procurement as a means for state capture (Ch et al., 2019; Canen et al., 2023). These findings underscore the extent to which private influence continues to undermine accountability in Benin's public procurement processes.

In addition to these governance challenges, there is substantial passive waste in local governments due to poor coordination and limited information flow across multiple layers of the bureaucracy. This results in low execution rates of municipal budgets. Between 2015–2018, budget execution at the municipal level averaged only 60 percent, with execution rates as low as 35 percent for capital expenditures (Bedasso, 2021a).

3.2 Experimental design

This study was implemented in collaboration with the Institute for Empirical Research in Political Economy (IERPE), the African School of Economics (ASE), and the Government of Benin. The intervention was conducted at scale within the municipal administrative system in 20 randomly selected municipalities (out of 77), spanning 8 of the 12 departments (see Figure 1).

The experimental design emerged from a series of consultations with the National Commission on Local Finances (CoNaFiL), with the aim of introducing structured deliberation as a tool for strengthening bureaucratic self-accountability. The intervention centers on creating a formal space for collective discussion of performance information, with the objective of improving how such information is interpreted, shared, and acted upon within the bureaucracy. The intervention did not introduce new incentives, but instead structured how performance-relevant information was brought into a common forum for discussion.

Municipalities assigned to the deliberation intervention convened deliberation sessions focused on performance and service delivery. *All* bureaucrats within a municipality were invited to participate, but attendance was not mandatory. The deliberation environment was standardized across sites to ensure comparability: bureaucrats retained control over the agenda—rather than politicians or citizens—and discussions were restricted to administrative performance and service delivery outcomes.

At the same time, bureaucrats were permitted to invite additional participants, such as municipal councilors, civil society representatives, central government officials, or media personnel, where relevant to the agenda of a given session,

allowing flexibility in participation while preserving bureaucratic ownership of the process.

Each session was designed to last approximately 90 minutes, with guidance to allocate time roughly equally between presentation of information and open discussion, a structure that was broadly reflected in observed time allocations (see Table 1).

Enumerators followed a structured observation protocol. They recorded attendance, including participants' roles and whether they belonged to the bureaucracy or to external groups (e.g., media or political actors). They also documented process measures, including the duration of presentations and discussions and the number of interventions (questions, comments, and criticisms). Following each session, enumerators provided standardized assessments of whether discussions were interactive and informative, and whether the overall atmosphere was collaborative or tense (see Table 1).

Figure I describes the timeline of the experiment. In August 2016, we compiled the information that would be presented during the deliberation sessions. In September, October, and November 2016, the first, second, and third collective deliberation meetings, respectively, were held. In December 2016, the endline survey of bureaucrats was carried out. We describe each of these steps below.

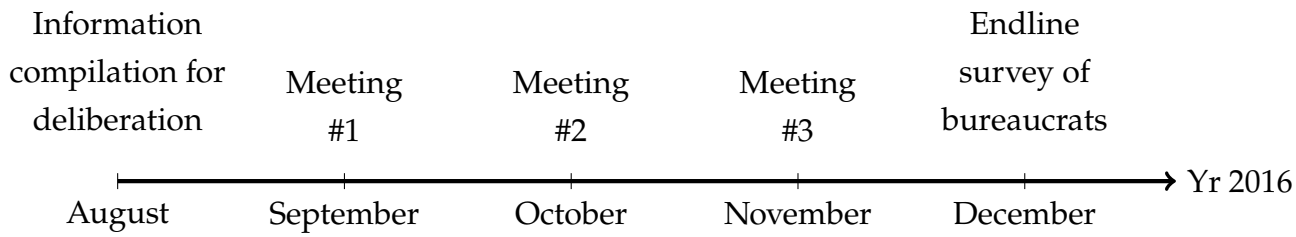


Figure I: Experiment timeline

173 out of the 299 surveyed bureaucrats in the treated municipalities reported to have attended at least one of the three meetings i.e. a take up rate of 57.86%.[6] Given this first stage, we focus below on the intent-to-treat (ITT) estimates, which can be interpreted as the average treatment effects corresponding to an approximately 60% complete implementation. The ITT is the relevant policy parameter, as participation in such deliberation exercises is inherently voluntary.

[6] Among the 299 surveyed, 116 (38.8%), 118 (39.46%) and 120 (40.13%) reported to have attended the first, second and third meeting respectively.

Generating information for the deliberation exercises. For the deliberation exercises to be meaningful, they were anchored in information that was externally generated and not produced within the deliberation process itself. In particular, a key component—municipal audit reports— was published by the central oversight body, CoNaFiL, and was already available through existing administrative processes. To provide a broader picture of constraints on public service delivery, these were complemented by two additional sources: an independent survey of bureaucrats and a survey of citizens conducted across all experimental municipalities, including those in the control group.

The audit reports and citizen survey results were presented at the municipality level, while bureaucrat survey responses were presented only in aggregate form, and respondents were informed of this at the time of data collection to mitigate concerns about social desirability.

The presentation materials consisted of slide-based summaries of key indicators and findings. In the audit-based meeting, information was organized along administrative and financial dimensions and presented as observations and associated risks raised by auditors; in the survey-based meetings, results were presented as summary statistics (e.g., shares of respondents reporting specific knowledge or experiences). Representative slides from the deliberation meetings are provided in Appendix C.

The survey of bureaucrats included questions on their knowledge of rules and regulations, their administrations and ongoing projects in the municipality, their perceptions of the role of the administration in improving education and health institutions, markets, maintenance of roads, and provision of water and sanitation, as well as the nature of their interactions with colleagues.

For the survey of citizens, a representative sample of 100 individuals was drawn from each municipality, with the exception of Ouinhi, where 112 citizens were surveyed. The survey focused on citizens' involvement in local politics and development affairs, their demand for better services, their perceptions of local governance and corruption, their trust in local institutions, and their satisfaction with the performance of local government. Across all meetings, the presentations emphasized descriptive statistics and factual findings rather than recommendations, after which participants engaged in open discussion.

Meeting #1: Deliberation on audit reports and recommendations. The first meeting, held in September 2016, focused on observations from the audit reports and the auditors' recommendations. On average, 31 participants attended, of whom 22 were bureaucrats (see Table 1). The General Secretary was present in all municipalities except one. Figure 2 provides details on participation across meetings.

Meetings lasted approximately 100 minutes on average, with 34% of the time devoted to presenting key findings and 47% to discussion. Figures 3 and 4 show the distribution of time across topics. Administrative functioning and financial autonomy received the greatest attention, both in presentation and discussion.

Participant engagement was highest in this meeting, with an average of seven contributions or suggestions. The number of criticisms and questions was also higher than in the second meeting, which focused on bureaucrat survey results. Enumerators assessed the meeting as both interactive and informative, and reported no tensions.

Meeting #2: Deliberation on the experiences of bureaucrats. The second meeting, held in October 2016, focused on results from the bureaucrat survey conducted across the 20 municipalities. On average, 32 participants attended, of whom 24 were bureaucrats (see Table 1; Figure 2). The General Secretary was present in all municipalities.

Meetings lasted approximately 92 minutes on average, with 42% of the time devoted to presenting key findings and 41% to discussion. Figure 3 shows that the presentation focused primarily on bureaucrats' understanding of rules and procedures, including public procurement processes and the availability of administrative and financial reports. As shown in Figure 4, discussion time was more evenly distributed across topics, with relatively limited time devoted to interpersonal relationships.

Engagement was more muted than in the first meeting: participants asked an average of three questions, and criticisms (6) exceeded suggestions (4).

Meeting #3: Deliberation on service delivery and the experiences of citizens. The third meeting, held in November 2016, focused on results from the citizen survey. On average, 34 participants attended, of whom 23 were bureaucrats (see Table 1; Figure 2). The General Secretary was present in all municipalities.

Meetings lasted approximately 94 minutes on average, with 42% of the time devoted to presenting key findings and 48% to discussion. As shown in Figures 3 and 4, discussions focused primarily on citizens' trust in the state, corruption, and perceptions of service delivery.

Enumerators assessed this as the least informative of the three meetings, with only 60% reporting that the session was informative. While the atmosphere was not reported to be tense, this meeting generated the highest number of criticisms and questions and the fewest suggestions.

Implementation and experimental integrity. Figure 5 presents balance tests using baseline survey data from bureaucrats. The figure includes both bureaucrat-level characteristics and municipality-level characteristics as reported by bureaucrats. All variables are standardized, and we plot mean differences by treatment status with 95% confidence intervals, with standard errors clustered at the municipality level. Overall, covariates are well balanced across treated and control municipalities at conventional significance levels.[7]

Figure 6 reports balance tests using administrative municipality-level data. Consistent with the survey-based measures, we find no systematic differences between treated and control municipalities at baseline.

All municipalities complied with the treatment protocol: deliberation meetings were conducted only in treated municipalities, with no such activities in control areas. Bureaucrats are typically not transferred across municipalities; nevertheless, 1.7% of surveyed bureaucrats in control municipalities (5 out of 296) reported attending at least one meeting in treated municipalities.[8] Given this limited crossover, and the incomplete take-up described above, all survey-based analysis is conducted using intention-to-treat estimates.

To minimize observer-induced distortions in behavior, discussions were not recorded, allowing interactions to proceed as naturally as possible in the field (Levitt and List, 2007; Spradley, 1980). Enumerators attended meetings unobtrusively and completed observation protocols only after sessions concluded. Participants were informed that there would be no real-time monitoring.

[7]Only two of the 16 covariates differ at the 10% level: indicators for the presence of yearly accountability sessions and procedures for equipment acquisition, both of which are lower in treated municipalities.

[8]Three bureaucrats from Za-Kpota reported attending the last two meetings; one bureaucrat each from Toviklin and Kouandé reported attending only the first meeting and all meetings, respectively.

In addition, bureaucrats were assured that survey responses would remain confidential and would be reported only in aggregate. No individual responses were disclosed during the meetings or afterwards, and access to the raw data was restricted to the principal investigators.

4 Data

We use two primary data sources to measure municipal performance and examine the mechanisms through which collective deliberation affects bureaucratic behavior: administrative audit data on municipal performance and an original endline survey of bureaucrats.

4.1 Municipal Audit-Based Performance

Our primary outcome is municipal performance, measured using annual audit reports produced by CoNaFiL. We digitize these reports to construct a municipality-year panel covering the period 2014–2019.

The audit reports assign scores across multiple dimensions of municipal administration. We focus on three core components—record keeping, transparency in public procurement, and overall administrative management—which capture key aspects of bureaucratic performance within the municipal administration.[9]

The record keeping component evaluates the maintenance of accounting records, the existence and functionality of archives, and the availability of quarterly and annual statements of financial and physical execution. The transparency component assesses compliance with public procurement regulations, the functionality of procurement control bodies, the exhaustiveness of published reports, and the traceability of fund transfers. Finally, the overall management component measures the effectiveness of the General Secretary in coordinating local services and the management of civil status documents.

We construct a municipal performance score defined as the total score achieved by a municipality expressed as a percentage of the maximum attainable score in that year, ensuring comparability across years. In a small number of cases

[9]Other audit components, including measures related to rule following by the local council, are analyzed separately and reported in Appendix Table A.8.

where a score is missing from the audit report for a given year, we recover the value using information from subsequent audit reports that document year-to-year performance changes. Additional details on the underlying components and data construction are provided in Appendix Section D.

Audit-based performance is positively correlated with citizens' reported experiences with local government, as measured using Afrobarometer data. Appendix Table A.2 shows that at baseline (2014), the correlation is 0.62 ($p = 0.01$).[10] This pattern is consistent with the audit-based measure capturing meaningful variation in administrative performance and service delivery. In addition, existing studies document features of the audit process that limit the scope for collusion between auditors and municipal bureaucrats (Riedl and Dickovick, 2014; Behanzin et al., 2018; Bedasso, 2021b; Imbert et al., 2018).

4.2 Bureaucrat Survey

To examine the mechanisms underlying the effects of deliberation, we conducted an endline survey of bureaucrats in December 2016, one month after the conclusion of the deliberation meetings. To ensure that the sample is representative of the municipal administration, we randomly sampled approximately 30 bureaucrats from the personnel roster of each municipality, irrespective of participation in the deliberation meetings. The final sample consists of 595 bureaucrats, with 299 from treated municipalities and 296 from control municipalities.

The survey collected information on bureaucrats' beliefs, attitudes, workplace interactions, and meeting practices within the administration. We measure beliefs about rule enforcement using responses to questions on whether procedures governing public procurement, staff recruitment, and the acquisition of office equipment are followed. Beliefs about corruption are measured by asking respondents to assess the prevalence of corrupt practices among the municipal council, private vendors, and their colleagues. These measures are combined into a standardized belief index, where higher values correspond to stronger perceived rule compliance and lower perceived corruption.

We also measure trust and workplace interactions. Trust is measured using responses on trust in the municipal council and in different types of colleagues

[10]The construction of the citizen experience measure is described in section 7.

within the administration, and is aggregated into a standardized index. Workplace interactions are captured through the reported frequency of collaborative behaviors, including working together to solve problems, organizing professional activities, resolving conflicts directly, and supporting colleagues when needed. To assess information exchange within the workplace, we use responses on the frequency with which staff meet to discuss work-related issues. These measures are combined into a standardized index

In addition, the survey collects information on meeting practices within the administration. Respondents report the frequency of internal meetings, as well as the extent to which these meetings involve discussions of key administrative issues, including audit findings, corruption, and service delivery. These measures allow us to distinguish between changes in the quantity of meetings and changes in their content, and provide proxies for monitoring-related interactions within the workplace.

Additional details on survey questions and variable construction are provided in Appendix Section E.

5 Effects on Performance

We begin by presenting estimates of the impact of deliberation on overall performance outcomes, before turning to mechanisms.

5.1 Estimation Strategy

Because the treatment is assigned at the municipality level and covers the entire bureaucracy, performance is measured using the municipality's audit score, expressed as a fraction of the maximum attainable score. Using municipality-year panel data, we employ a difference-in-differences (DiD) strategy that compares the evolution of performance in treated municipalities—those assigned to the deliberation meetings—relative to control municipalities before and after the intervention. Specifically, for municipality m in year t, we estimate:

$$y_{mt} = \alpha_m + \alpha_t + \beta(T_m \times \text{Post}_t) + \varepsilon_{mt} \tag{2}$$

where y_{mt} denotes the municipality's audit-based performance score (expressed as a percentage of the maximum possible score). The variable T_m is an indicator equal to one if the municipality was assigned to the deliberation arm and zero otherwise. Post_t is an indicator equal to one for the post-treatment period (2017–2019) and zero for the pre-treatment period.[11] The terms α_m and α_t denote municipality and year fixed effects, respectively, absorbing time-invariant municipal characteristics and common temporal shocks. Standard errors are clustered at the municipality level, corresponding to the level of treatment assignment (Abadie et al., 2017).

The coefficient of interest, β, captures the intention-to-treat (ITT) effect of the deliberation intervention. Given the relatively small number of clusters (20 municipalities), conventional cluster-robust inference may be unreliable. We therefore complement these estimates with p-values based on 1,000 permutations of treatment assignment across municipalities. In each permutation, the treatment indicator is reassigned while preserving the number of treated units, and the difference-in-differences specification is re-estimated. This generates a placebo distribution of the coefficient on the treatment $\times$ post interaction under the null of no effect. The p-value is computed as the fraction of permuted estimates that are at least as large in absolute value as the estimate obtained under the observed assignment. This approach exploits the known assignment mechanism and remains valid in difference-in-differences settings with few clusters (Fisher, 1935; MacKinnon and Webb, 2020).

5.2 Results

Table 2 reports the results. We find that collective deliberation leads to an improvement in municipal audit scores. As shown in Column (2), the average municipal audit score in treated municipalities increases by 6.8 percentage points relative to the control group in the post-treatment period (2017–2019). Relative to a control mean of 77.8%, this corresponds to an 8.7% increase in audit-based performance. The estimate is statistically significant at conventional levels using cluster-robust standard errors ($p < 0.05$) and remains significant when assessed using randomization inference ($p < 0.01$). This suggests that deliberation can improve perfor-

[11] Although the intervention was implemented between September and November 2016, we define the post-treatment period as starting in 2017, the first full administrative year following the intervention.

mance even in the absence of changes to formal bureaucratic incentives. Given that audit scores are positively correlated with citizen experiences and are based on independent assessments, these estimates reflect meaningful improvements in administrative performance.

We next assess whether the results are driven by a small number of municipalities. Appendix Figure B.7 presents treatment effects for each of the treated municipalities. The improvements are not concentrated among a few outliers but instead reflect more broadly distributed gains across treated municipalities.

We then examine heterogeneity by baseline performance, measured by whether municipalities fall above or below the mean audit score in the pre-treatment period. The positive impact of deliberation is concentrated among municipalities with below-mean baseline performance. In contrast, municipalities with above-mean baseline performance show no significant additional gains, consistent with more limited scope for improvement at higher levels of performance.

To better understand the sources of the aggregate effect, we decompose the audit score into its three core sub-components. While our main specification uses a composite index to improve statistical power and address multiple testing concerns (Anderson, 2008a), examining the underlying components provides insight into which dimensions of administrative performance are most affected.

While estimates at the sub-component level are less precise, consistent with the higher statistical power of the composite index, the results indicate improvements across multiple dimensions of administrative performance. Record keeping scores increase by 9 percentage points in treated municipalities (12% of the control mean), reflecting improvements in accounting and documentation practices. Procurement transparency improves by 6 percentage points (7.4% of the control mean), consistent with greater adherence to procurement procedures. The largest point estimate is observed in overall management, which increases by 14 percentage points (17.2% of the control mean), capturing improvements in the coordination of municipal services.

6 Mechanisms

Guided by the conceptual framework, we examine how deliberation affected beliefs, trust, monitoring and workplace interactions within the organization. The

framework highlights two channels through which belief changes may affect performance: a peer effects channel, operating through workers' beliefs about each other, and a monitoring channel, operating through changes in managerial beliefs and oversight.

6.1 Peer Effects: Beliefs and Trust

Using endline survey data from bureaucrats, we estimate the following specification for each individual bureaucrat i in municipality m:

$$y_{im} = \gamma + \phi T_m + u_{im}. \tag{3}$$

The outcome y_{im} includes measures of bureaucrats' beliefs about corruption practices among councilors, other bureaucrats, and city hall vendors, as well as their beliefs about the enforcement of rules governing the acquisition of office equipment, public procurement, and recruitment. We also examine outcomes related to bureaucrats' trust in councilors, office colleagues, colleagues within the same department, and city hall colleagues more broadly.

For both sets of outcomes—beliefs and trust—we construct standardized indices such that coefficients can be interpreted in standard deviation units relative to the control group. These composite indices improve statistical power and address concerns related to multiple hypothesis testing (Anderson, 2008a). For the aggregate index, which combines beliefs about corruption and rule enforcement, corruption measures are reverse-coded so that higher values indicate stronger perceived enforcement and lower perceived corruption.

Table 3 reports the effects of deliberation on bureaucratic beliefs. Deliberation leads to a large and statistically significant negative update in beliefs about the rule of law. The aggregate index decreases by 0.40 standard deviations in treated municipalities relative to control ($p = 0.006$; RI $p = 0.006$), indicating a substantial decline in perceived rule compliance.

The sub-components show a consistent pattern. On the corruption dimension, deliberation increases perceived corruption across actors. The largest effect is for city hall vendors: bureaucrats in treated municipalities are 0.47 standard deviations more likely to report corruption among vendors relative to the control group ($p < 0.01$; RI $p < 0.01$). Estimated effects for corruption among councilors

and other bureaucrats are also positive, though smaller in magnitude and not statistically significant at conventional levels.

On the enforcement dimension, deliberation reduces beliefs that administrative rules are followed. Bureaucrats in treated municipalities are 0.27 standard deviations less likely to report that procurement rules are enforced ($p < 0.10$; RI $p < 0.10$). Estimated effects for equipment acquisition procedures and hiring rules are similar in magnitude (approximately -0.27) but are not statistically significant at conventional levels.

Figure 8 shows that the effects are not confined to the mean but are across the belief distribution, with negative effects from the 10th to the 60th percentile. This suggests a broad-based shift in beliefs.

More importantly, Figure 9 further shows that these belief changes extend throughout the municipality, including among bureaucrats who did not attend the meetings. The distributions of beliefs for bureaucrats in treated municipalities who attended the meetings and those who did not are statistically indistinguishable (Kolmogorov-Smirnov (K-S) test p-value>0.10).

Table 4 reports the effects of deliberation on bureaucratic trust. The aggregate trust index is 0.35 standard deviations lower in treated municipalities relative to the control group ($p = 0.039$; RI $p = 0.060$), indicating a substantial reduction in interpersonal trust following deliberation.

The decline is observed across relational categories. Trust in councilors is 0.38 standard deviations lower in treated municipalities ($p = 0.031$; RI $p = 0.020$), representing the largest effect. Trust in departmental colleagues is 0.28 standard deviations lower than the control group and statistically significant ($p = 0.064$; RI $p = 0.060$). Estimates for trust in office colleagues and city hall colleagues more broadly are similar in magnitude and direction but are not statistically significant at conventional levels.

Consistent with the results for beliefs, the effects on trust are not confined to the mean. Figure 11 shows negative effects across the distribution of the trust index, although the confidence intervals are wider. Figure 12 further indicates that these declines extend throughout the organization, including among bureaucrats who did not attend the meetings. The distributions of trust in treated municipalities who attended the meetings and those who did not are statistically indistinguishable (Kolmogorov-Smirnov (K-S) test p-value>0.10).

