

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

Shan Aman-Rana[†] Leonard Wantchekon[‡] Lazare Kovo[§]

May 23, 2026

[LATEST VERSION CLICK HERE](#)

Abstract

Bureaucratic performance depends not only on individual incentives and ability, but also on beliefs about how others within the organization behave. We examine whether changing these beliefs through deliberation — an institutional process in which objective, performance-relevant information about the bureaucracy as a whole is made public and discussed collectively — can improve performance. We conduct a randomized controlled trial with municipal administrations in Benin. Using administrative audit data from 2014 to 2019, we find that deliberation increases municipal performance by 9 percent of the control mean, with larger effects among municipalities with lower baseline performance. Deliberation also reshapes the organizational environment: bureaucrats negatively update their beliefs about corruption and rule enforcement, these beliefs become less dispersed within municipalities, managerial attention shifts toward audit findings and corruption, and interpersonal trust declines. These results highlight the potential for low-cost institutional changes to improve bureaucratic performance, while pointing to a fundamental feature of weakly institutionalized settings: improving performance may come at the cost of disrupting interpersonal trust and relational arrangements.

JEL codes: D73, H11, O1

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

*We are grateful to Michael Best, Nick Bloom, Gaurav Chiplunkar, Guido Friebel, Virginia Minni, Clement Minaudier, Priyama Majumdar, Diana Moreira, John Van Reenen, Raffaella Sadun, Megha Patnaik, Silvia Vannutelli, Gerard Padro i Miquel, Paola Sapienza, Sheetal Seekhri, Ivo Welch, Arthur Zeyang Yu, Sandip Sukhtankar, Shaoda Wang, Yanhui Wu and participants at seminars and workshops for comments and suggestions. Sanjog Rajiv, Saad Hassan, 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 a central determinant of economic development (Besley and Persson, 2009, 2010; Evans, 1995; Evans and Rauch, 1999; Kohli, 2004). A large literature studies bureaucratic underperformance by focusing on the role of individual incentives and ability in shaping bureaucratic behavior.¹

While these approaches have generated important insights, they largely operate within a fixed institutional environment. In practice, bureaucratic behavior is embedded within organizations, where actions depend on beliefs about how others behave. Underperformance may therefore persist not only because of hidden action, but because it reflects an equilibrium sustained by organizational expectations and informal relational arrangements. This perspective points to a complementary policy lever: modifying institutional processes, or the “rules of the game,” within organizations as a way to improve performance (Callen et al., 2024).

In this paper, we conduct a field experiment with municipal bureaucracies in Benin to test whether changing organizational norms of performance can improve outcomes. In treated municipalities, we introduce deliberation as an institutional process in which objective, performance-relevant information about the bureaucracy as a whole, rather than about individual officials, is made public and discussed collectively. We find results on two dimensions. First, deliberation leads to meaningful improvements in municipal performance. Second, these improvements are accompanied by shifts in the organizational environment within the bureaucracy: bureaucrats negatively update their beliefs about corruption and rule enforcement, managers hold fewer meetings but discussions shift toward audit findings and corruption, and interpersonal trust declines. These findings highlight the potential for low-cost institutional changes to improve bureaucratic performance, while pointing to a fundamental feature of weakly institutionalized settings: performance improvements can come at the cost of disrupting trust (Tirole, 1986).

Municipal bureaucrats in Benin play a central 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,

¹There is a large literature on bureaucratic incentives (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, 2025; Dipoppa and Gulzar, 2023; Deserranno et al., 2025a,b; Brown, 2022; Khan, 2025; Janvry et al., 2023), selection (Moreira and Pérez, 2024; Dal Bó et al., 2013; Ashraf et al., 2020; Bergeron et al., 2022; Jia et al., 2015; Aman-Rana, 2025; Riaño, 2021; Fenizia, 2022; Xu, 2018; Agte and Bedoya, 2023), and monitoring and political control (Olken, 2007; Lichand et al., 2016; Gulzar and Pasquale, 2017; Iyer and Mani, 2012; Akhtari et al., 2022; Dasgupta and Kapur, 2020).

and local infrastructure such as roads, schools, and health centers. Like many Weberian bureaucracies (Weber, 1922), these bureaucrats operate under weak formal incentive structures: wages are fixed, dismissal is rare, and promotions are largely seniority-based.