The joint pattern of results—improvements in performance (Table 2), negative updating of beliefs (Table 3), and a decline in trust (Table 4)—is consistent with an environment in which efforts are strategic substitutes, as characterized in Lemma 1. In such settings, a negative update in beliefs about peers' compliance increases individual effort even as trust declines.

This interpretation is consistent with theoretical work highlighting how interpersonal relationships can sustain collusive equilibria within organizations. When monitoring is imperfect, trust and social cohesion can facilitate implicit agreements that tolerate low effort or rule violations (Tirole, 1986; Milgrom and Roberts, 1988). In such environments, higher trust reduces the perceived risk of being singled out for shirking, weakening individual incentives to exert effort. Conversely, a decline in trust disrupts these informal arrangements, increasing individual accountability and inducing higher effort.

To assess whether the Beninese municipal bureaucracy operates in such an environment, we examine the relationship between trust and performance at baseline. We consider two measures of performance: municipal audit-based performance and a citizen experience index constructed from the Afrobarometer. The results are presented in Figure B.9a and Figure B.9b. Consistent with the mechanism above, the trust index is negatively correlated with both measures of performance (p-value < 0.10 for audit-based performance and p-value $= 0.14$ for citizen experience). While a large literature emphasizes the role of trust in facilitating cooperation and improving organizational performance (Brown et al., 2015; Keefer and Vlaicu, 2024), a parallel literature highlights that social capital can also sustain inefficient or collusive outcomes (Basu, 1986; Banfield, 1958; Gambetta, 1993; Alesina and Giuliano, 2014). In this setting, higher trust appears to be associated with tolerance of low effort rather than productive cooperation, and reductions in trust reflect a shift away from such equilibria toward greater individual discipline.

6.2 Monitoring Channel

The conceptual framework in Section 2 highlights a monitoring (penalty) channel through which deliberation may affect effort. A negative update in beliefs about aggregate performance can induce the principal to increase monitoring intensity, thereby raising the expected penalty for shirking and increasing effort. Under

this mechanism, we would expect changes in managerial beliefs to translate into increased monitoring in practice.

We begin by examining whether deliberation altered managerial beliefs. A natural limitation is that managers constitute a smaller share of the bureaucracy, and we therefore have less statistical power to detect effects for this group. That said, Figures B.10–B.11 show that managers exhibit similar directional, though statistically insignificant, negative updates in their beliefs about the rule of law. Managers also do not significantly update their beliefs about corruption among councilors.

A similar pattern emerges for trust (Figures B.12–B.13), although the effects are more muted. In particular, managers do not exhibit declines in trust in city hall colleagues. Overall, the effects for managers are attenuated across outcomes, but we cannot reject that they are statistically similar to those for workers.

In the conceptual framework, the monitoring channel requires not only changes in managerial beliefs but also that these translate into stronger oversight. We therefore examine whether deliberation affected monitoring-related interactions within the administration. While we do not directly observe monitoring intensity or specific checks undertaken by managers, we use bureaucrats' reported meeting attendance as a proxy for exposure to accountability-related activities within the administration.

This proxy is informative under the institutional feature that monitoring in this setting is largely mediated through formal and semi-formal interactions where performance is evaluated and expectations are communicated. Under this structure, an increase in monitoring intensity would plausibly manifest as an increase in the frequency of such interactions, or in bureaucrats' participation in them.

Table 5 reports the effects of deliberation on the number of meetings attended by bureaucrats in the past three months, both including and excluding the deliberation meetings. When deliberation meetings are included, the estimates are small and statistically insignificant, indicating no meaningful change in overall meeting attendance. When deliberation meetings are excluded, meeting attendance is lower in treated municipalities. This pattern suggests that deliberation did not lead to an expansion of monitoring-related interactions, and if anything, substituted for routine administrative meetings rather than increasing overall exposure to oversight.

We next examine whether deliberation affected the content of meetings. While

the overall number of meetings did not increase, Figure 13 shows that their content shifted. In treated municipalities, discussions of audit reports and corruption increase relative to the control group, with no corresponding change in discussions related to taxes or finances, other administrative challenges, or broader issues of governance (Figure 14). This pattern suggests that deliberation redirected attention toward performance deficiencies, particularly in domains highlighted by audit information, rather than generating a broad-based expansion in monitoring or administrative oversight.

Taken together, while managerial beliefs move in a direction consistent with the monitoring channel, we do not observe corresponding changes in our proxies for monitoring behavior. This suggests that belief updates at the managerial level do not translate into stronger oversight along the dimensions we are able to measure.

We cannot rule out the possibility that monitoring changed along margins not captured by our data. However, this interpretation appears less likely in the institutional context we study, where formal sanctions are limited. Bureaucrats face few explicit penalties beyond career concerns and are typically neither dismissed nor reassigned across municipalities, constraining the scope for monitoring-based discipline. In such a setting, even if managerial beliefs become more pessimistic, the ability to translate these into meaningful increases in monitoring is likely to be limited.

These results, combined with those in subsection 6.1, indicate that the performance gains are unlikely to operate primarily through the monitoring channel. Instead, taken together with the evidence on beliefs and trust, they point to the peer effects channel as the more dominant mechanism driving the improvements in performance.

6.3 Alternative Explanations

Performance Salience and Hawthorne Effects. We examine the temporal dynamics of the treatment effect to assess whether the results reflect short-term salience or persistent changes in behavior. Figure B.8 presents event-study estimates for each year from 2014 to 2019. Each coefficient is estimated using comparisons across 10 treated and 10 control municipalities, which limits precision. In the pre-treatment period (2014–2016), the magnitude of the differences between treated and control municipalities are indistinguishable from zero. Following the

intervention, audit scores in treated municipalities increase in 2017 and remain elevated through 2018 and 2019. While the estimates are imprecise, the absence of pre-trends and the pattern of the magnitude of the post-treatment coefficients suggest that the observed gains are unlikely to be driven solely by short-term salience. Results suggests a shift to a new behavioral equilibrium rather than a transient response to attention.

Learning and Coordination. One alternative interpretation is that deliberation improved performance through a learning or coordination channel. We assess this possibility along two dimensions.

First, while the deliberation meetings involved information sharing, they were not designed as technical training workshops and did not provide procedural instruction or skill-building. The content of the meetings focused on presenting audit findings and survey results rather than teaching bureaucrats how to perform specific administrative tasks. That said, it remains possible that deliberation facilitated informal learning or improved coordination among bureaucrats following the meetings.

Second, we examine whether deliberation led to greater coordination or collaboration within the administration after the meetings. Figure 15 shows that one month after the deliberation ended, there is no evidence of an increase in workplace interactions, including collaboration, problem-solving, or coordination among bureaucrats. This suggests that deliberation did not lead to sustained changes in day-to-day coordination or collective activity.

Moreover, a learning or coordination-based mechanism would not readily explain the observed decline in interpersonal trust (Table 4) or the negative updating of beliefs regarding rule of law (Table 3). If improved performance were driven by enhanced collaboration or knowledge sharing, we would expect trust to remain stable or increase.

Bureaucratic Autonomy and Empowerment. A third possibility is that deliberation improved performance by increasing bureaucrats' sense of voice or autonomy. While the meetings may have provided a platform for expression, the data are not consistent with an empowerment-based mechanism. Deliberation is associated with a decline in interpersonal trust (Table 4) and no increase in cooperative workplace interactions, patterns that run counter to improved coordination or morale.

7 Downstream effects on citizen experience

To assess downstream effects on public service delivery, we use data from the Afrobarometer, a nationally representative cross-sectional survey. We use the 2014 (Round 6) and 2017 (Round 7) waves to correspond to the pre- and post-treatment periods. While the audit data span 2014–2019, Afrobarometer data are available only for 2014 and 2017, capturing short-run changes in citizen experience. The survey provides independent measures of citizens' interactions with local government, including access to services and exposure to corruption.

We construct a *citizen experience index* that aggregates responses across multiple dimensions of service delivery, including perceived living conditions, perceptions of corruption, and direct experiences in accessing services such as official documents, water, and sanitation, as well as whether they paid bribes. Variables reflecting worse outcomes (e.g., corruption or difficulty accessing services) are reverse-coded so that higher values indicate better experiences. The index is standardized, and coefficients are expressed in standard deviation units relative to the control group. We estimate effects using the same specification as in Equation 2, replacing the outcome with citizen experience.

Table A.10 presents the results. Following deliberation, citizen experience increases by 0.17 standard deviations, though the estimate is not statistically significant at conventional levels. Consistent with the audit-based results, we observe stronger improvements among municipalities with lower baseline performance. In these municipalities, citizen experience increases by 0.29 standard deviations ($p < 0.1$, RI $p = 0.185$) (Table A.11). Estimates for individual components of the index (Table A.12) follow a similar pattern.

The direction of the estimates particularly in lower-performing municipalities and the strong baseline correlation between audit scores and citizen experiences (Table A.2) is consistent with improvements in administrative performance translating into better service delivery outcomes.

7.1 Discussion: External Validity and Barriers to Adoption

A natural question is whether structured deliberation can be replicated in other bureaucratic settings and, if so, why similar practices do not emerge organically. While the intervention is low-cost and yields meaningful improvements in per-

formance, its adoption may be constrained by organizational and institutional frictions.

First, the results suggest that the pre-intervention environment is characterized by high levels of interpersonal trust and shared expectations that limit individual accountability. In such settings, informal arrangements can sustain low effort or weak enforcement of rules. Initiating a process that reveals collective shortcomings may therefore be both costly and professionally risky, as it disrupts existing workplace norms. As a result, no single bureaucrat has the incentive to coordinate such a shift, making external facilitation important for moving the organization toward a higher-performance equilibrium.

Second, the effectiveness of deliberation relies on the credibility of the information environment. In our setting, discussions were anchored in objective, third-party data—namely audit reports and independent surveys of bureaucrats and citizens—which are difficult to dismiss and help align beliefs across participants. Importantly, audit reports already existed within the administrative system, typically as hard-copy documents, but were not routinely collated or used to facilitate collective discussion within the bureaucracy.

Deliberation therefore operated not just by generating new information, but by structuring how existing information was brought into a common forum for interpretation and discussion. In the absence of such structured aggregation and presentation, similar meetings may fail to generate meaningful belief updating and instead devolve into unproductive discussions.

Third, implementation requires administrative and political support. In our context, deliberation was authorized by the hierarchy but structured to allow bureaucrats to engage in candid discussion, including by limiting the role of elected officials. Replicating this balance—top-down authorization combined with protected space for internal deliberation—may be challenging in more politicized or hierarchical environments.

Taken together, while structured deliberation is financially inexpensive and scalable, its effectiveness depends on institutional conditions that facilitate credible information sharing, protect open discussion, and overcome organizational inertia.

8 Conclusion

This paper studies whether structured collective deliberation among bureaucrats can improve organizational performance. In a field experiment across municipal administrations in Benin, we show that deliberation leads to meaningful improvements in audit-based performance, with effects concentrated among lower-performing municipalities.

To understand the mechanisms underlying these effects, we examine how deliberation shapes beliefs, trust, and interactions within the bureaucracy. We find that deliberation induces a substantial negative update in beliefs about rule enforcement and reduces interpersonal trust. These changes are broad-based and extend beyond those who directly participate in the meetings. At the same time, we find no evidence of increased monitoring or greater coordination within the administration.

The joint pattern of results is consistent with an environment in which effort choices are strategic substitutes. In such settings, higher interpersonal trust can sustain informal arrangements that tolerate low effort, while a reduction in trust increases individual accountability and induces higher effort. Our findings therefore highlight the importance of belief-based and peer-driven mechanisms in shaping bureaucratic performance, particularly in contexts where formal incentives and monitoring are limited.

More broadly, the results suggest that relatively simple institutional changes in how information is shared and discussed within organizations can have meaningful impacts on performance. In our setting, relevant performance information—such as audit reports—was already available but not routinely used to facilitate collective reflection. Structured deliberation provided a mechanism to aggregate this information and create common knowledge around performance.

These findings point to deliberation as a potentially scalable and low-cost tool for improving state capacity. At the same time, the intervention highlights important trade-offs. While performance improves, this comes alongside a decline in interpersonal trust, raising questions about the longer-run implications for organizational cohesion. Understanding how to design deliberation processes that enhance accountability without undermining coordination remains an important area for future research.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge, "When should you adjust standard errors for clustering?," Technical Report, National Bureau of Economic Research 2017.

Akhtari, Mitra, Diana Moreira, and Laura Trucco, "Political turnover, bureaucratic turnover, and the quality of public services," *American Economic Review*, February 2022, *112* (2), 442–493.

Alesina, Alberto and Paola Giuliano, "Family ties," in "Handbook of Economic Growth," Vol. 2, Elsevier, 2014, pp. 177–215.

Aman-Rana, Shan, "Does Ability Matter for Discretionary Promotions in Bureaucracies? Evidence from Pakistan," *Working Paper*, 2023.

Anderson, Michael L, "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 2008, *103* (484), 1481–1495.

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

Ashraf, Nava and Oriana Bandiera, "Social incentives in organizations," *Annual Review of Economics*, 2018, *10* (1), 439–463.

_ , _ , and Scott S Lee, "Awards unbundled: Evidence from a natural field experiment," *Journal of Economic Behavior & Organization*, 2014, *100*, 44–63.

_ , _ , Edward Davenport, Scott Lee et al., "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services," *American Economic Review*, May 2020, *110* (5), 1355–94.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti, "Active and passive waste in government spending: evidence from a policy experiment," *American Economic Review*, September 2009, *99* (4), 1278–1308.

_ , **Iwan Barankay, and Imran Rasul**, "Social preferences and the response to incentives: Evidence from personnel data," *The Quarterly Journal of Economics*, August 2005, pp. 917–962.

_ , _ , **and** _ , "Social incentives in the workplace," *The Review of Economic Studies*, April 2010, *77* (2), 417–458.

_ , **Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat**, "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats," Technical Report, National Bureau of Economic Research 2020.

Banerjee, Abhijit, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh, "Can institutions be reformed from within? Evidence from a randomized experiment with the Rajasthan police," Technical Report, National Bureau of Economic Research 2012.

Banerjee, Abhijit V, Esther Duflo, and Rachel Glennerster, "Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system," *Journal of the European economic association*, May 2008, *6* (2-3), 487–500.

Banfield, Edward C., *The Moral Basis of a Backward Society*, New York: Free Press, 1958.

Basu, Kaushik, "One kind of power," *Oxford Economic Papers*, July 1986, *38* (2), 259–282.

Becker, Gary S and George J Stigler, "Law enforcement, malfeasance, and compensation of enforcers," *The Journal of Legal Studies*, 1974, *3* (1), 1–18.

Bedasso, Biniam, "Institutional and personnel capabilities for fiscal decentralisation in Benin," Technical Report, Collaborative Africa Budget Reform Initiative (CABRI) 2021.

Bedasso, Binyam, "Public Investment Management in Benin," IMF Working Paper, International Monetary Fund 2021.

Behanzin, Précieux Christian, Manuela Hounvi, Casilda Dossou-Yovo, Steven Mariano, and Stella Towanou, "Social capital and the practices of governance in the municipality of Sèmè-Podji in Benin," *Environmental and Water Sciences, Public Health and Territorial Intelligence Journal*, 2018, *2* (4), 31–40.

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu, "The Glittering Prizes: Career Incentives and Bureaucrat Performance," *The Review of Economic Studies*, May 2019, *87* (2), 626–655.

Besley, Timothy and John McLaren, "Taxes and bribery: the role of wage incentives," *The economic journal*, 1993, *103* (416), 119–141.

_ **and Torsten Persson**, "The origins of state capacity: Property rights, taxation, and politics," *American Economic Review*, September 2009, *99* (4), 1218–44.

_ **and** _ , "State capacity, conflict, and development," *Econometrica*, 2010, *78* (1), 1–34.

Bicchieri, Cristina, *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge, UK: Cambridge University Press, 2006.

_ , *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*, Oxford, UK: Oxford University Press, 2016.

Bidwell, Kelly, Katherine Casey, and Rachel Glennerster, "Debates: Voting and expenditure responses to political communication," *Journal of Political Economy*, 2020, *128* (8), 2880–2924.

Bloch, Francis, Garance Genicot, and Debraj Ray, "Reciprocity in groups and the limits to social capital," *American Economic Review*, May 2007, *97* (2), 65–69.

Bloom, Nicholas and John Van Reenen, "Measuring and explaining management practices across firms and countries," *The Quarterly Journal of Economics*, November 2007, *122* (4), 1351–1408.

_ **, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts**, "Does management matter? Evidence from India," *The Quarterly Journal of Economics*, November 2013, *128* (1), 1–51.

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

Brown, Sarah, Daniel Gray, Jolian McHardy, and Karl Taylor, "Employee trust and workplace performance," *Journal of Economic Behavior & Organization*, 2015, *116*, 361–378.

Burgess, Robin, Matthew Hansen, Benjamin A Olken, Peter Potapov, and Stefanie Sieber, "The political economy of deforestation in the tropics," *The Quarterly journal of economics*, 2012, *127* (4), 1707–1754.

Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan, and Arman Rezaee, "The political economy of public sector absence," *Journal of Public Economics*, 2023, *218*, 104787.

Canen, Nathan, Rafael Ch, and Leonard Wantchekon, "Political uncertainty and the forms of state capture," *Journal of Development Economics*, 2023, *160*, 102972.

Ch, Rafael, Mathias Hounkpe, and Leonard Wantchekon, "campaign Finance and State Capture," Technical Report, Economic Development and Institutions 2019.

Collier, Paul and Pedro C Vicente, "Votes and violence: Evidence from a field experiment in Nigeria," *The economic journal*, 2014, *124* (574), F327–F355.

Corbacho, Ana, Daniel W Gingerich, Virginia Oliveros, and Mauricio Ruiz-Vega, "Corruption as a self-fulfilling prophecy: evidence from a survey experiment in Costa Rica," *American Journal of Political Science*, 2016, *60* (4), 1077–1092.

Dasgupta, Aditya and Devesh Kapur, "The political economy of bureaucratic overload: Evidence from rural development officials in India," *American Political Science Review*, 2020, *114* (4), 1316–1334.

Dasgupta, Partha, "Economics of social capital," *Economic Record*, 2005, *81*, S2–S21.

de Janvry, Alain, Guojun He, Elisabeth Sadoulet, Shaoda Wang, and Qiong Zhang, "Subjective Performance Evaluation, Influence Activities, and Bureaucratic Work Behavior: Evidence from China," *American Economic Review*, March 2023, *113* (3), 766–799.

Debnath, Sisir, Mrithyunjayan Nilayamgode, and Sheetal Sekhri, "Information Bypass: Using Low-cost technological innovations to curb leakages in welfare programs," *Journal of Development Economics*, 2023, p. 103137.
Decentralisation and Communal Development Support Programme

***Decentralisation and Communal Development Support Programme**, Technical Report, KfW Development Bank 2021.*

Dhaliwal, Iqbal and Rema Hanna, *"The devil is in the details: The successes and limitations of bureaucratic reform in India," Journal of Development Economics, 2017, 124, 1–21.*

Dipoppa, Gemma and Saad Gulzar, *"Administrative Incentives Impact Crop-Residue Burning and Health in South Asia," Working Paper, 2023.*

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, *"Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India," The Quarterly Journal of Economics, 2013, 128 (4), 1499–1545.*

_ , _ , _ , and _ , *"The value of regulatory discretion: Estimates from environmental inspections in India," Econometrica, 2018, 86 (6), 2123–2160.*

Evans, Peter and James E Rauch, *"Bureaucracy and growth: A cross-national analysis of the effects of" Weberian" state structures on economic growth," American sociological review, 1999, pp. 748–765.*

Finan, Frederico, Benjamin A Olken, and Rohini Pande, *"The personnel economics of the developing state," in "Handbook of Economic Field Experiments," Vol. 2, Elsevier, 2017, pp. 467–514.*

Fisher, Ronald A., *The Design of Experiments, Edinburgh: Oliver and Boyd, 1935.*

Fujiwara, Thomas and Leonard Wantchekon, *"Can informed public deliberation overcome clientelism? Experimental evidence from Benin," American Economic Journal: Applied Economics, October 2013, 5 (4), 241–255.*

Gambetta, Diego, *"The Mafia: A ruinous rationality," 1993.*

Goold, Susan Dorr, Michael A Neblo, SY Kim, R De Vries, Gene Rowe, and Peter Muhlberger, *"What is good public deliberation," Hastings Center Report, 2012, 42 (2), 24–26.*

Gulzar, Saad and Benjamin J Pasquale, *"Politicians, bureaucrats, and development: Evidence from India," American Political Science Review, 2017, 111 (1), 162–183.*

Hadfield, Gillian K and Stephen Macedo, *"Rational reasonableness: Toward a positive theory of public reason," The Law & Ethics of Human Rights, 2012, 6 (1), 7–46.*

Imbert, Bruno, Mouhamadou Sy, Bacari Koné, Louis d'Humières, Pierre Roumegas, Nicoletta Feruglio, Mathias Gogohounga, and Angelo Donou,

"Benin Technical Assistance Report - Public Investment Management Assessment," Technical Report, International Monetary Fund 2018.

Iyer, Lakshmi and Anandi Mani, *"Traveling agents: political change and bureaucratic turnover in India," Review of Economics and Statistics, August 2012, 94 (3), 723–739.*

Jia, Ruixue, Masayuki Kudamatsu, and David Seim, *"Political selection in China: The complementary roles of connections and performance," Journal of the European Economic Association, August 2015, 13 (4), 631–668.*

Kandel, Eugene and Edward P Lazear, *"Peer pressure and partnerships," Journal of Political Economy, 1992, 100 (4), 801–817.*

Keefer, Philip and Razvan Vlaicu, *"Employee trust and performance constraints in public sector organizations," European Journal of Political Economy, 2024, p. 102503.*

Khan, Adnan Q, Asim I Khwaja, and Benjamin A Olken, *"Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," The Quarterly Journal of Economics, October 2016, 131 (1), 219–271.*

_ , Asim Ijaz Khwaja, and Benjamin A Olken, *"Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings," Technical Report, National Bureau of Economic Research 2018.*

Khan, Muhammad Yasir, *"Mission motivation and public sector performance: Experimental evidence from pakistan," Unpublished manuscript, 2023.*

Kohli, Atul, *State-directed development: political power and industrialization in the global periphery, Cambridge University Press, 2004.*

Lee, David S, *"Training, wages, and sample selection: Estimating sharp bounds on treat," 2009.*

Levitt, Steven D. and John A. List, *"What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?," Journal of Economic Perspectives, 2007, 21 (2), 153–174.*

Lichand, Guilherme, Marcos FM Lopes, and Marcelo C Medeiros, *"Is corruption good for your health," scholar. harvard. edu/glichand/publications/job-market-paper. Acessado em, 2016, 25 (04), 2016.*

López-Moctezuma, Gabriel, Leonard Wantchekon, Daniel Rubenson, Thomas Fujiwara, and Cecilia Pe Lero, *"Policy deliberation and voter persuasion: Experimental evidence from an election in the Philippines," American Journal of Political Science, 2022, 66 (1), 59–74.*

MacKinnon, James G and Matthew D Webb, *"Randomization inference for difference-in-differences with few treated clusters," Journal of Econometrics, 2020, 218 (2), 435–450.*

Mas, Alexandre and Enrico Moretti, *"Peers at work," The American Economic Review, March 2009, 99 (1), 112–145.*

Mendelberg, Tali, *"The deliberative citizen: Theory and evidence," Political decision making, deliberation and participation, 2002, 6 (1), 151–193.*

Milgrom, Paul and John Roberts, *"An economic approach to influence activities in organizations," American Journal of Sociology, 1988, 94, S154–S179.*

Muralidharan, Karthik and Paul Niehaus, *"Experimentation at scale," Journal of Economic Perspectives, November 2017, 31 (4), 103–124.*

Niehaus, Paul and Sandip Sukhtankar, *"The marginal rate of corruption in public programs: Evidence from India," Journal of Public Economics, 2013, 104, 52–64.*

Olken, Benjamin A, *"Monitoring corruption: evidence from a field experiment in Indonesia," Journal of political Economy, 2007, 115 (2), 200–249.*

Parkinson, John and Jane Mansbridge, *Deliberative systems: Deliberative democracy at the large scale, Cambridge University Press, 2012.*

Peter, B. Evans, *Embedded autonomy: states and industrial transformation, Vol. 25, Princeton, NJ: Princeton University Press, 1995.*

Porta, Rafael La, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny, *"The quality of government," Journal of Law, Economics, and Organization, March 1999, 15 (1), 222–279.*

Portes, Alejandro and Patricia Landolt, *"The downside of social capital," 1996.*

Rasul, Imran and Daniel Rogger, *"Management of bureaucrats and public service delivery: Evidence from the nigerian civil service," The Economic Journal, 2018, 128 (608), 413–446.*

Riaño, Juan Felipe, *"Bureaucratic nepotism," Available at SSRN: https://ssrn.com/abstract=3995589 or http://dx.doi.org/10.2139/ssrn.3995589, November 2021.*

Riedl, Rachel Beatty and J Tyler Dickovick, *"Party systems and decentralization in Africa," Studies in Comparative International Development, September 2014, 49, 321–342.*

Spradley, James P., *Participant Observation, New York: Holt, Rinehart and Winston, 1980.*

Tella, Rafael Di and Ernesto Schargrodsky, *"The role of wages and auditing during a crackdown on corruption in the city of Buenos Aires," Journal of Law and Economics, 2003, 46 (1), 269–292.*

Tirole, Jean, *"Hierarchies and bureaucracies: On the role of collusion in organizations," The Journal of Law, Economics, & Organization, October 1986, 2 (2), 181–214.*

Wantchekon, Leonard and Jenny Guardado, *"Deliberation and Ethical Voting Behavior: Evidence from a Campaign Experiment in Benin," American Political Science Review, Forthcoming.*

Tables

Table 1: Description of deliberation meetings

	Meeting 1		Meeting 2		Meeting 3		Overall	
	Mean	N	Mean	N	Mean	N	Mean	N
Meetings' Characteristics								
Total Participants	31.30	10	31.60	10	34.10	10	32.33	30
Bureaucrat Participants	21.80	10	23.80	10	22.90	10	22.83	30
General Secretary Attended	0.90	10	1.00	10	1.00	10	0.97	30
Duration (in minutes)	100.00	10	91.80	10	94.10	10	95.30	30
Duration of Presentation (in minutes)	32.70	10	36.40	10	38.20	10	35.77	30
Duration of Discussion (in minutes)	48.90	10	37.50	10	44.20	10	43.53	30
Duration of Presentation (in percentage)	0.34	10	0.42	10	0.42	10	0.39	30
Duration of Discussion (in percentage)	0.47	10	0.41	10	0.48	10	0.46	30
Number of Contributions	7.75	8	4.30	10	3.40	10	4.96	28
Number of Critics	7.20	5	5.60	10	9.67	6	7.14	21
Number of Questions	4.00	8	2.89	9	4.83	6	3.78	23
Perception of Enumerators								
Meeting is Interactive	1.00	10	1.00	9	1.00	10	1.00	29
Meeting is Informative	1.00	10	1.00	9	0.60	10	0.86	29
Meeting Atmosphere is Good	1.00	10	1.00	9	1.00	10	1.00	29
Meeting Atmosphere is Tense	0.00	10	0.00	9	0.00	10	0.00	29

Note: The data is compiled from the meeting notes taken by enumerators. Meeting 1 discussed the audit reports of the municipalities, while meetings 2 and 3 described results from a survey of bureaucrats and citizens, respectively. General Secretary attended is a dummy variable that turns on 1 of a General Secretary of the municipality attended the meeting. Criticism, Contribution, and Question are the number of times these items were recorded during each meeting across municipalities. The enumerators also reported their perceptions regarding the atmosphere of the meetings. These are dummies for whether the meetings were interactive or informative and whether the atmosphere was tense or good.

Table 2: Did Deliberation Affect Municipal Performance?

Dependent Variables	Pre-Period Control Mean	DID Estimate	Standard Error	P-Value	Randomization Inference P-Value	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Municipal Audit-Based Performance	77.835	6.8286**	2.9555	0.032	0.000	112

Note: The unit of observation is municipality-year. The data is from the Audit Reports of 2014-2019. Municipal Audit-Based Performance is a score constructed from 3 sub-components: (i) Record Keeping (ii) Transparency of Public Procurement (iii) Overall Management. Record Keeping is constructed from scores on: maintenance of accounting records, existence of documentation room/archive accessible to the public and equipped with storage furniture, statements of transferred resources, updated physical registers and stock-taking by the authorized officers. Transparency of Public Procurement is constructed from scores given on: Secretariat of the organizations responsible for public procurement is functional, the number and exhaustiveness of the reports published by them, details on FaDeC funds and spending, compliance with public procurement procedures, quality of accounts and traceability of transfers. Overall Management is constructed from scores given on effectiveness of the role of coordinator of local services by the Secretary General, and management of certificates of civil status (certificates of birth, death, marriage and divorce). We report *p*-value in Column (4) and *p*-value from Randomization Inference with 1,000 repetitions in Column (5). All regressions include municipality and year fixed effects. Standard Errors are clustered at the municipality level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table 3: Did Deliberation Affect Bureaucrats' Beliefs About Corruption and Enforcement of Rules?

Dependent Variables	ITT	Standard Error	P-Value	RI P-Value	Obs.
	(1)	(2)	(3)	(4)	(5)
Aggregate Index					
Beliefs about Rule of Law	-0.4045***	0.1299	0.006	0.006	580
Subcomponent: Bureaucrats' beliefs about corruption among					
Councilors	0.2031	0.1333	0.144	0.142	455
Other Bureaucrats	0.2172	0.1434	0.146	0.142	510
City Hall Vendors	0.4714***	0.1464	0.005	0.004	374
Subcomponent: Bureaucrats' beliefs about enforcement of rules for					
Acquisition of Office Equipment	-0.2658	0.1584	0.110	0.109	431
Public Procurement	-0.2707*	0.1476	0.082	0.071	401
Hiring Staff	-0.2720	0.1893	0.167	0.154	411

Note: The unit of observation is a bureaucrat. The Belief about Rule of Law Index is a standardized aggregate of beliefs about corruption and enforcement of rules as described in the lower panels. Trust is measured on a four-point scale ranging from "not at all" to "very much," with higher values indicating greater trust. We report *p*-value in Column (3) and *p*-value from Randomization Inference with 1,000 repetitions in Column (4). Standard errors are clustered at the municipality level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table 4: Did Deliberation Affect Bureaucrats' Trust in Others?

Dependent Variables	ITT	Standard Error	P-Value	RI P-Value	Obs.
	(1)	(2)	(3)	(4)	(5)
Trust Index	-0.3511**	0.1585	0.039	0.060	595
Bureaucrats' Trust in					
Councilors	-0.3769**	0.1613	0.031	0.020	581
Office Colleagues	-0.2732	0.1589	0.102	0.140	538
Colleagues in Same Dept	-0.2788*	0.1415	0.064	0.060	574
City Hall Colleagues	-0.2170	0.1476	0.158	0.200	582

Note: The unit of observation is a bureaucrat. The Trust Index is a standardized aggregate of trust in councilors, office colleagues, colleagues in the same department, and city hall colleagues. We report *p*-value in Column (3) and *p*-value from Randomization Inference with 1,000 repetitions in Column (4). Standard errors are clustered at the municipality level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

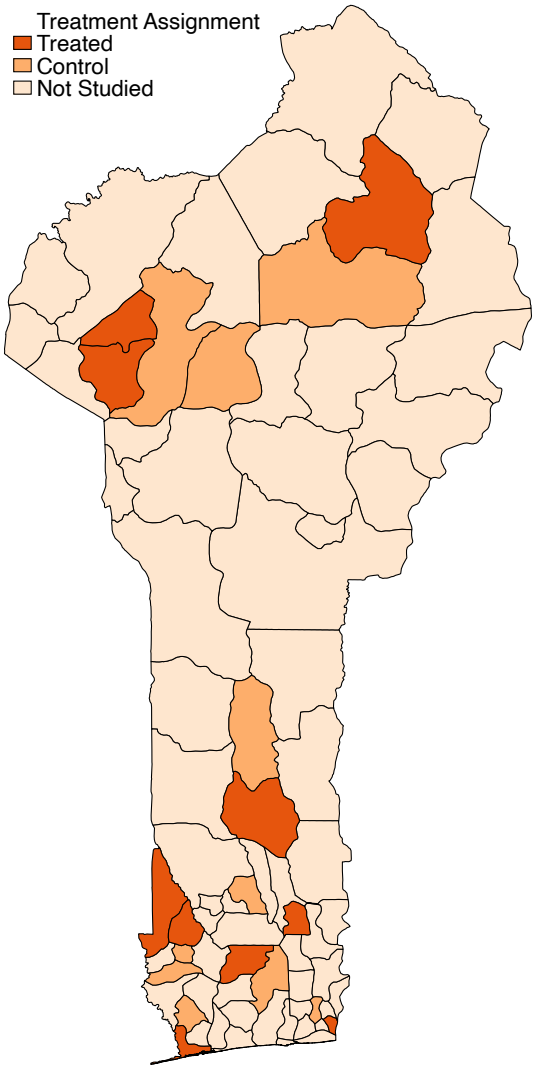
Table 5: Number of meetings held

	Number of City-hall Meetings attended in the last 3 months	
	(1)	(2)
	Overall	Excluding Deliberation
Treatment	-0.460	-1.610*
	(0.896)	(0.847)
	{0.640}	{0.0800}
Control	6.132	6.098*
	(0.625)	(0.622)
	{0.380}	{0.0600}
N	595	595

Note: The unit of analysis is a bureaucrat. Standard errors in parenthesis are clustered at the municipality level. Randomization inference *p*-value is in square brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

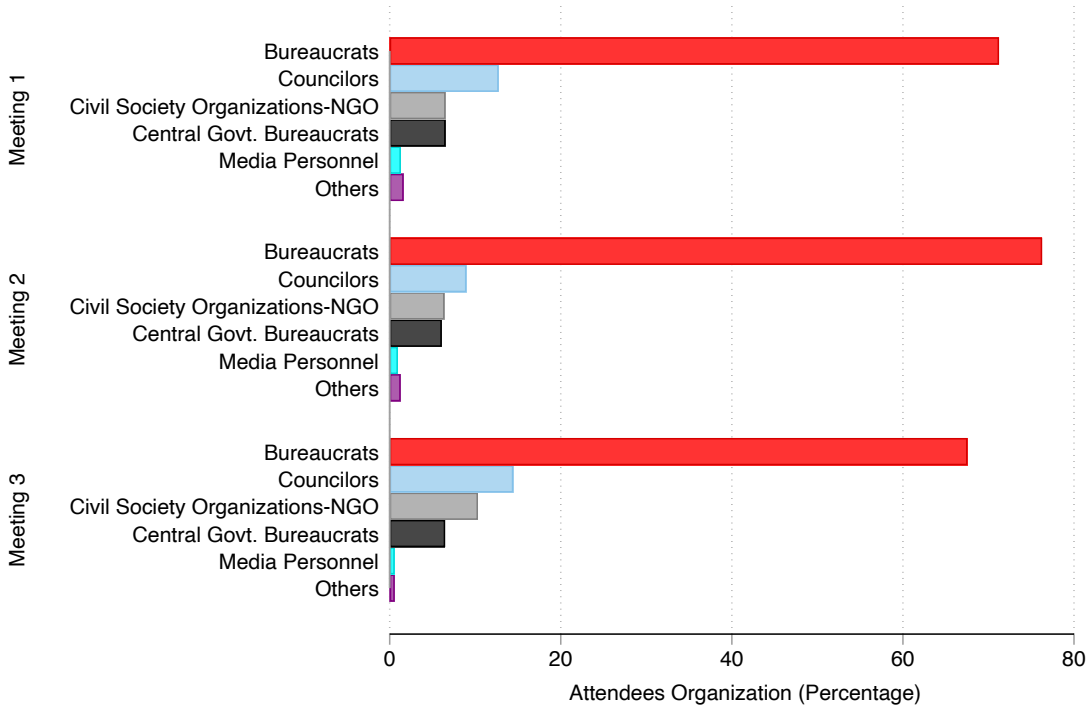
Figures

Figure 1: Spatial Distribution of Treatment Assignment



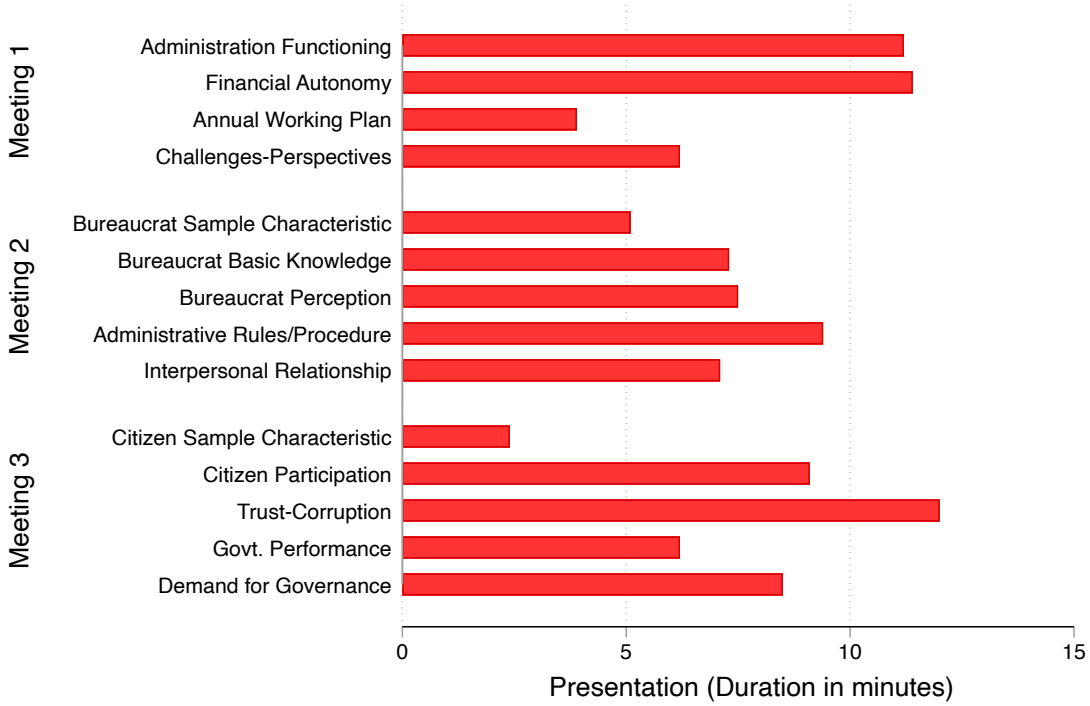
Notes: The figure shows the spatial distribution of the 20 municipalities in the sample across. Shading indicates treatment and control assignment.

Figure 2: Attendance by Types of Actors in the Collective Deliberation Meetings



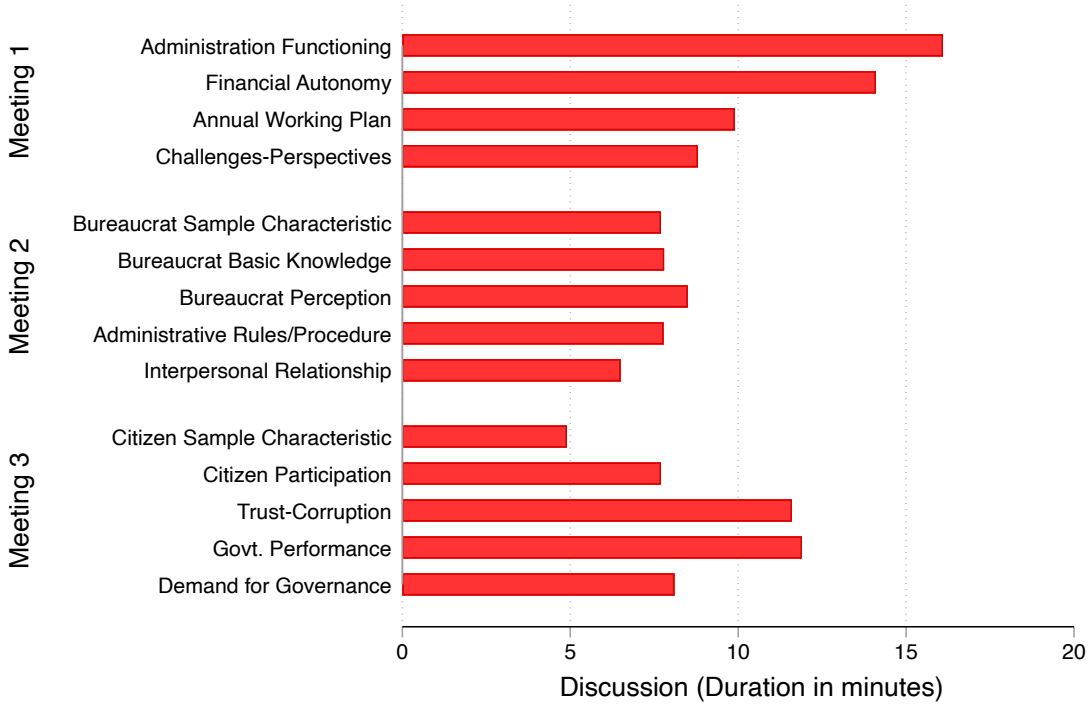
Notes: Data are collected by enumerators before the start of the meeting.

Figure 3: Presentation topics and their duration



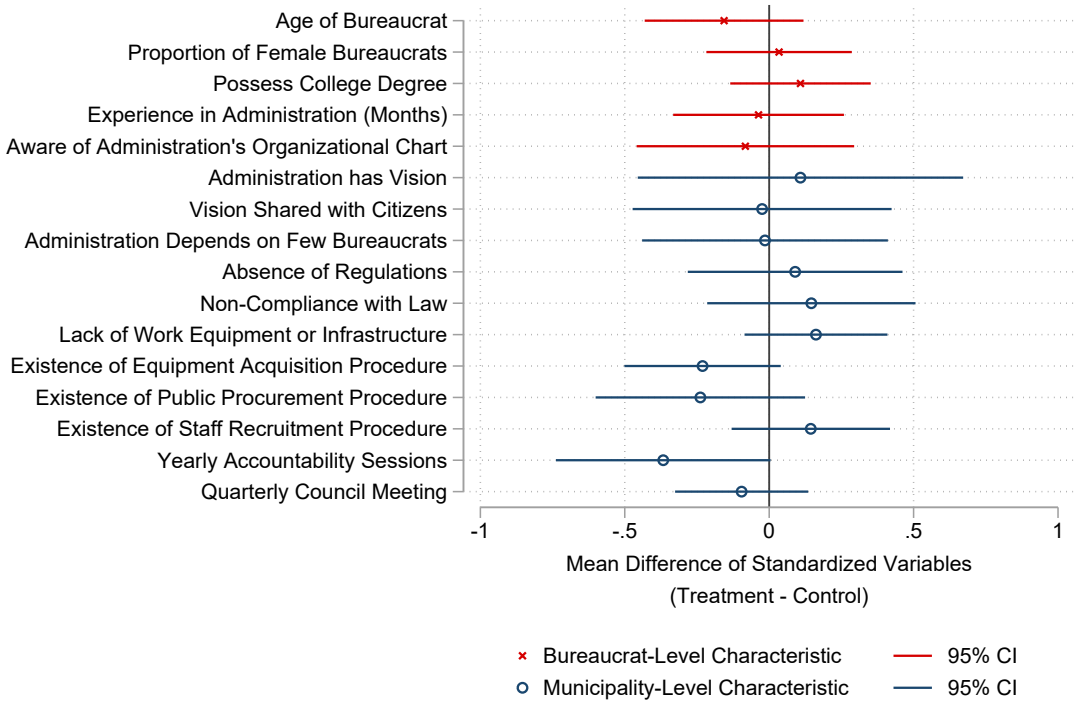
Notes: Data are from meeting notes taken by enumerators.