To better understand the organizational environment and relational arrangements within the bureaucracy, we focus on interpersonal trust and study its relationship with performance.² We measure interpersonal trust using a trust index composed of trust in councilors, office colleagues, colleagues within the same department, and city hall colleagues. We find that higher interpersonal trust is negatively associated with both municipal audit scores (-0.55 , p -value <0.1) and citizen experience with the government (-0.53 , p -value $=0.14$) in the control municipalities. These patterns suggest that in this setting, stronger interpersonal trust may coexist with weak accountability and low performance.

In partnership with the Government of Benin, we conducted a field experiment across 20 municipalities (ten treatment and ten control) randomly selected from the country's 77 jurisdictions. The sample spans 8 of the 12 departments in Benin.³

Our intervention was designed around four principles to ensure the deliberation was credible, inclusive, and conducive to honest reflection. First, performance data was generated by an independent third party to prevent participants from dismissing the findings as biased or politically motivated. Second, we employed a standardized 90-minute agenda to maintain comparability across treatment sites. Third, the sessions were inclusive of all municipal bureaucrats, with an aim for the information to reach the entire organization simultaneously. Finally, to foster candid communication the mayor were excluded from the sessions, and bureaucrats were tasked with guiding the deliberation. This structure created a protected space for transparent reflection on organizational deficiencies with a lower risk of repercussions from the mayor.

In the treated municipalities three deliberation meetings were held at monthly intervals, beginning in September 2016. 58% of all bureaucrats in the treated municipality attended at least one of the three sessions. Each meeting focused on a distinct theme: the first on the central government auditors' report for the municipality, the second on survey results from bureaucrats, and the third on survey results from citizens. All information was presented in aggregated form.⁴ The audit reports were already available within the

²Following Macchiavello and Morjaria (2022), we interpret interpersonal trust as partly reflecting informal organizational relationships within the bureaucracy.

³Implementing the experiment at 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).

⁴Citizen survey results were presented as municipality-level aggregates, while bureaucrat survey re-

administrative system but were not typically discussed collectively across the bureaucracy. The two surveys, conducted by the research team one month prior to the intervention, complemented these reports.

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

To guide interpretation, we develop a simple framework in which managers choose monitoring intensity based on beliefs about aggregate bureaucratic performance, while workers choose effort based on expected monitoring and informal relational arrangements within the bureaucracy. In environments where information about organizational performance is not collectively aggregated or discussed, organizational failures may remain fragmented and privately observed rather than collectively recognized. These conditions can sustain interpersonal trust and stable relational arrangements that weaken individual accountability and reduce incentives for monitoring or high effort. Deliberation acts as an organization-wide information shock by making performance-relevant information publicly available and subject to collective discussion. As organizational failures become collectively visible, managers may intensify monitoring and workers become less confident that low effort and weak rule enforcement will remain informally tolerated within the organization. Existing relational arrangements weaken, individual accountability rises, and effort increases even without changes in formal incentives. The same process disrupts the interpersonal trust and relational arrangements that previously sustained weak accountability and low performance within the bureaucracy.

We measure bureaucratic performance using municipality-level administrative audit data from 2014 to 2019. The audits assess core dimensions of bureaucratic functioning, including record keeping, transparency in public procurement, and overall administrative management. 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.⁵ Our primary outcome is a municipal-year panel of annual audit scores assigned to each municipality, expressed as a percentage of the maximum attainable score based on

sponses were aggregated across the twenty municipalities to limit social desirability bias.

⁵CoNaFiL's 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).

rules. Audit scores are strongly correlated at baseline with an index of citizen experiences with government services from Afrobarometer data (0.62, p -value = 0.01), indicating that the measure captures administrative practices that affect service delivery directly.

Our main result is that deliberation led to improvements in performance. After deliberation, municipal audit scores increase by 6.8 percentage points (8.7 percent relative to the control mean; randomization inference p -value < 0.05) in the treated municipalities relative to the control municipalities. These improvements are broadly distributed across treated units rather than driven by outliers, and are larger among municipalities that performed worse at baseline. Although we have less statistical power to detect effects for each year separately, the estimated performance effects remain economically meaningful for up to three years after treatment, suggesting that the results are unlikely to reflect only temporary responses to the deliberation sessions.

To examine how deliberation reshaped the organizational environment within the bureaucracy, we complement audit data with survey measures of bureaucrats' beliefs, trust, and workplace interactions, including the number and content of meetings held within the municipal administrations. Deliberation leads to a substantial negative update in beliefs about rule enforcement and corruption: bureaucrats in treated municipalities are 0.4 standard deviations less likely than those in the control group to report rule of law in the municipality, as measured by a combined index of perceived corruption and rule enforcement (randomization inference p -value < 0.01). These effects are not confined to the mean but extend across the belief distribution and to bureaucrats who did not attend the meetings. Deliberation also reduced the dispersion of beliefs within municipalities, with beliefs becoming significantly more aligned among attendees, particularly along the corruption dimension.

Managers were significantly more likely to attend the deliberation sessions and exhibit similar negative updates in beliefs about corruption and rule enforcement as workers. To examine whether these belief changes translated into changes in managerial behavior, we study the frequency and content of meetings held within the municipality, which provide a proxy for managerial oversight and accountability-related interactions within the bureaucracy. Excluding the deliberation sessions themselves, treated municipalities hold significantly fewer meetings, a decline driven primarily by managers rather than workers. This suggests that workers' day-to-day interactions remained broadly unchanged, while managers reduced general administrative meetings. At the same time, the content of meetings shifts toward discussions of audit findings and corruption, with no comparable change in other topics. Together, these patterns suggest not an increase in oversight

through more frequent interaction, but a reallocation of managerial attention toward accountability-related concerns.

We next examine the effects of deliberation on interpersonal trust among bureaucrats. Consistent with the conceptual framework, deliberation erodes trust within the bureaucracy, with trust falling by 0.35 standard deviations in the treated group relative to control (randomization inference $p = 0.06$). Trust declines are observed among both workers and managers. Similar to the effects on beliefs, these effects are not confined to the mean but extend across the distribution of trust and to bureaucrats who did not attend the meetings. These results suggest that in weakly institutionalized settings, improving performance may require disrupting informal relational arrangements and trust.

Our results contribute to three strands of the literature: state capacity and bureaucracies, institutions, and public deliberation.

We contribute to the literature on state capacity building (Besley and Persson, 2009, 2010; Bardhan, 2016; Pepinsky et al., 2017; Finan et al., 2017; Page and Pande, 2018; Besley et al., 2022; Muralidharan, 2024). While much of this literature has focused on incentives, selection, monitoring, and political accountability, we instead focus on organizational norms and shared expectations within bureaucracies. A growing body of work studies how information can improve monitoring and accountability by providing performance information about individual bureaucrats to principals (Dhaliwal and Hanna, 2017; Raffer, 2022; Muralidharan et al., 2021; Debnath et al., 2023; Callen et al., 2020; Dal Bó et al., 2021; Mattsson, 2022).⁶ We instead show that making aggregated information about bureaucratic performance public within the organization and subject to collective discussion can shift beliefs about organizational functioning and improve performance. Within the literature on audits and accountability, providing audit information publicly has been shown to discipline politicians through electoral or legal channels (Ferraz and Finan, 2008; Bobonis et al., 2016; Avis et al., 2018). We instead highlight a distinct internal mechanism: when municipality-level audit information is shared publicly within the bureaucracy and discussed collectively, it can reshape shared expectations and, in turn, improve bureaucratic performance. More broadly, our findings contribute to a literature emphasizing organizational design as a determinant of state performance (Myerson, 1995; Olken, 2010; Burgess et al., 2012; Deserranno et al., 2019; Bandiera et al., 2020; Garfias and Sellars, 2021; Vannutelli, 2022; Mastrococco and Teso, 2023; Fenske et al., 2023; Angelucci et al.,

⁶Within the literature on state capacity, several studies have focused on bureaucratic accountability and the role of internal organizational frictions in public sector performance (Bandiera et al., 2009; Becker and Stigler, 1974; Besley and McLaren, 1993; La Porta et al., 1999; Niehaus and Sukhtankar, 2013; Corbacho et al., 2016; Mahadevan, 2024; Dhillon et al., 2017).

2024) by showing how the internal organization of information and discussion within bureaucracies can shape state performance.