Figure 4: Discussion topics and their duration



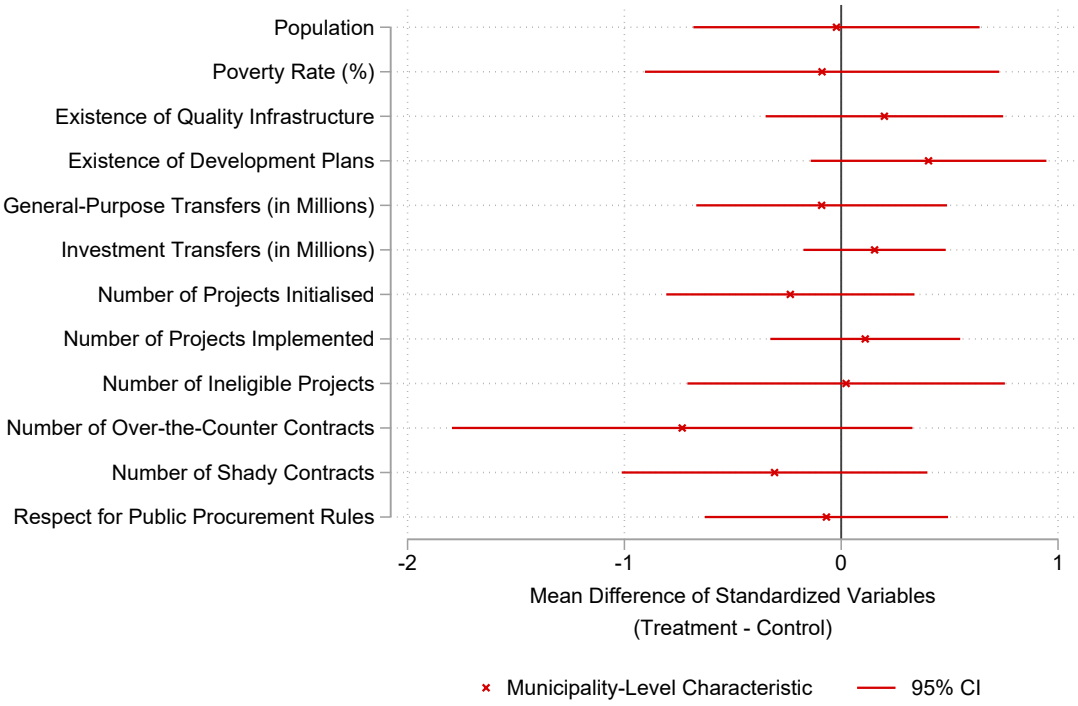
Notes: Data are from meeting notes taken by enumerators.

Figure 5: Balance test on characteristics of bureaucrats



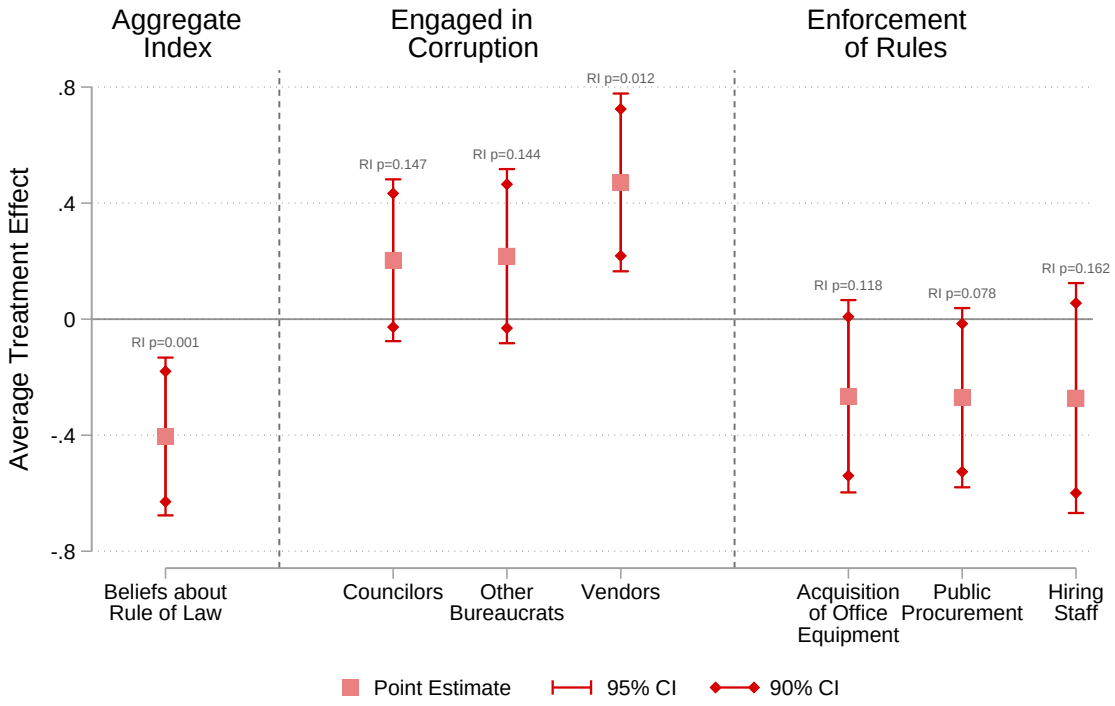
Notes. The variables have been standardized and their mean difference between treatment and control group is plotted along with 95% CIs. The standard errors are clustered at the municipality level. The source of the data is bureaucrat survey.

Figure 6: Balance test on municipal characteristics



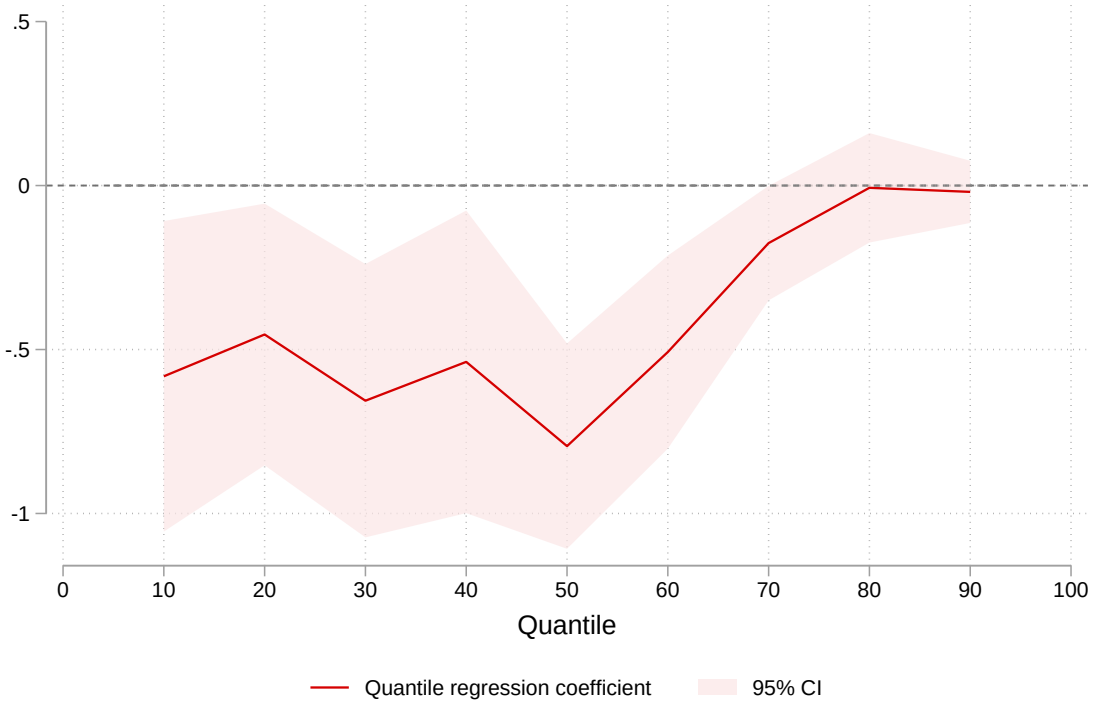
Notes. The variables have been standardized and their mean difference between treatment and control group is plotted along with 95% CIs. The standard errors are clustered at the municipality level. The source of the data are the municipal audit reports, which are available on the National Commission for Local Finance (CoNaFiL) website.

Figure 7: Did Deliberation Affect Bureaucratic Beliefs about Corruption and Enforcement of Rules



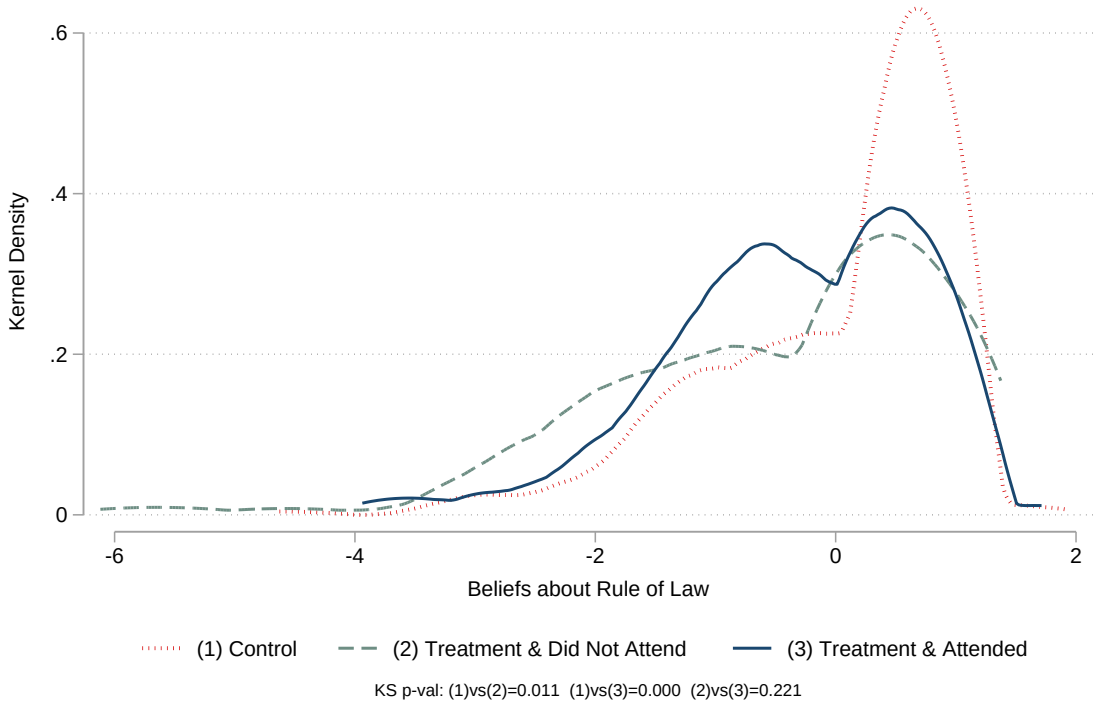
Notes: The figure reports estimates of the effect of deliberation on bureaucrats' beliefs about corruption and enforcement of rules, with 90% and 95% confidence intervals. Randomization inference p-values (1,000 repetitions) are shown above each estimate. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality level.

Figure 8: Quantile Regression Estimates: Did Deliberation Affect Bureaucratic Beliefs about Corruption and Enforcement of Rules



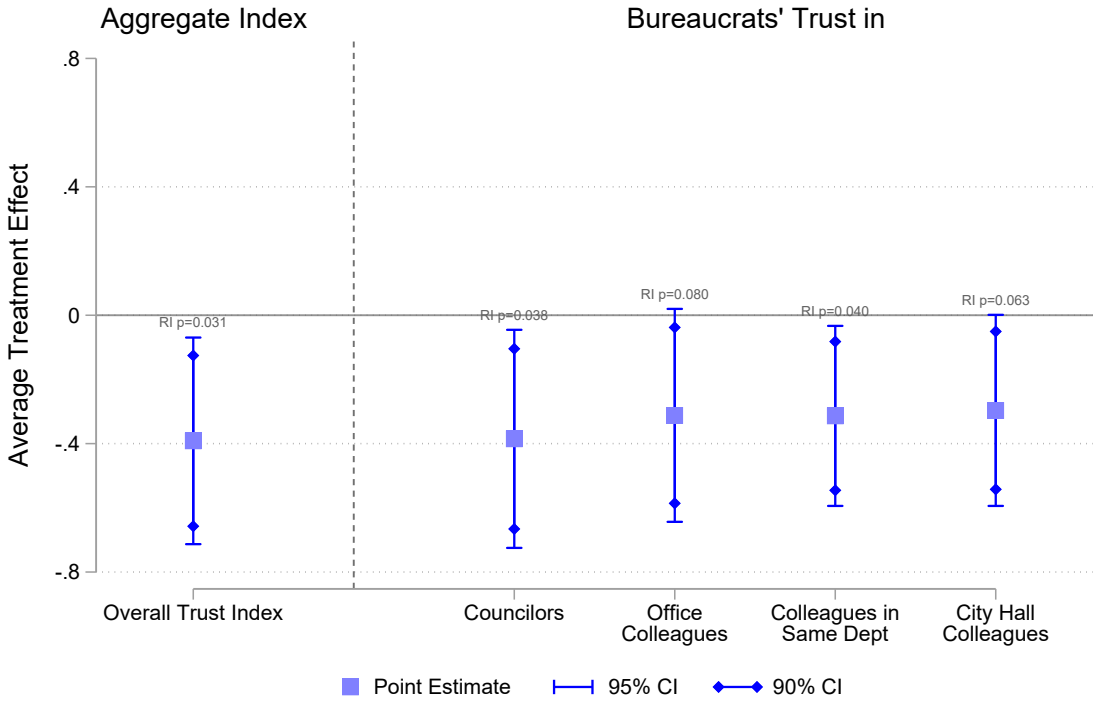
Notes: The figure reports quantile regression estimates of the treatment effect on the index of beliefs about rule of law across quantiles 10 to 90. The shaded area represents the 95% confidence interval. Standard errors are clustered at the municipality level.

Figure 9: Kernel Density of Bureaucratic Beliefs about Corruption and Enforcement



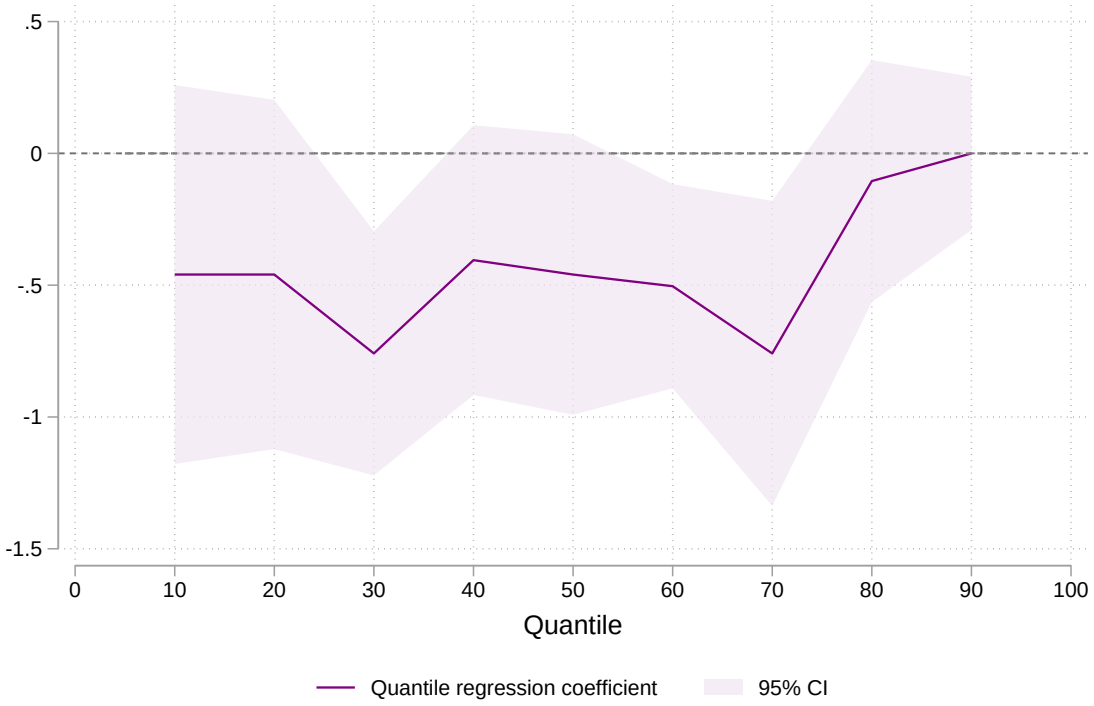
Notes: The figure shows kernel density estimates of the Beliefs about Rule of Law index by treatment and attendance status. Kolmogorov-Smirnov test p-values for pairwise comparisons are reported below the figure.

Figure 10: Did Deliberation Affect Bureaucrats' Trust in Others?



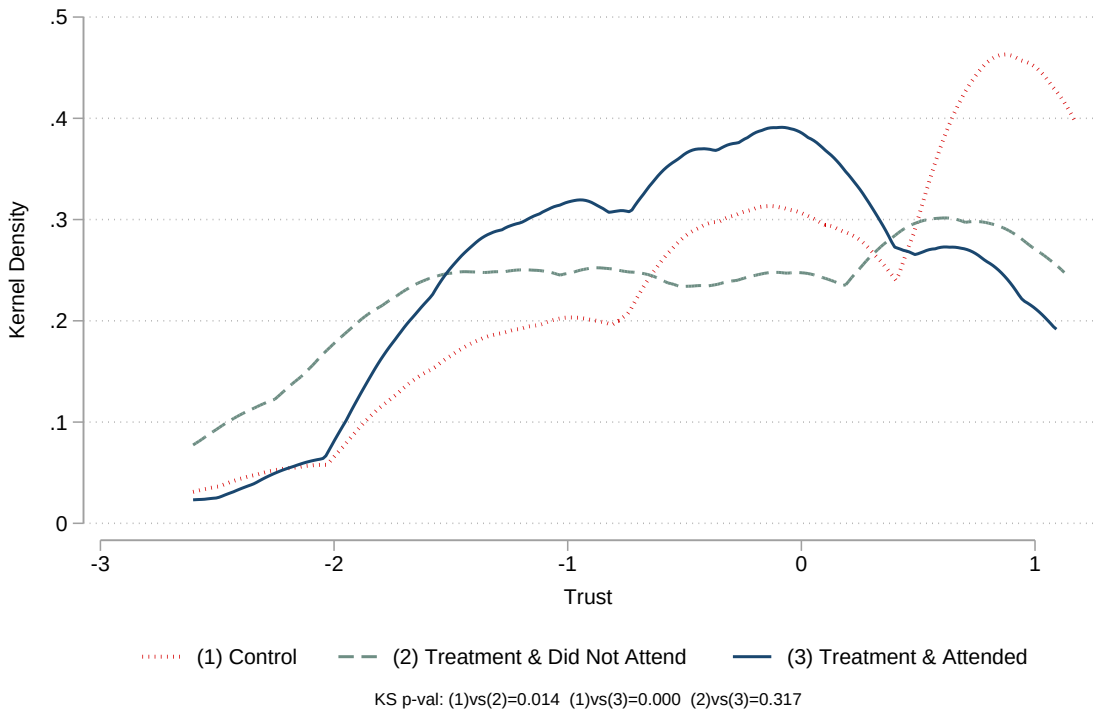
Notes: The figure reports estimates of the effect of deliberation on bureaucrats' trust, with 90% and 95% confidence intervals. Randomization inference p-values (1,000 repetitions) are shown above each estimate. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality level.

Figure 11: Quantile Regression Estimates: Did Deliberation Affect Bureaucrats' Trust in Others



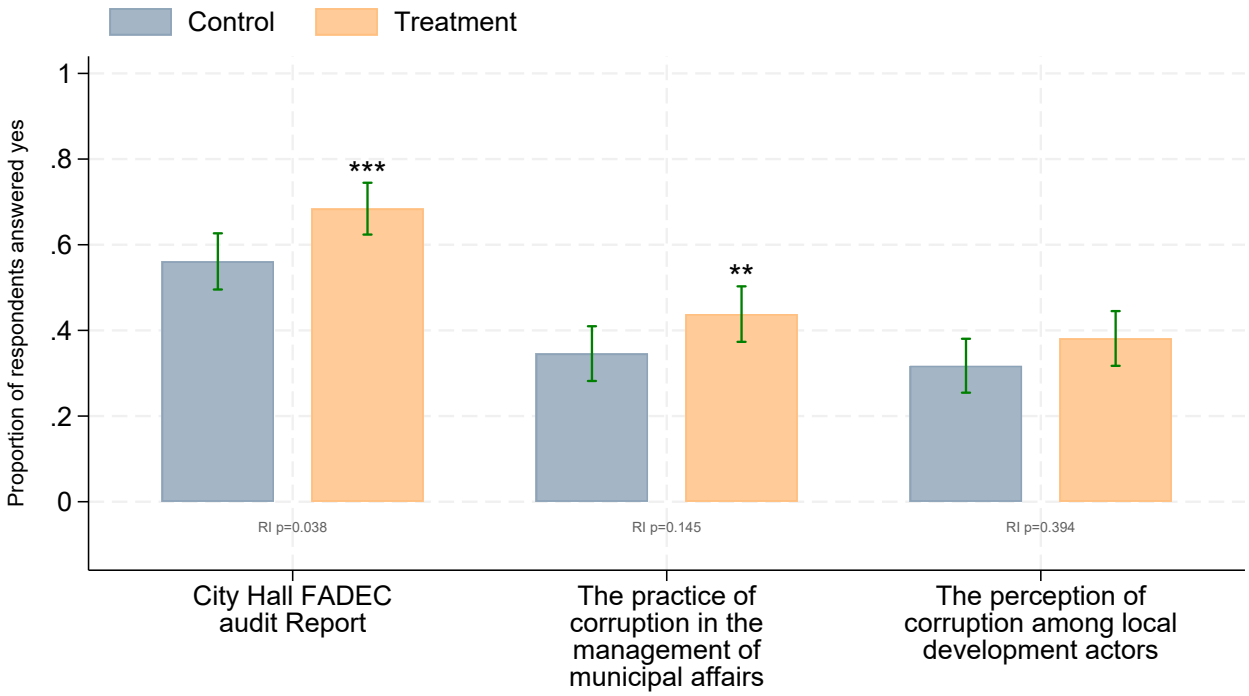
Notes: The figure reports quantile regression estimates of the treatment effect on the Trust index across quantiles 10 to 90. The shaded area represents the 95% confidence interval. Standard errors are clustered at the municipality level.

Figure 12: Kernel Density of Bureaucratic Trust Index



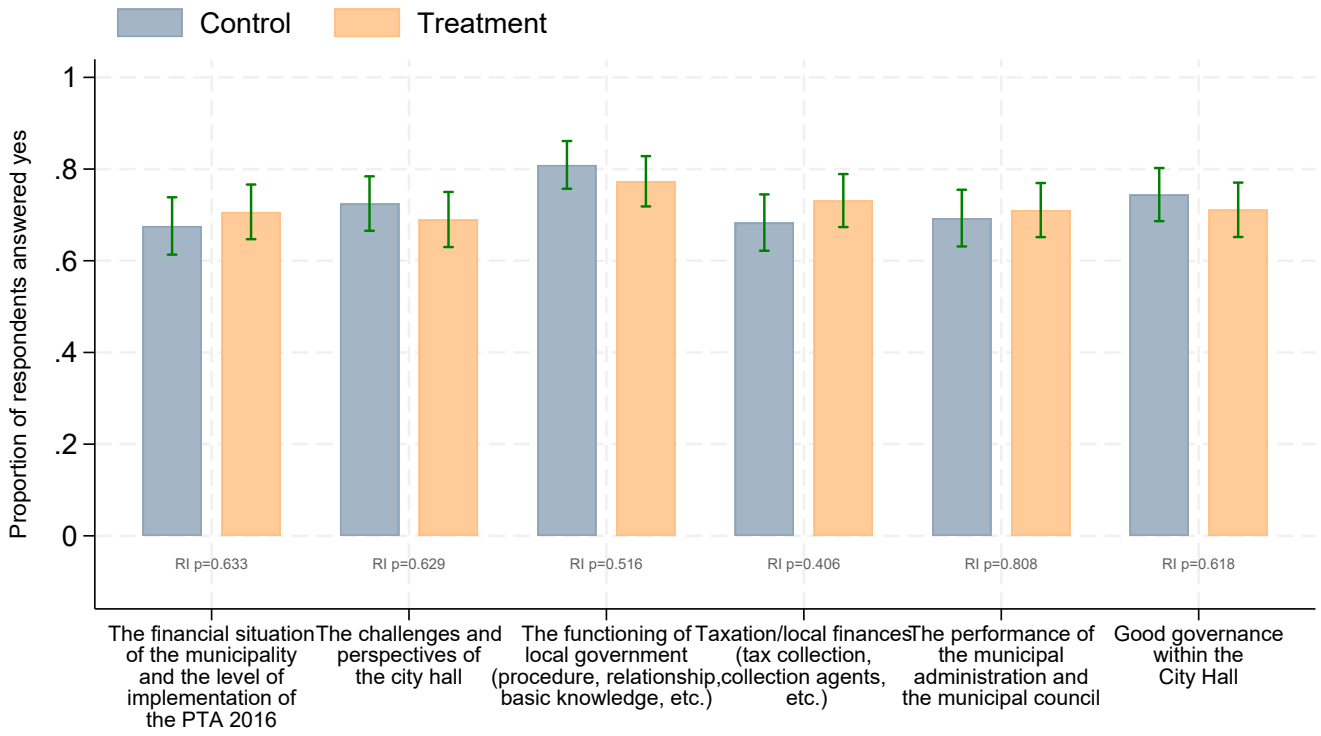
Notes: The figure shows kernel density estimates of the Trust index by treatment and attendance status. Kolmogorov-Smirnov test p-values for pairwise comparisons are reported below the figure.

Figure 13: Topics Discussed During General Meetings in the Municipality



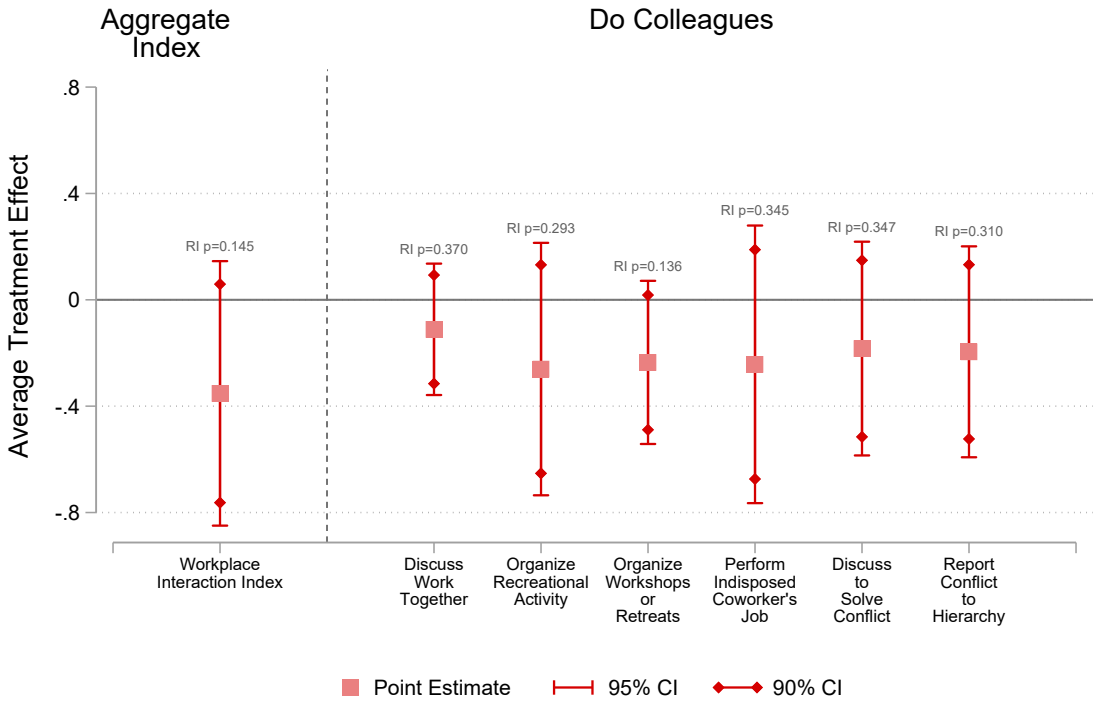
Notes: The figure shows the proportion of respondents who reported that audit and corruption-related topics were discussed during general meetings in the municipality, by treatment status. Error bars represent 95% confidence intervals. Stars denote statistical significance of the difference in means: * p<0.10, ** p<0.05, *** p<0.01. Randomization inference p-values (1,000 repetitions) are reported below.

Figure 14: Topics Discussed During General Meetings in the Municipality



Notes: The figure shows the proportion of respondents who reported that various non-audit topics were discussed during the meetings, by treatment status. Error bars represent 95% confidence intervals. Stars denote statistical significance of the difference in means: * p<0.10, ** p<0.05, *** p<0.01. Randomization inference p-values (1,000 repetitions) are reported below.

Figure 15: Deliberation and Workplace Learning and Coordination: Effects on Bureaucrats' Workplace Interactions



Notes: The figure reports estimates of the effect of deliberation on workplace interaction outcomes, with 90% and 95% confidence intervals. Randomization inference p-values (1,000 repetitions) are shown above each estimate. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality level.

Appendices

A Appendix Tables

Table A.1: Relation between previous year's audit-based performance and current year's FaDeC investment transfers

	Dependent Variable: FaDeC Investment Transfers (t)			
	(1)	(2)	(3)	(4)
Audit-Based Performance (t-1)	1615995**	1564939*	1546852*	2029669**
	(735236)	(754951.5)	(763319.3)	(905340.7)
	[0.0406]	[0.0520]	[0.0570]	[0.0371]
Control for Population Size	No	Yes	Yes	Yes
Control for Poverty Rate	No	No	Yes	Yes
Municipality Fixed Effects	No	No	No	Yes
Observations	97	97	97	97

Note: The unit of observation is municipality-year. The data is from the Audit Reports of 2014-2019. The dependent variable is the amount of FaDeC investment transfers in year t. Audit-based performance (t-1) is last year's performance score of the municipality as evaluated by the auditors. These are measured in percentages. Column (1) has no controls. Column (2) controls for population size of the municipalities in year t, while Column (3) controls for their poverty rates. Finally, Column (4) includes municipality fixed effects. The standard errors are reported in parentheses and are clustered at the municipality level. The p-values are reported in square brackets. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.2: Baseline correlation between municipal audit-based performance and citizen experience index (2014)

	Municipal Audit-Based Performance		
	Correlation Coefficient	P-Value	Obs.
Citizen Experience Index	0.6232***	0.010	16

Note: Citizen Experience Index is a standardized index of the several indicators at the citizen level. These indicators are: whether the citizens feel that (i) there are corrupt government officials (ii) these officials go unpunished (iii) their present living conditions are good (iv) their living conditions are better compared to others (v) it is difficult to access identity document and household services (vi) bribes are required to access them. In the construction of the index, only the signs of all indicators except (iii) and (iv) were flipped so that higher values are better.

Table A.3: Sub-components of Performance: Did Collective Deliberation Affect Municipal Audit-based Performance?

Dependent Variables	Pre-Period Control Mean	DID Estimate	Standard Error	P-Value	Randomization Inference P-Value	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Record Keeping	78.493	9.3016	7.7031	0.242	0.220	112
Transparency of Public Procurement	78.297	5.7866	4.6223	0.226	0.240	112
Overall Management	83.500	14.4186	13.4617	0.298	0.200	72

Note: The unit of observation is municipality-year. The data is from the Audit Reports of 2014-2019. Municipal Audit-Based Performance is a score constructed from 3 sub-components: (i) Record Keeping (ii) Transparency of Public Procurement (iii) Overall Management. Record Keeping is constructed from scores on: maintenance of accounting records, existence of documentation room/archive accessible to the public and equipped with storage furniture, statements of transferred resources, updated physical registers and stock-taking by the authorized officers. Transparency of Public Procurement is constructed from scores given on: Secretariat of the organizations responsible for public procurement is functional, the number and exhaustiveness of the reports published by them, details on FaDeC funds and spending, compliance with public procurement procedures, quality of accounts and traceability of transfers. Overall Management is constructed from scores given on effectiveness of the role of coordinator of local services by the Secretary General, and management of certificates of civil status (certificates of birth, death, marriage and divorce). We report *p*-value in Column (4) and *p*-value from Randomization Inference with 1,000 repetitions in Column (5). All regressions include municipality and year fixed effects. Standard Errors are clustered at the municipality level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.4: Attrition: Bureaucrats' Beliefs About Corruption and Enforcement of Rules

1(Missing)	Treatment Effect (1)	Standard Error (2)	P-Value (3)	Non-Missing Obs. (4)
Aggregate Index				
Beliefs about Rule of Law	-0.0103	0.0136	0.455	580
Subcomponent: Bureaucrats' beliefs about corruption among				
Councilors	-0.0158	0.0515	0.762	455
Other Bureaucrats	-0.0720*	0.0412	0.096	510
City Hall Vendors	-0.0340	0.0752	0.656	374
Subcomponent: Bureaucrats' beliefs about enforcement of rules for				
Acquisition of Office Equipment	0.0241	0.0798	0.766	431
Public Procurement	0.0102	0.0814	0.902	401
Hiring Staff	0.0171	0.0807	0.835	411

Note: The table reports treatment effects on an indicator for whether the dependent variable is missing. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.5: Lee Bounds: Bureaucrats' Beliefs About Corruption and Enforcement of Rules

Dependent Variables	Treatment Effect	Lee Bounds		Obs.
		Lower Bound	Upper Bound	
	(1)	(2)	(3)	(4)
Aggregate Index				
Beliefs about Rule of Law	-0.3969***	-0.4252***	-0.3549***	580
	(0.1048)	(0.0924)	(0.1011)	
Subcomponent: Bureaucrats' belief about corruption among				
Councilors	0.2227	0.2031	0.2191**	455
	(0.1539)	(0.1812)	(0.1071)	
Other Bureaucrats	0.1800	0.0588	0.2849***	510
	(0.1916)	(0.1222)	(0.1017)	
City Hall Vendors	0.4372***	0.3421*	0.5336***	374
	(0.1263)	(0.1942)	(0.1387)	
Subcomponent: Bureaucrats' beliefs about enforcement of rules for				
Acquisition of Office Equipment	-0.2209	-0.3390***	-0.2428**	431
	(0.1508)	(0.1160)	(0.1066)	
Public Procurement	-0.3381**	-0.3388	-0.2614**	401
	(0.1321)	(0.2205)	(0.1097)	
Hiring Staff	-0.2284	-0.3507***	-0.2558**	411
	(0.1683)	(0.1254)	(0.1124)	

Note: Column (1) reports the regression results from main results and the standard errors are clustered at the municipality level. Column (2) and (3) report the estimates of lower and upper Lee Bounds respectively (Lee, 2009). No controls are used and the standard errors are not clustered. The unit of observation is a bureaucrat. Significance levels are coded as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.6: Attrition - Trust

1(Missing)	Treatment Effect (1)	Standard Error (2)	P-Value (3)	Non-Missing Obs. (4)
Aggregate Index				
Trust Index	-	-	-	595
Subcomponent: Bureaucrats' trust in				
Councilors	-0.0070	0.0129	0.596	581
Office Colleagues	0.0091	0.0474	0.850	538
Same Dept Colleagues	-0.0037	0.0205	0.858	574
City Hall Colleagues	-0.0036	0.0131	0.788	582

Note: The table reports treatment effects on an indicator for whether the dependent variable is missing. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.7: Lee Bounds - Trust

Dependent Variables	Treatment Effect (1)	Lee Bounds Lower Bound (2)	Lee Bounds Upper Bound (3)	Obs. (4)
Aggregate Index				
Trust Index	-0.2363	-	-	595
	(0.1864)			
Subcomponent: Bureaucrats' trust in				
Councilors	-0.4028**	-0.3852***	-0.3641***	581
	(0.1875)	(0.0869)	(0.0886)	
Office Colleagues	-0.0340	-0.2732**	-0.2649***	538
	(0.1813)	(0.1132)	(0.0937)	
Same Dept Colleagues	-0.0729	-0.2831***	-0.2709***	574
	(0.1505)	(0.0856)	(0.0897)	
City Hall Colleagues	-0.0532	-0.2219***	-0.2170**	582
	(0.1306)	(0.0858)	(0.0871)	

Note: Column (1) reports the regression results from main results and the standard errors are clustered at the municipality level. Column (2) and (3) report the estimates of lower and upper Lee Bounds respectively (Lee, 2009). No controls are used and the standard errors are not clustered. The unit of observation is a bureaucrat. Significance levels are coded as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.8: Auditor's Evaluation of Local Government Council

Dependent Variables	Pre-Period Control Mean	DID Estimate	Standard Error	P-Value	Randomization Inference P-Value	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Overall Performance	67.818	3.2039	2.5163	0.218	0.160	120
Council's Performance	55.774	-2.0356	5.3483	0.708	0.560	112
Sub-Components of Council's Performance						
Local Government Finance	44.231	-3.0958	6.0608	0.615	0.400	112
Rule Following by the Council	96.167	-0.0217	7.2104	0.998	1.000	112

Notes: The unit of observation is municipality-year. Overall Performance is a score assigned by CoNaFiL to a municipality annually based on several criteria including both Municipal and Council's Performance. Council's Performance is a score constructed from 2 sub-components : (i) Local Government Finance and (ii) Rule Following by the Council. Local Government Finance is in turn constructed from scores given based on increase in own revenue over the year, self-financing of the investment, self-financing capacity, expenditure on maintenance of movable and immovable assets compared to operating expenditure, infrastructure maintenance expenditure, evolution of own revenue, and level of consumption of available resources. Rule Following by the Council is in turn constructed from scores given based on adoption of the initial budget within the legal deadlines, budget vote date, date of voting of the administrative account, administrative account voting date, and availability of the administrative account within the legal deadlines. We report *p*-value in Column (4) and *p*-value from Randomization Inference with 1,000 repetitions in Column (5). Municipality and year fixed effects are included. Standard Errors are clustered at the municipality level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.9: Heterogeneity based on baseline audit-based performance - Did Deliberation Affect Municipal Performance?

Dependent Variables	Below Mean Municipality Control Mean	Treatment	Treatment × Above Mean	Obs.	P-Value: T + T × Above Mean = 0
	(1)	(2)	(3)	(4)	(5)
Municipal Audit-Based Performance	63.138	10.779*** (2.605) {0.060}	-7.572*** (2.225) {0.960}	112	0.312
Sub-Components of Municipal Audit-Based Performance					
Record Keeping	63.800	18.583* (9.860) {0.400}	-18.562** (8.692) {0.800}	112	0.997
Transparency of Public Procurement	68.238	11.061* (5.623) {0.100}	-10.110 (6.057) {0.620}	112	0.848
Overall Management	65.000	39.935** (17.788) {0.140}	-42.201** (17.554) {0.380}	72	0.825

Note: The unit of observation is municipality-year. The data is from the Audit Reports of 2014-2019. Municipal Audit-Based Performance is a score constructed from 3 sub-components: (i) Record Keeping (ii) Transparency of Public Procurement (iii) Overall Management. Record Keeping is constructed from scores on: maintenance of accounting records, existence of documentation room/archive accessible to the public and equipped with storage furniture, statements of transferred resources, updated physical registers and stock-taking by the authorized officers. Transparency of Public Procurement is constructed from scores given on: Secretariat of the organizations responsible for public procurement is functional, the number and exhaustiveness of the reports published by them, details on FaDeC funds and spending, compliance with public procurement procedures, quality of accounts and traceability of transfers. Overall Management is constructed from scores given on effectiveness of the role of coordinator of local services by the Secretary General, and management of certificates of civil status (certificates of birth, death, marriage and divorce). Above and below mean performing municipalities are based on baseline data. We report *p*-value in Column (4) and *p*-value from Randomization Inference with 1,000 repetitions in Column (5). All regressions include municipality and year fixed effects. Standard errors are reported in parenthesis, while the Randomization Inference P-Values (1,000 reptitions) are reported in square braces. Standard Errors are clustered at the municipality level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.10: Did Bureaucratic Deliberation Affect Citizens' Outcomes (2014, 2017)?