Our findings also highlight the social dimension of state capacity and organizations (Aman-Rana and Minaudier, 2026; Aman-Rana et al., 2025; Overbeck and Lungu, 2025; Henn, 2020; Ashraf and Bandiera, 2018), including the role of organizational norms, informal relational arrangements, and repeated interactions within organizations (Baker et al., 2002; MacLeod, 2007; Levin, 2003; Macchiavello and Morjaria, 2022; Gibbons and Henderson, 2012; Blader et al., 2020; Fahn et al., 2023). While work linking information flows to trust finds a positive relationship (Fisman and Khanna, 1999), and trust and cohesion are typically viewed as supporting organizational performance (Brown et al., 2015; Keefer and Vlaicu, 2024; Nguyen, 2026), our findings suggest that in weakly institutionalized settings trust and shared expectations may also sustain low-effort equilibria, consistent with theoretical work by Tirole (1986) and Milgrom and Roberts (1988). Deliberation can disrupt these relational arrangements and improve performance.⁷

Our paper relates to the broader literature on institutions (North, 1990; Ostrom, 1990; Acemoglu et al., 2001; Dell, 2010; Dell et al., 2018; Wantchékon and García-Ponce, 2013; Nunn, 2008; Nunn and Wantchekon, 2011; Khan, 2018). While much of this work emphasizes the persistence of historical institutions and the role of critical junctures in overcoming path dependence, we show that governments can implement what we term “institutional edits”: marginal, low-cost modifications to existing rules of the game that generate meaningful performance gains. We provide experimental evidence that collective deliberation is one such tool, shifting shared expectations and, in turn, the micro-institutions that govern bureaucratic behavior. In this sense, our paper contributes to a growing literature that studies institutions as the primary object of interest using field experiments (Callen et al., 2024).

The paper also contributes to the literature on public deliberation. Deliberative institutions are widely viewed as foundational to democratic governance (Bidwell et al., 2020; Mansuri and Rao, 2013; Mendelberg, 2002; Parkinson and Mansbridge, 2012; Hadfield and Macedo, 2012). While existing work focuses on how deliberation shapes politician-voter relationships and political outcomes (Fujiwara and Wantchekon, 2013; López-Moctezuma et al., 2022; Wantchekon and Guardado, Forthcoming; Majumdar and Fatemipour, 2024; Collier and Vicente, 2014), we provide evidence that introducing the institution of deliberation within a public sector bureaucracy can improve performance.

⁷On the downsides of social capital more generally, see (Basu, 1986; Dasgupta, 2005; Bloch et al., 2007; Alesina and Giuliano, 2014; Banfield, 1958; Gambetta, 1993; Portes and Landolt, 1996).

2 Conceptual Framework

This section provides a simple framework to interpret the empirical results.

Consider a bureaucracy in which a principal (manager) sets monitoring intensity based on her beliefs about average effort in the organization, and workers choose effort based on both the monitoring environment and their beliefs about how peers behave. Each worker receives a fixed wage unrelated to effort, bears a cost of exerting effort, faces an expected penalty for shirking that rises with monitoring intensity, and derives utility from informal relational arrangements sustained through repeated interaction and interpersonal trust. Following [Macchiavello and Morjaria \(2022\)](#), we interpret interpersonal trust as reflecting the strength of informal organizational relationships within the bureaucracy.

In environments where information about organizational performance is not collectively aggregated, shared, or discussed, organizational failures may remain fragmented and privately observed rather than collectively recognized within the bureaucracy. These conditions can sustain high levels of interpersonal trust and stable relational arrangements within the organization. In settings with weak formal incentives and limited sanctions, such arrangements may weaken individual accountability by fostering informal tolerance of low effort, non-compliance, or weak enforcement of rules. Workers may come to rely on stable workplace relationships and shared expectations that organizational failures will not trigger substantial scrutiny or punishment. Managers, in turn, may rely on stable relational arrangements and informal understandings in place of intensive monitoring or formal oversight, limiting individual accountability within the organization.

Deliberation introduces an organization-wide information shock by making performance-relevant information publicly available and subject to collective discussion. Information about audit failures, corruption, citizen dissatisfaction, and organizational weaknesses becomes collectively visible within the bureaucracy. As managers and workers update negatively about organizational performance and peer behavior, existing relational arrangements weaken and organizational expectations shift.

As managers become more pessimistic about organizational functioning, monitoring intensifies and the expected penalty for shirking rises. At the same time, workers become less confident that low effort and weak rule enforcement will remain informally tolerated within the organization. Organizational failures become harder to dismiss as isolated incidents, increasing perceived individual accountability and weakening reliance on existing relational arrangements. Effort therefore rises even without changes in formal incentives.

The same process erodes trust within the organization by weakening the relational ar-

rangements and shared expectations that previously sustained low accountability. Deliberation may therefore simultaneously improve bureaucratic performance while reducing interpersonal trust.