Dependent Variables	Pre-Period Control Mean	DID Estimate	Standard Error	P-Value	Randomization Inference P-Value	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Aggregate Index						
Citizen Experience Index	-	0.1668	0.1498	0.281	0.420	456
Panel A : Corruption						
Corrupt among Govt Officials	0.960	0.0005	0.0407	0.990	0.980	444
Officials Go Unpunished	0.892	-0.0196	0.0472	0.683	0.640	449
Panel B : Present Living Conditions						
Are Good	0.279	0.0516	0.1127	0.653	0.620	456
Better Compared to Others	0.233	0.0961	0.0705	0.191	0.180	455
Panel C : Difficulty in Access to						
Identity Document Services	0.652	-0.0675	0.1595	0.678	0.700	206
Household Services	0.857	0.0159	0.1434	0.913	0.900	88
Panel D : Paid Bribes for						
Identity Document Services	0.304	-0.0141	0.1250	0.912	0.880	206
Household Services	0.143	-0.4373	0.2576	0.110	0.120	88

Note: The unit of observation is a citizen-year. The analysis is based on Afrobarometer data for the years 2014 and 2017. Panel A consists of indicators of whether the citizens perceive that there is corruption in council and government officials and whether the latter go unpunished. Panel B consists of indicators for whether the citizens feels that their present living conditions are good and are better compared to others. Panel C & D consist of indicators of whether citizens faced difficulties in accessing (i) identity documents like a birth certificate, driver's license, passport, voter's card or permit from government and (ii) household services (which includes water, electricity and sanitation and had to pay bribes for it respectively). We report *p*-value in Column (4) and *p*-value from Randomization Inference with 10,000 repetitions in Column (5). Standard Errors are clustered at the municipality level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.11: Heterogeneity based on baseline citizen experience

Dependent Variables	Below Mean Municipality Control Mean	Treatment	Treatment × Above Mean	Obs.	P-Value: T + T × Above Mean = 0
	(1)	(2)	(3)	(4)	(5)
Aggregate Index					
Citizen Experience Index	-0.407	0.285*	-0.256***	432	0.845
		(0.149)	(0.0790)		
		{0.185}	{0.726}		

Note: The unit of observation is a citizen-year. The analysis is based on Afrobarometer data for the years 2014 and 2017. Panel A consists of indicators of whether the citizens perceive that there is corruption in council and government officials and whether the latter go unpunished. Panel B consists of indicators for whether the citizens feels that their present living conditions are good and are better compared to others. Panel C & D consist of indicators of whether citizens faced difficulties in accessing (i) identity documents like a birth certificate, driver's license, passport, voter's card or permit from government and (ii) household services (which includes water, electricity and sanitation and had to pay bribes for it respectively). Above and below mean municipalities are based on baseline data (2014) from the Afrobarometer. We report *p*-value in Column (4) and *p*-value from Randomization Inference with 10,000 repetitions in Column (5). Standard Errors are clustered at the municipality level. Significance levels are denoted as: * p<0.1, ** p<0.05, *** p<0.01.

Table A.12: Heterogeneity for sub-components of citizen experience index based on baseline citizen experience

Dependent Variables	Below Mean Municipality Control Mean	Treatment	Treatment × Above Mean	Obs.	P-Value: T + T × Above Mean = 0
	(1)	(2)	(3)	(4)	(5)
Panel A: Corruption					
Corrupt among Govt Officials	1	-0.0293	0.0646*	420	0.427
		(0.0385) {0.630}	(0.0335) {0.591}		
Officials Go Unpunished	0.894	-0.0381	0.0402	425	0.970
		(0.0545) {0.626}	(0.0566) {0.895}		
Panel B : Present Living Conditions					
Are Good	0.0833	0.146 (0.102) {0.430}	-0.206* (0.114) {0.720}	432	0.674
Better Compared to Others	0.125	0.163**	-0.145*	431	0.832
		(0.0744) {0.0630}	(0.0845) {0.228}		
Panel C : Difficulty in Access to					
Identity Document Services	0.680	-0.0119	-0.121	195	0.375
		(0.188) {0.965}	(0.119) {0.756}		
Household Services	0.778	0.0118	0.0105	86	0.896
		(0.178) {0.945}	(0.200) {0.907}		
Panel D : Paid Bribes for					
Identity Document Services	0.200	0.0858	-0.218*	195	0.340
		(0.114) {0.533}	(0.114) {0.284}		
Household Services	0.222	-0.380	-0.147	86	0.0520
		(0.288) {0.150}	(0.210) {0.730}		

B Appendix Figures

Figure B.1: Collective Deliberation Meetings

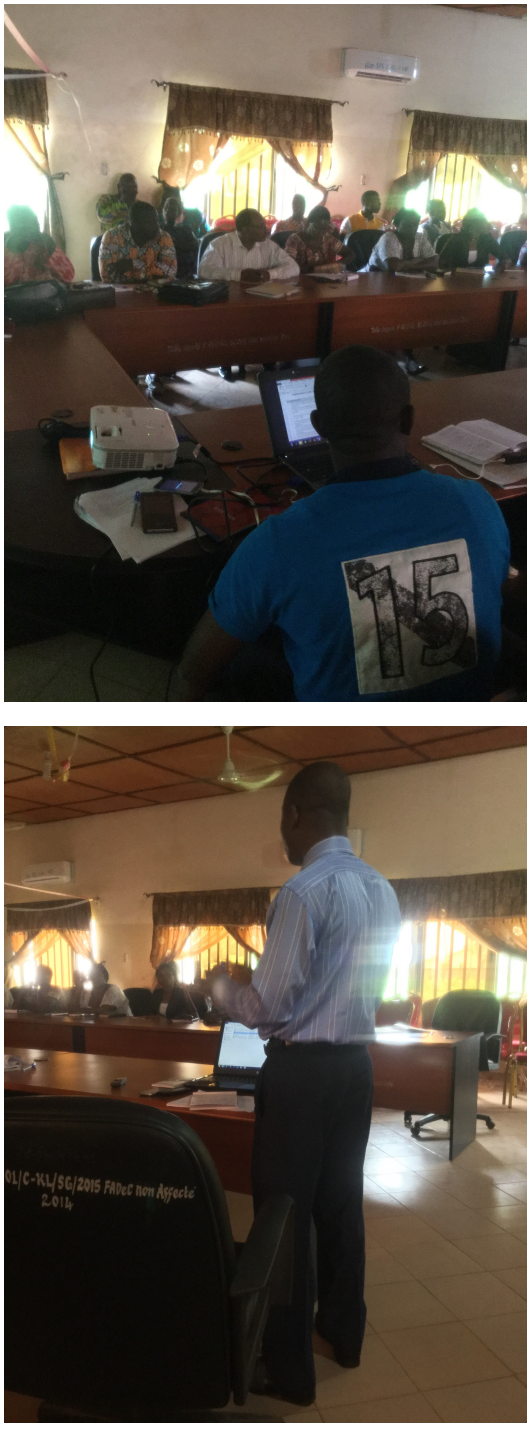
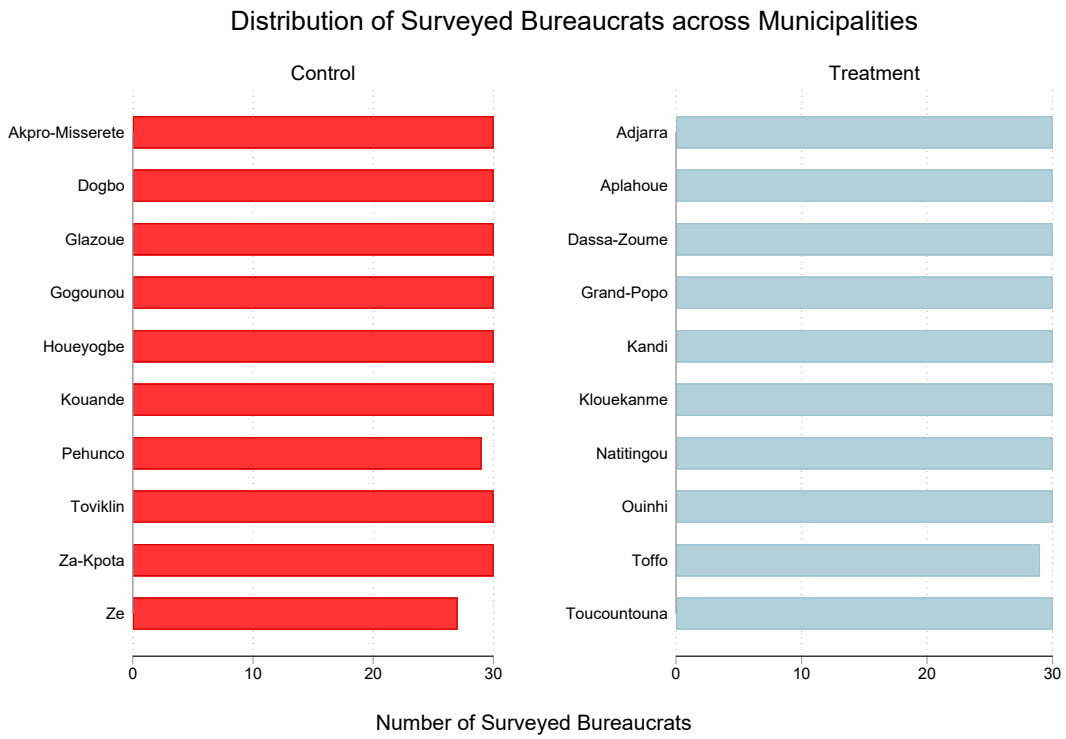
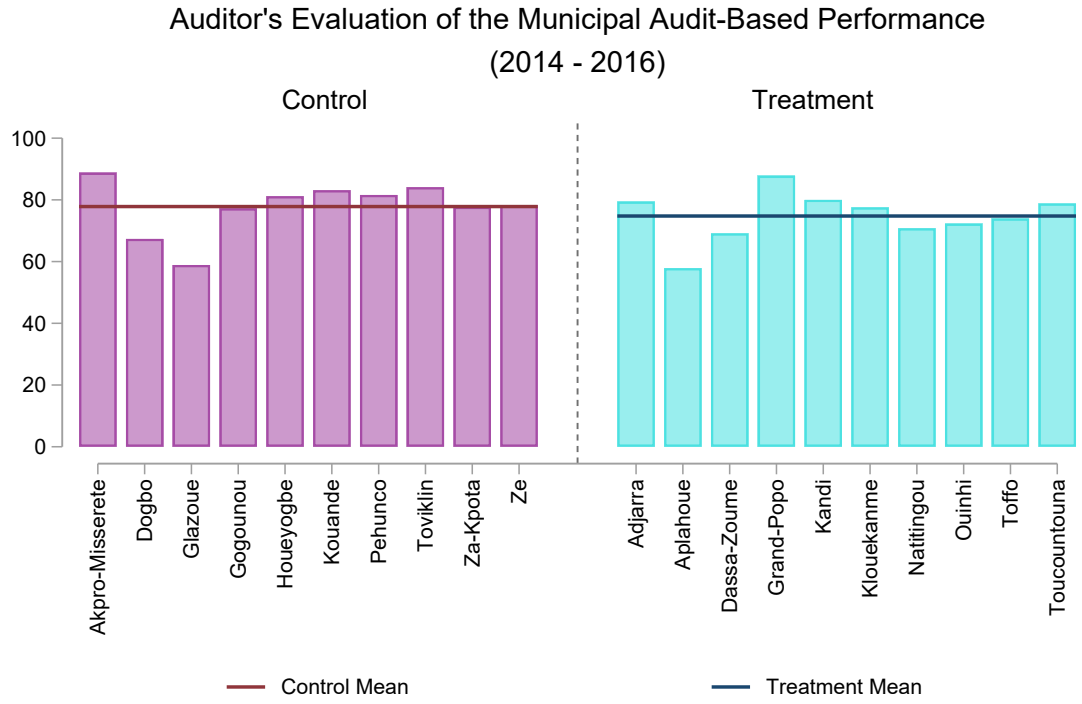


Figure B.2: Number of bureaucrats surveyed across treatment and control municipalities in December 2016



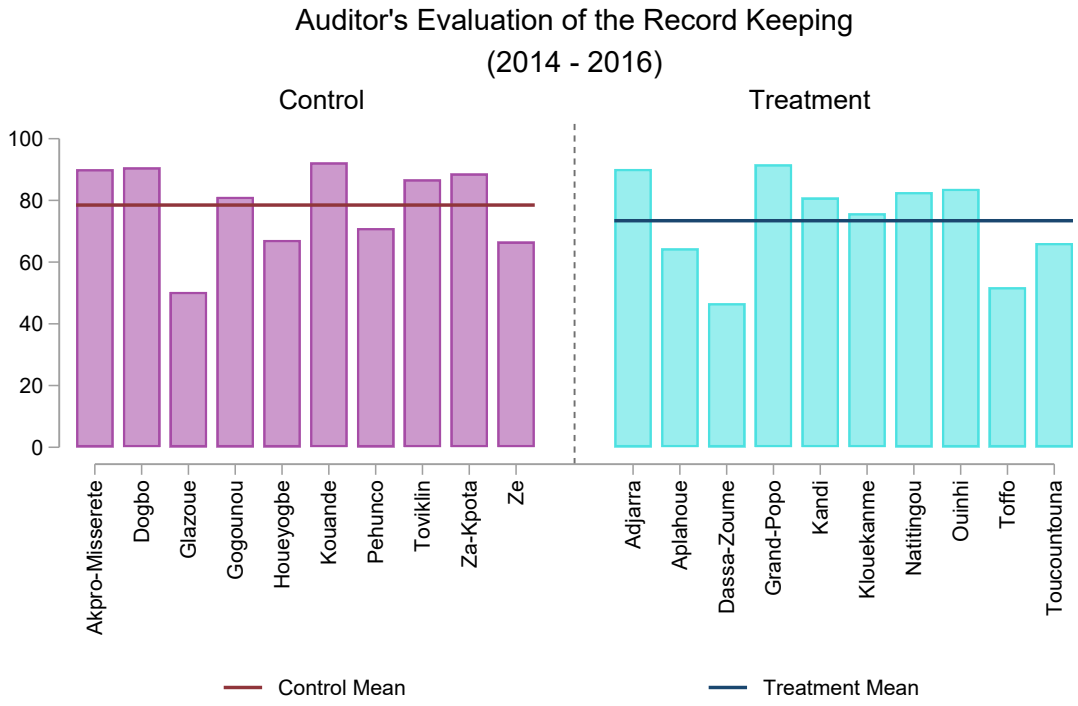
Notes. For the endline survey conducted in December 2016, 30 bureaucrats were randomly selected within each of the 20 municipalities. However, there were some exceptions: 27 in Ze and 29 each in Pehunco & Toffo, bringing the total number of observations to 595.

Figure B.3: Baseline Aggregate Average of Municipal Audit-Based Performance



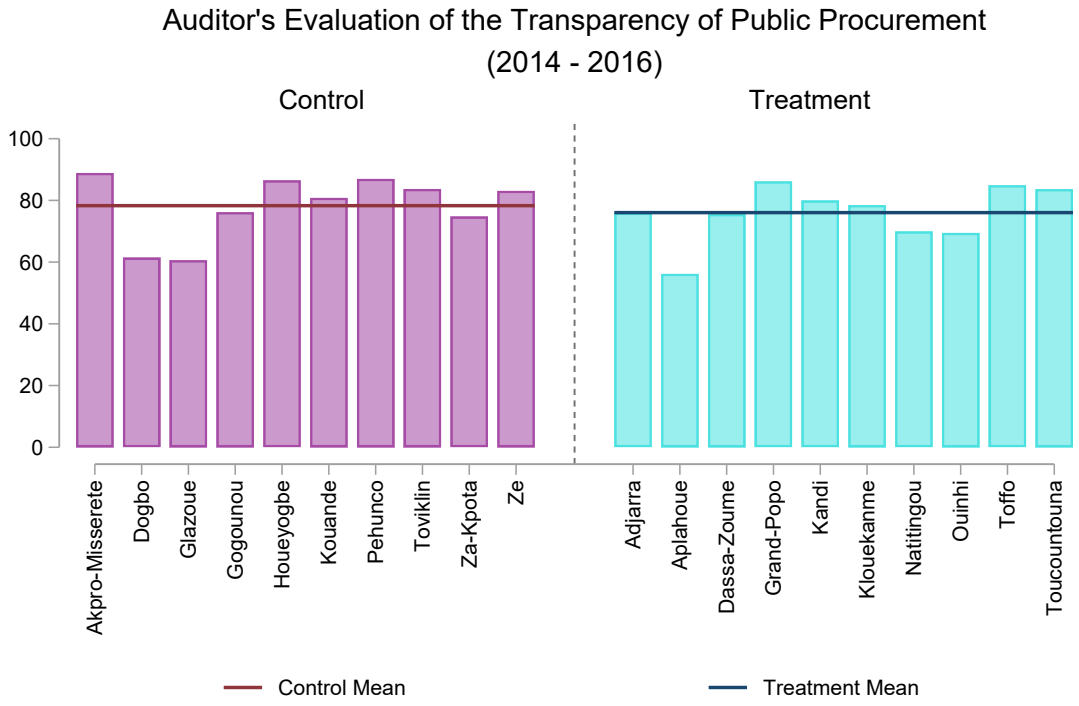
Notes. Municipal Audit-Based Performance is a score constructed from three sub-components: (i) Record Keeping (ii) Transparency of Public Procurement (iii) Overall Management. Record Keeping is constructed from scores on: maintenance of accounting records, existence of documentation room/archive accessible to the public and equipped with storage furniture, statements of transferred resources, updated physical registers and stock-taking by the authorized officers. Transparency of Public Procurement is constructed from scores given on: Secretariat of the organizations responsible for public procurement is functional, the number and exhaustiveness of the reports published by them, details on FaDeC funds and spending, compliance with public procurement procedures, quality of accounts and traceability of transfers. Overall Management is constructed from scores given on effectiveness of the role of coordinator of local services by the Secretary General, and management of certificates of civil status (certificates of birth, death, marriage and divorce.

Figure B.4: Baseline Average of Auditor's Evaluation of Record Keeping by Bureaucrats



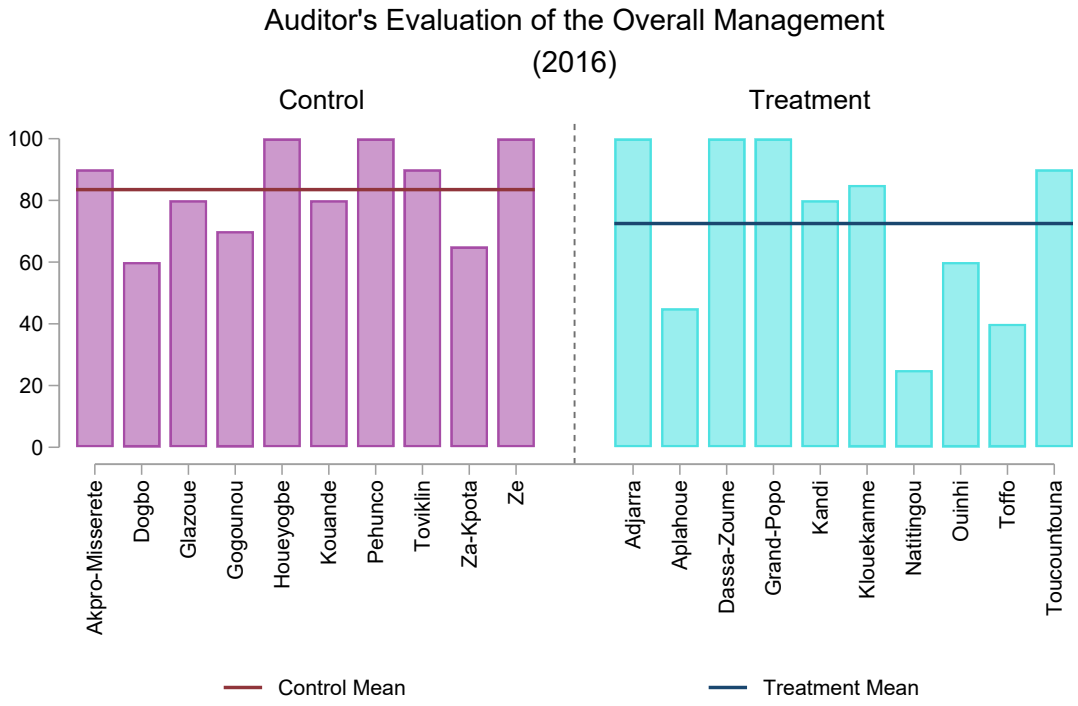
Notes. Record Keeping is constructed from scores on: maintenance of accounting records, existence of documentation room/archive accessible to the public and equipped with storage furniture, statements of transferred resources, updated physical registers and stock-taking by the authorized officers.

Figure B.5: Baseline Average of Auditor's Evaluation of Transparency of Public Procurement by Bureaucrats



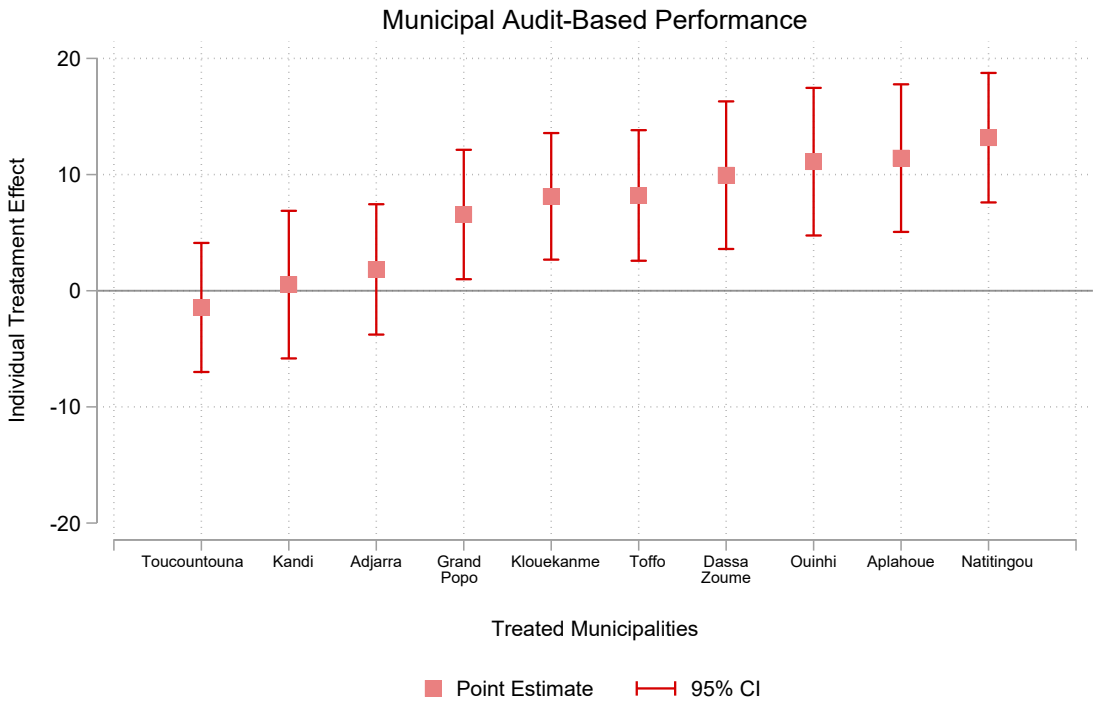
Notes. Transparency of Public Procurement is constructed from scores given on: Secretariat of the organizations responsible for public procurement is functional, the number and exhaustiveness of the reports published by them, details on FaDeC funds and spending, compliance with public procurement procedures, quality of accounts and traceability of transfers.

Figure B.6: Baseline Average of Auditor's Evaluation of Overall Management by Bureaucrats



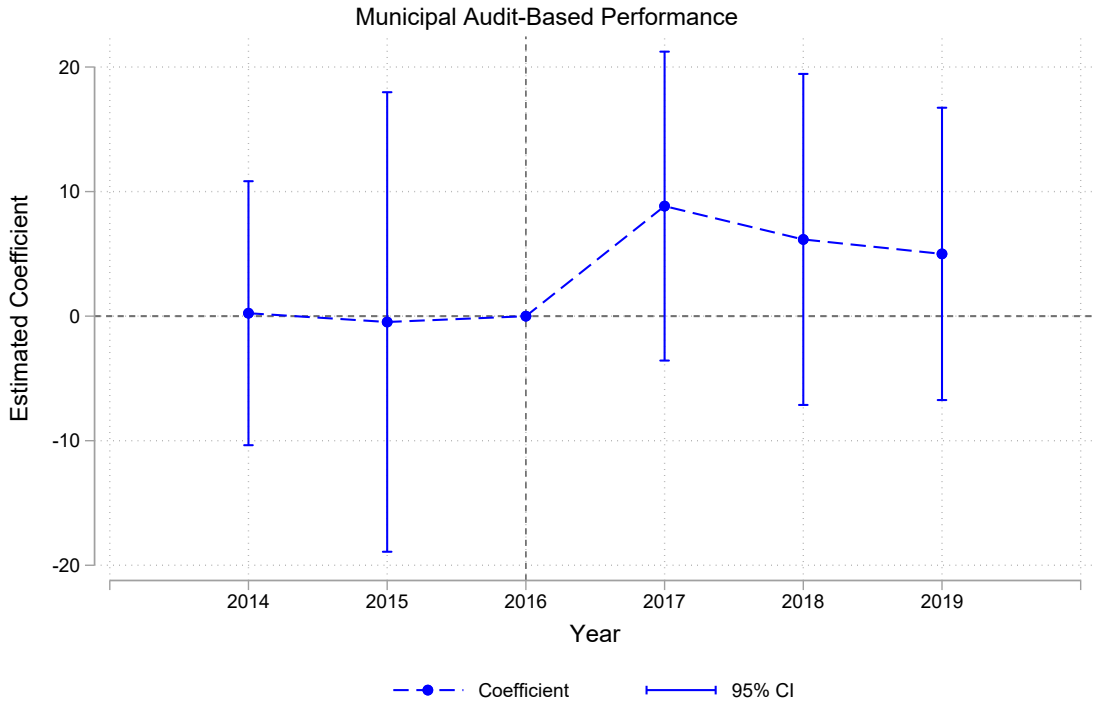
Notes. Overall Management is constructed from scores given on effectiveness of the role of coordinator of local services by the Secretary General, and management of certificates of civil status (certificates of birth, death, marriage and divorce)

Figure B.7: Treatment Effects in the different Treated Municipalities



Notes: The figure shows estimated treatment effects on municipal audit-based performance for each treated municipality individually, using a difference-in-differences specification with municipality and year fixed effects.

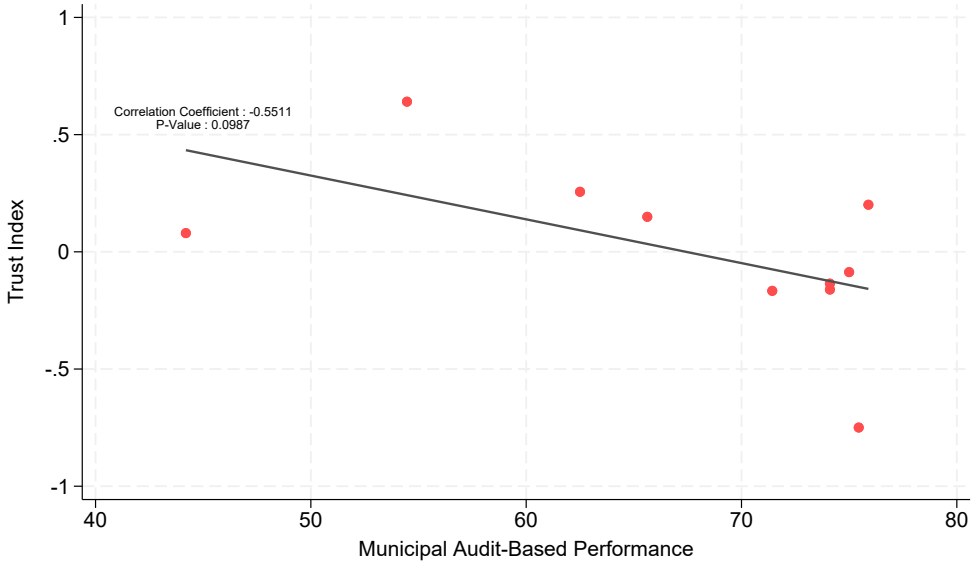
Figure B.8: Event Study Plot of Municipal Audit-Based Performance



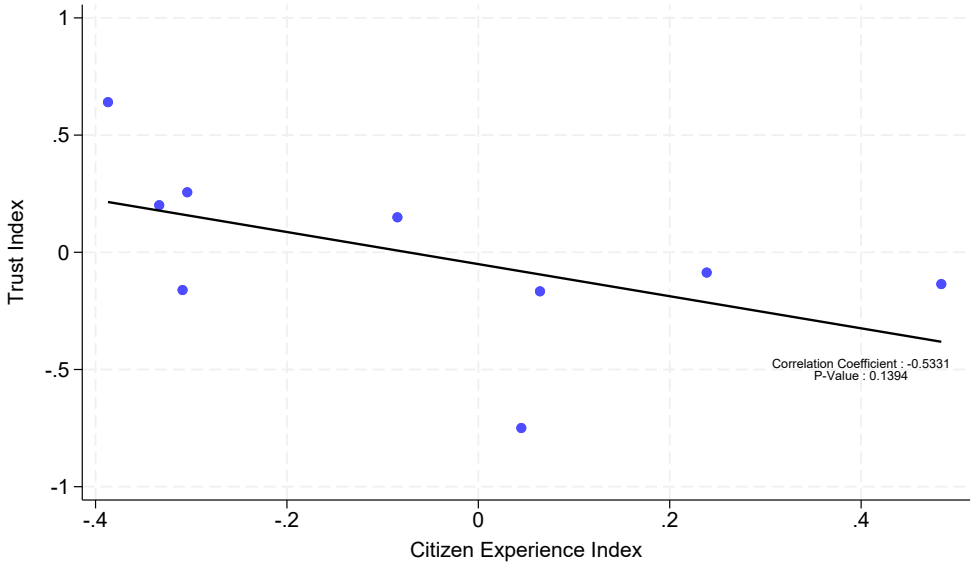
Notes. Municipal Audit-Based Performance is a score built from the audit reports of FADeC. It score is constructed from three sub-components: (i) Record Keeping (ii) Transparency of Public Procurement (iii) Overall Management. Record Keeping is constructed from scores given based on availability of quarterly and annual execution statements of transferred resources, functionality of an archiving and documentation device, updating of physical records and existence of stock records, updating of physical registers by the authorizing officer, and proper maintenance of accounting records. Transparency of Public Procurement is constructed from scores given based on legal existence and functionality of the main PM bodies and functionality of the S/PRMP, compliance with public procurement procedures, functionality of public procurement award and control bodies, availability of the FADeC enforcement point, quality of the administrative account and traceability of transfers, and execution of public orders. Overall Management is constructed from scores given based on effectiveness of the role of coordinator of local services by the Secretary General, and management of civil status documents.

Figure B.9: Baseline Relationship between Trust and Performance Outcomes

(a) Trust and Audit-Based Performance

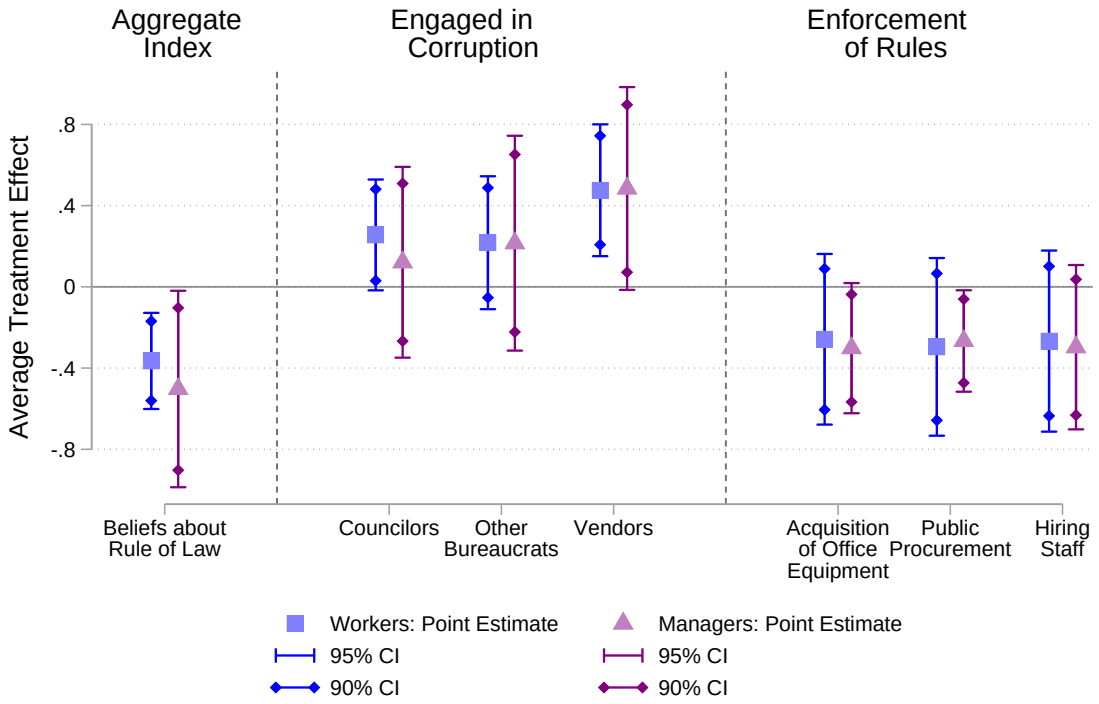


(b) Trust and Citizen Experience



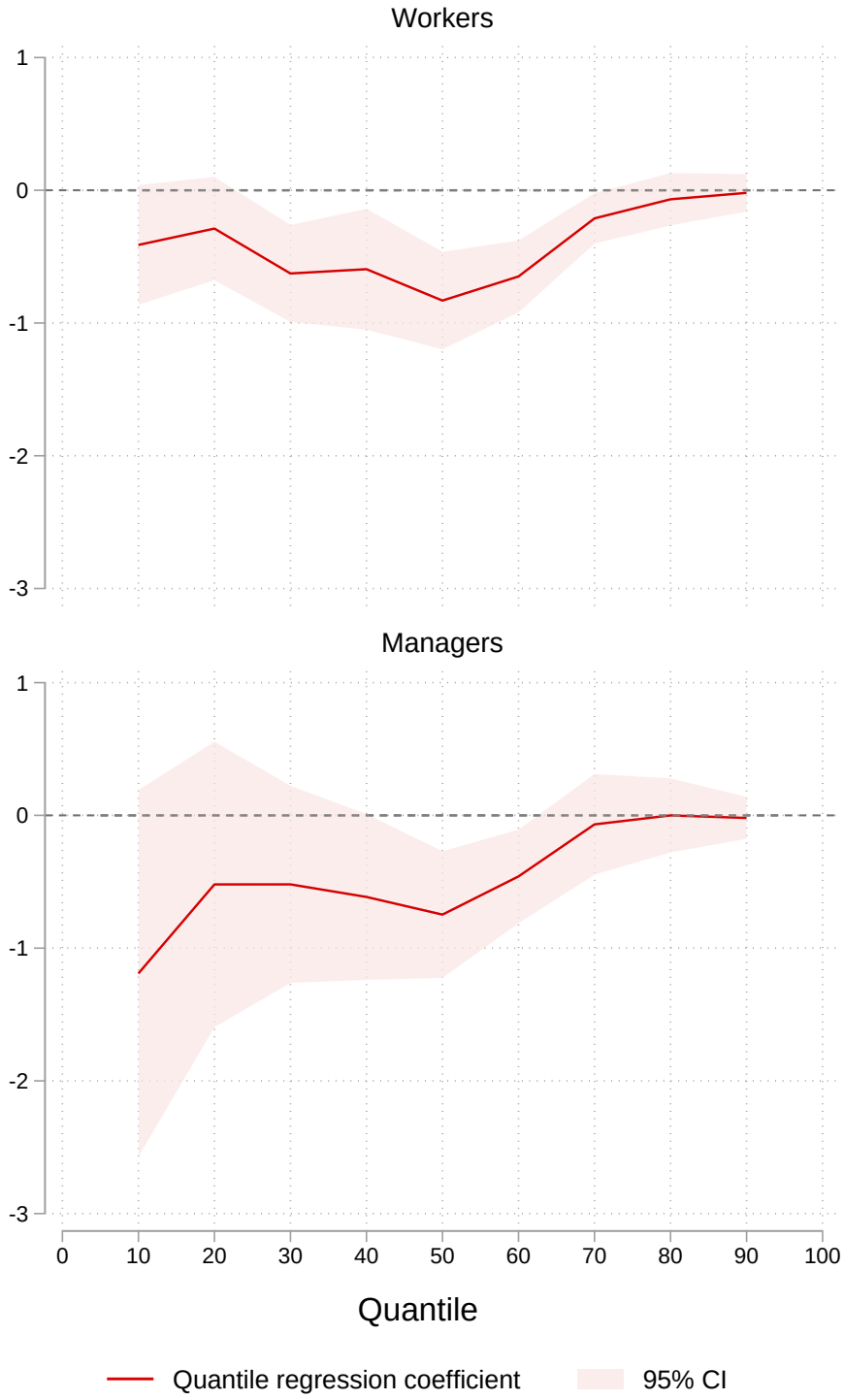
Notes: The top panel plots the relationship between the trust index and municipal audit-based performance. The bottom panel plots the relationship between the trust index and the citizen experience index.

Figure B.10: Workers and managers: Effects of deliberation on bureaucratic beliefs about rule of law



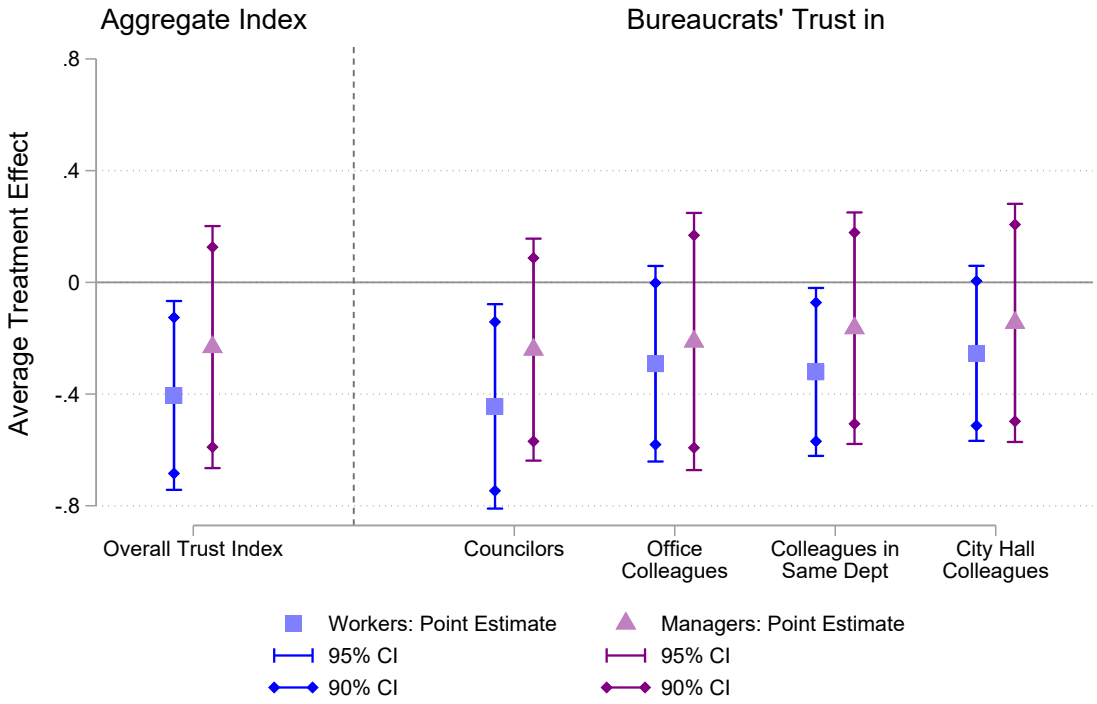
Notes: The figure reports estimates of the effect of deliberation on beliefs about corruption and enforcement of rules, separately for workers and managers, with 90% and 95% confidence intervals. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality level.

Figure B.11: Workers and managers: Quantile estimates of the effects of deliberation on bureaucratic beliefs about rule of law



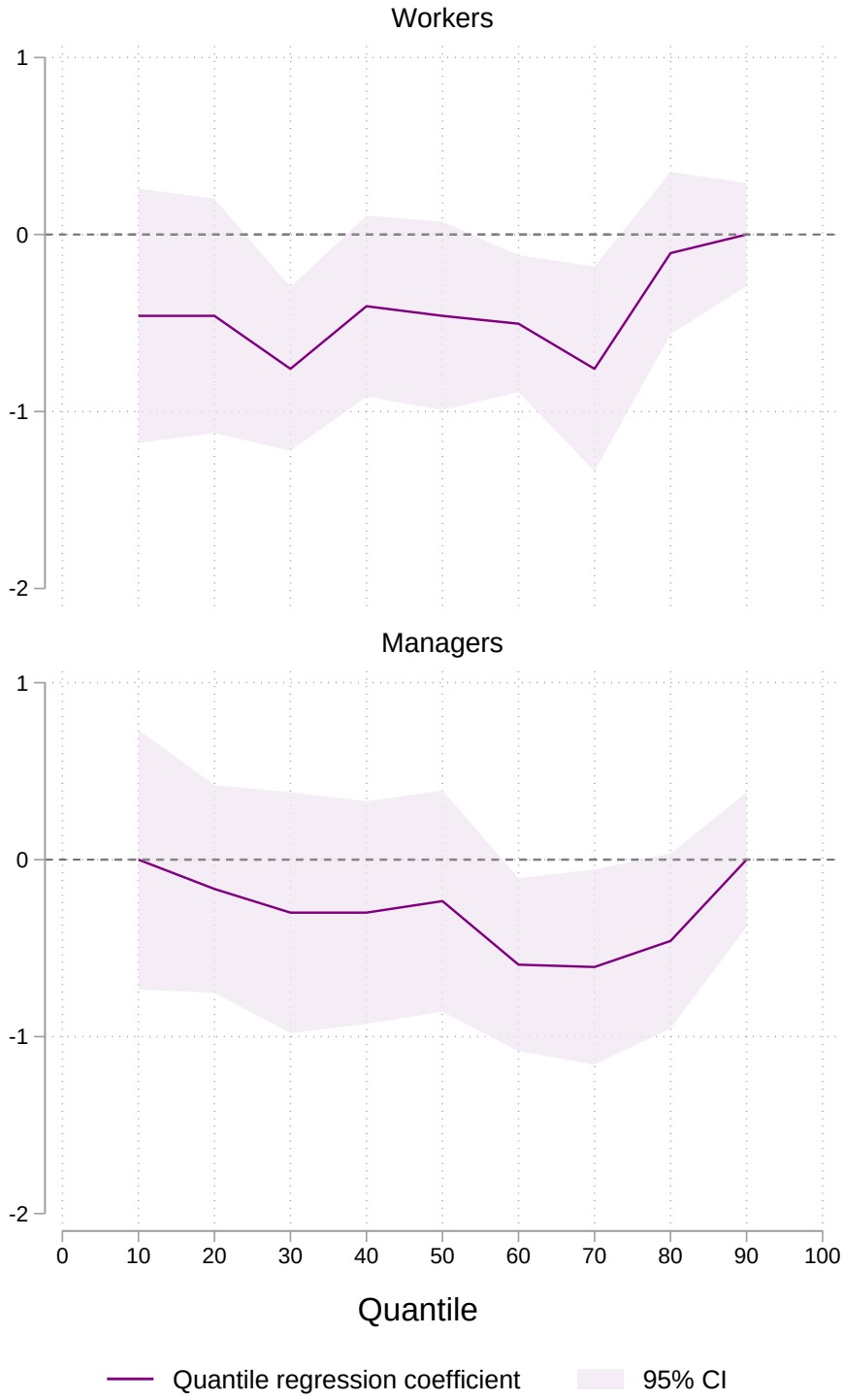
Notes: The figure reports quantile regression estimates of the treatment effect on the Beliefs about Enforcement index, separately for workers and managers, across quantiles 10 to 90. The shaded area represents the 95% confidence interval. Standard errors are clustered at the municipality level.

Figure B.12: Workers and managers: Effects of deliberation on bureaucratic trust in others



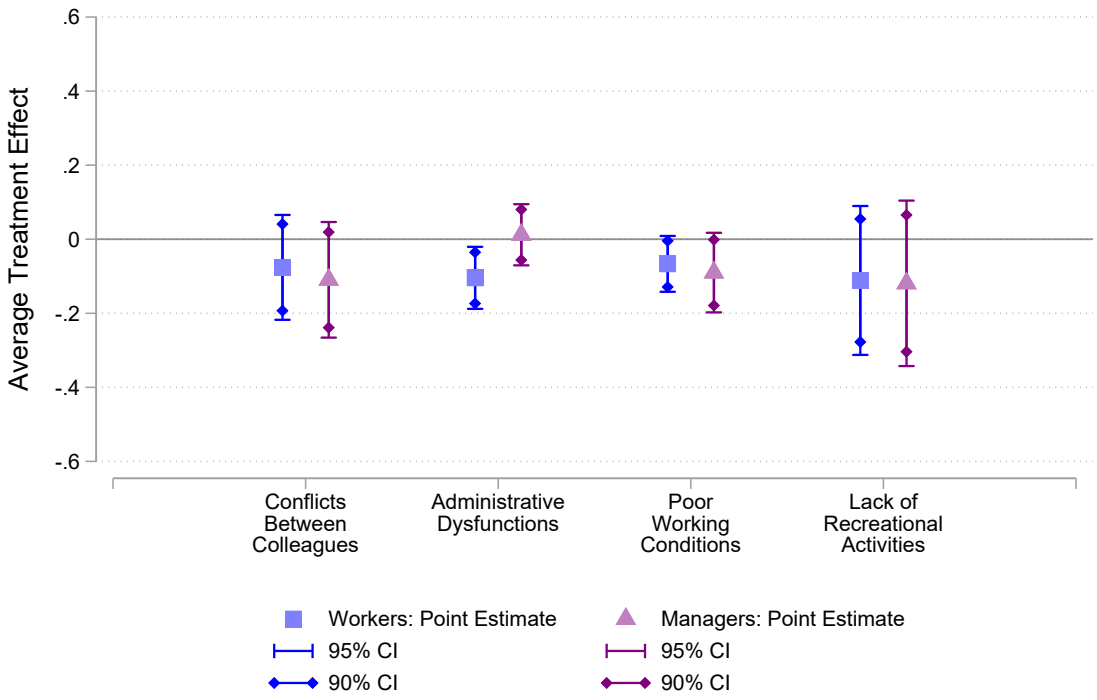
Notes: The figure reports estimates of the effect of deliberation on trust outcomes, separately for workers and managers, with 90% and 95% confidence intervals. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality level.

Figure B.13: Workers and managers: Quantile estimates of the effects of deliberation on bureaucratic trust in others



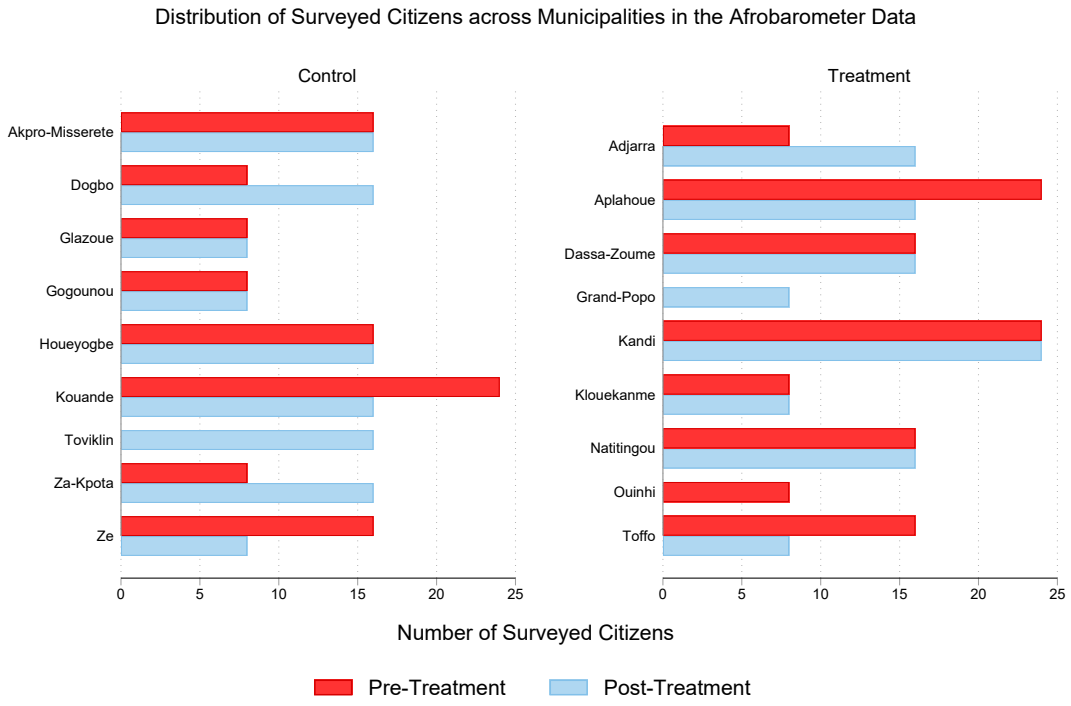
Notes: The figure reports quantile regression estimates of the treatment effect on the Trust index, separately for workers and managers, across quantiles 10 to 90. The shaded area represents the 95% confidence interval. Standard errors are clustered at the municipality level.

Figure B.14: Workers and managers: Issues meetings helped solve



Notes: The figure reports estimates of the effect of deliberation on whether meetings helped resolve workplace issues, separately for workers and managers, with 90% and 95% confidence intervals. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality level.

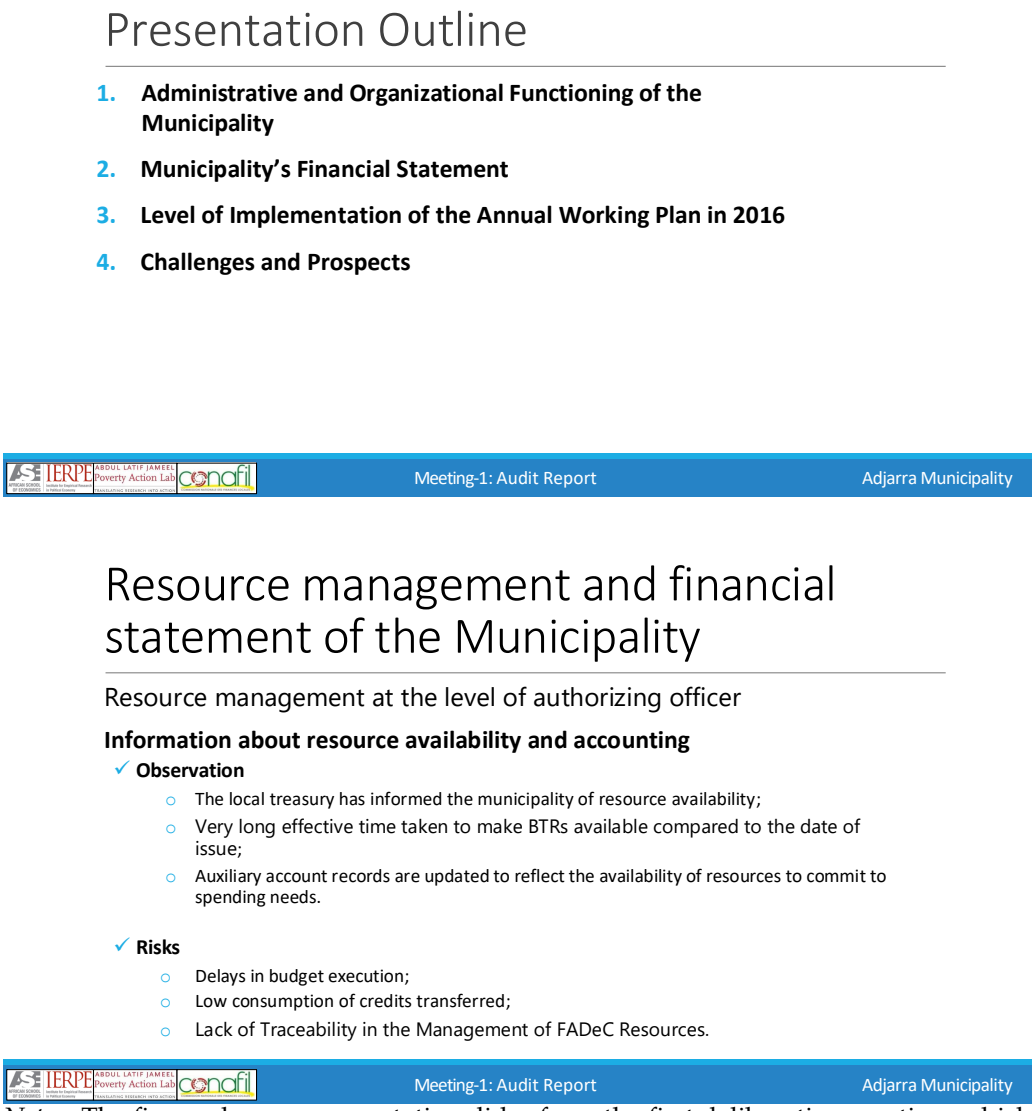
Figure B.15: Number of surveyed citizens across treatment and control municipalities in the Afrobarometer Data



Notes. The sampling frame is a region, constituency and urban-rural location (strata). Within each stratum, clustered random sampling takes place: (i) randomly selecting start points within each PSU, (ii) followed by randomly selecting households from each start point and (iii) finally randomly selecting respondents within a household. There are 8 households surveyed per PSU in each round. We use round 6 (2014) and round 7 (2017) as pre and post treatment periods respectively for our analysis. Note that there are no observations sampled for Pehunco and Toucountouna for either of the rounds. No observations were sampled for round 6 for Toviklin and Grand-Popo.

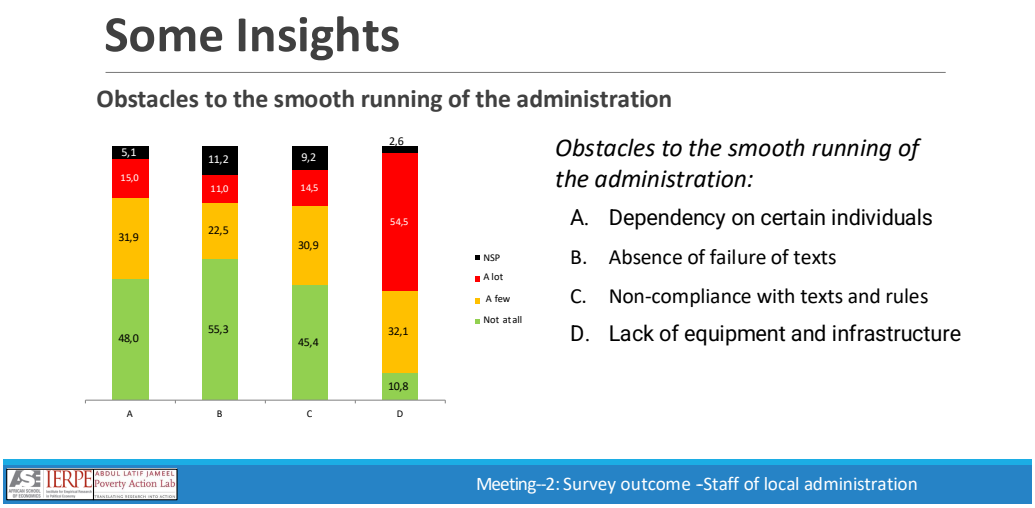
C Deliberation Meeting Materials

Figure C.16: Meeting 1: Audit-based presentation



Notes: The figure shows representative slides from the first deliberation meeting, which presented municipality-specific audit findings. Information was organized along administrative and financial dimensions and presented as observations and associated risks.

Figure C.17: Meetings 2 and 3: Survey-based presentations



Notes: The figure shows representative slides from the second and third deliberation meetings. Meeting 2 presented aggregate results from the bureaucrat survey, including staff knowledge, administrative processes, and internal constraints. Meeting 3 presented municipality-level results from the citizen survey, including perceptions of governance, corruption, trust, and service delivery. Across meetings, the materials emphasized descriptive statistics rather than prescriptive recommendations.

D Audit-Based Performance Measures

This appendix provides details on the construction of the municipal audit-based performance measure used in the main analysis.

D.1 Data Source and Digitization

The audit data are drawn from annual reports produced by the National Commission for Local Finance (CoNaFiL). These reports are available in PDF format and were manually digitized to construct a municipality-year panel covering the period 2014–2019 for all municipalities in the experimental sample.

Digitization involved extracting component-level scores and total scores for each municipality and year. All entries were cross-checked for consistency across reports.

D.2 Performance Components

The audit reports evaluate municipal administration across multiple dimensions. Our primary performance measure aggregates three components:

1. *Record Keeping*
2. *Transparency in Public Procurement*
3. *Overall Administrative Management*

These components correspond to internal administrative processes within the municipal bureaucracy. Each component consists of multiple criteria evaluated by auditors, which vary slightly across years.

Tables D.13, D.14, and D.15 provide detailed descriptions of the criteria and scoring structure for each component.

D.3 Construction of the Performance Measure

For each municipality-year, we compute a performance score by aggregating the scores across the three components described above.

$$Municipal\ Audit - Based\ Performance_{mt} = \frac{RawScore_{mt}}{MaxScore_t} \times 100$$

where $RawScore_{mt}$ is the total score obtained by municipality m in year t, and $MaxScore_t$ is the maximum attainable score in that year. Table D.16 summarizes the number of criteria and maximum attainable scores across years.

In a small number of cases where scores are missing in the audit report for a given year, we recover values using information from subsequent audit reports. These reports include tables documenting year-to-year performance changes, which allow reconstruction of missing values.

D.4 Details of the criteria used by auditors

Table D.13: Details of the criteria used by auditors to audit the local government's record keeping

Year	Record Keeping Criterion	Maximum Score	Total Score
2014	Proper maintenance of accounting records	5	5
2015	Proper maintenance of accounting records	5	5
2016	Features of the archiving and documentation system*	5	17‡
	Availability of quarterly and annual execution statements (financial and physical) of transferred resources	7†	
	Updation of physical registers by the authorizing officer	5	
2017	Features of the archiving and documentation system	5	16
	Availability of quarterly and annual execution statements (financial and physical) of transferred resources	6	
	Updation of physical registers by the authorizing officer	5	
2018	Features of the archiving and documentation system	5	16
	Availability of quarterly and annual execution statements (financial and physical) of transferred resources	6	
	Updation of physical records and existence of stock records	5	
2019	Features of the archiving and documentation system	5	16
	Availability of quarterly and annual execution statements (financial and physical) of transferred resources	6	
	Updation of physical records and existence of stock records	5	

*It involves existence of documentation room accessible to the public and equipped with storage furniture

†Exceptions: 10 for Kouande, Pehunco and Ze

‡Exceptions: 20 for Kouande, Pehunco and Ze

Table D.14: Details of the criteria used by auditors to audit the local government's transparency of public procurement

Year	Transparency of Public Procurement Criterion	Maximum Score	Total Score
2014	Existence & functioning of the main bodies as well as the Secretariat responsible for public procurement[§]	9	15
	Availability of details regarding the spending of the FaDeC funds for the auditors	6	
2015	Existence & functioning of the main bodies as well as the Secretariat responsible for public procurement	9	15
	Availability of details regarding the spending of the FaDeC funds for the auditors	6	
2016	Functioning of the bodies responsible for public procurement and control of public contracts[¶]	10	30
	Compliance with public procurement procedures	20	
2017	Functioning of the bodies responsible for public procurement and control of public contracts	6	35
	Compliance with public procurement procedures	10	
	Execution of public purchases	15	
	Quality of the administrative account and traceability of transfers	4	
2018	Functioning of the bodies responsible for public procurement and control of public contracts	6	35
	Compliance with public procurement procedures	10	
	Execution of public purchases	15	
	Quality of the administrative account and traceability of transfers	4	
2019	Functionality of public procurement award and control bodies	5	34
	Compliance with public procurement procedures	10	
	Execution of public purchases	15	
	Quality of the administrative account and traceability of transfers	4	

[§]There are 3 main bodies responsible for public procurement: Secretariat of the Person Responsible for Public Procurement, Public Procurement Commission and Public Procurement Control Unit. 3 points each for the functioning of each of these bodies.

[¶]Points are assigned on the number and exhaustiveness of the reports published by the aforementioned bodies.

Table D.15: Details of the criteria used by auditors to audit the local government's overall management

Year	Overall Management Criterion	Maximum Score	Total Score
2014	N/A	N/A	N/A
2015	N/A	N/A	N/A
2016	Effectiveness of the role of coordinator of local services by the Secretary General	5	5
2017	Effectiveness of the role of coordinator of local services by the Secretary General	5	5
2018	Effectiveness of the role of coordinator of local services by the Secretary General	5	5
2019	Effectiveness of the role of coordinator of local services by the Secretary General	5	10
	Management of Civil Status documents*	5	

*These are vital records of life events kept under governmental authority, including birth certificates, marriage licenses, divorce certificates and death certificates.

Table D.16: Summary of the criterion used by auditors across years

Indices / Sub-Indices		Year					
		2014	2015	2016	2017	2018	2019
Record Keeping	Number of Criteria	1	1	3	3	3	3
	Maximum Score	5	5	17†	16	16	16
Transparency of Public Procurement	Number of Criteria	2	2	2	4	4	4
	Maximum Score	15	15	30	35	35	34
Overall Management	Number of Criteria	0	0	1	1	1	2
	Maximum Score	N/A	N/A	5	5	5	10
Municipal Audit-Based Performance	Number of Criteria	3	3	6	8	8	9
	Maximum Score	20	20	52**	56	56	60

E Bureaucrat Survey Measures

The bureaucrat survey was conducted in December 2016 across all municipalities in the study. Respondents were randomly sampled from municipal personnel rosters, irrespective of participation in the deliberation meetings.

E.1 Belief Measures

Beliefs about Rule Enforcement. Respondents were asked:

"Would you say that the following rules are respected?"

†Exceptions: 20 for Kouande, Pehunco and Ze
**Exceptions: 55 for Kouande, Pehunco and Ze

- The rules governing the acquisition of office equipment on behalf of the municipal administration
- The rules governing public procurement
- The procedures for recruiting administrative staff

Responses are recorded on a three-point scale ranging from "not at all" to "fully."

Beliefs about Corruption. Respondents were asked:

> "Do you think the following people are engaged in corruption practices?"

- Members of the municipal council
- Colleagues within the same department
- Other colleagues within the town hall
- City Hall service providers (e.g., those who win public contracts at the City Hall)

Responses are recorded on a four-point scale ranging from "not at all" to "frequently."

These two sets of questions are combined to form one standardized belief index. All components are coded so that higher values correspond to stronger perceived rule compliance and lower perceived corruption.

E.2 Trust

Trust is measured using responses to the following question:

> "Do you trust. . . ?"

- The municipal council
- Your office colleagues

- Colleagues in the same department
- Other staff of the town hall

Responses are recorded on a four-point scale ranging from "not at all" to "very much." These variables are standardized and aggregated into a trust index, with higher values indicating greater trust.

E.3 Workplace Interactions

Workplace interactions are measured using responses to the following question:

> "Here is a list of actions that people sometimes carry out together as staff of the same office. For each of them, please tell us how often the City Hall's staff carries it out in general."

- Joining with others to address a City Hall problem
- Organizing recreational, extra-professional activities with colleagues (e.g., excursion, picnic, Christmas for staff children, etc.)
- Organizing professional activities outside the work between colleagues (e.g., strategic retreat, workshop, etc.)
- Offering to perform the task of an indisposed colleague
- In case of conflict or misunderstanding with a colleague, contact the colleague directly to discuss it and find a solution
- Refer to the hierarchy to report a problem between colleagues for resolution.

Responses are recorded on a four-point scale ranging from "never" to "always." These variables are standardized and aggregated into a workplace interaction index.

E.4 Meeting Practices

To capture meeting activity within the administration, respondents were asked:

> "In the last three months, how many meetings have you attended in your City Hall?"

E.5 Meeting Content

Meeting content is measured using responses to the following question:

> "What were the topics discussed during these meetings?"

For each topic, respondents indicate whether it was discussed:

- City Hall FADeC audit report
- Financial situation of the municipality and implementation of the PTA
- Challenges and perspectives of the City Hall
- Functioning of local government (procedures, relationships, knowledge)
- Taxation and local finances
- Performance of the municipal administration and council
- Good governance within the City Hall
- Corruption in the management of municipal affairs
- Perceptions of corruption among local development actors
- Other topics

These measures allow us to distinguish between changes in the quantity of meetings and changes in their content, and provide proxies for monitoring-related interactions within the workplace.

E.6 Index Construction

All indices are constructed by standardizing individual components to mean zero and unit variance and aggregating them using inverse-covariance weighting following Anderson (2008b).