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 including civil registration, infrastructure, water, and sanitation.

The administrative apparatus operates within a hierarchical structure led by the General Secretary (GS), who coordinates activities 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 many of whom interact directly with citizens in the course of service delivery.

Mayors retain authority over aspects of municipal administration, including task assignments, the allocation of personnel across departments, and the recruitment of some contractual and support staff. However, formal hiring, promotion, and inter-municipality transfers for many bureaucrats remain governed by the central civil service. Municipal administrations therefore operate under decentralized political authority while remaining embedded within a centrally managed civil service system in which local control over personnel decisions and performance incentives is limited (Riedl and Dickovick, 2014; Bedasso, 2021b). Bureaucrats operate under fixed wages, although career concerns and advancement within the civil service may still shape behavior, as in other bureaucracies.

Municipal audits. The National Commission for Local Finance (CoNaFiL), in collaboration with the Office of the Auditor General, conducts yearly audits across all municipalities. 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).⁸ Auditors evaluate the management of municipal funds, focusing on record-keeping, transparency in public procurement, compliance with procurement regulations, and the quality of administrative management. Details of the criteria used are provided in Appendix Sub-section D.4. Auditors also assess broader administrative processes, including adherence to budget adoption rules and financial reporting requirements by the elected council.

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-linked allocations in subsequent FADeC transfers.⁹ Municipalities that perform well receive larger 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 support administrative improvements. In Benin, CoNaFiL plays a central role in this process and is widely regarded as an important oversight institution with substantial technical capacity and authority (Riedl and Dickovick, 2014; Behanzin et al., 2018; Bedasso, 2021a). Consistent with this, the IMF’s Public Investment Management Assessment (PIMA) describes Benin’s regulatory framework for public investment as relatively strong by regional standards (Imbert et al., 2018).

However, despite the availability of detailed audit information and recommendations, municipalities lack standardized or institutionalized forums for collectively interpreting audit findings or coordinating responses across different levels of the bureaucracy. While some municipalities establish ad hoc committees or internal processes to follow up on audit recommendations (*Decentralisation and Communal Development Support Programme, 2021; Bedasso, 2021a*), these practices are uneven and not embedded within a broader

⁸To support municipalities in meeting their development needs, the central government introduced the Support Fund for the Development of Municipalities (FADeC) in 2008 (Decree No. 2008-276).

⁹Transfers 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. FADeC investment grants constitute a major source of financing for local public goods and consist of two components: an earmarked grant targeting specific policy areas (e.g., health and education), and a formula-based grant allocated using socio-economic indicators (e.g., population and poverty levels) together with indicators of prior administrative and financial performance derived from audit reports.

deliberative process. As a result, audit information may not translate into coordinated administrative action.

More broadly, despite the availability of audit information and formal oversight structures, municipal performance and accountability remain uneven. Existing evidence points to persistent weaknesses in transparency and procurement oversight, including limited enforcement of rules governing access to public information and public procurement procedures (Imbert et al., 2018; Ch et al., 2019; Canen et al., 2023). Studies also document the influence of private vendors in local procurement processes and broader concerns about state capture (Ch et al., 2019; Canen et al., 2023). At the same time, municipalities face substantial coordination and information frictions across layers of the bureaucracy, contributing to low budget execution rates. Between 2015–2018, municipal budget execution 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 for a map of Benin and the treatment assignment).

Municipalities assigned to the deliberation intervention convened deliberation sessions centered on administrative performance and service delivery. The intervention was designed to create a structured forum in which objective, performance-relevant information about the bureaucracy as a whole could be made public within the organization and discussed collectively. *All* bureaucrats within a municipality were invited to participate, although attendance was not mandatory. To ensure comparability across sites, the deliberation environment was standardized: bureaucrats retained control over the agenda, rather than politicians or citizens, and discussions were restricted to municipal administrative performance and overall service delivery.

To allow for candid reflection on the information presented, mayors were not invited or allowed to be a part of the sessions. However, 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 still 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 7).

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 7).

Figure I describes the timeline of the experiment. In August 2016, we compiled the information presented during the deliberation sessions. The three deliberation meetings, held in September, October, and November 2016 respectively, focused on municipal audit reports, survey responses from bureaucrats, and survey responses from citizens. In December 2016, the endline survey of bureaucrats was carried out. We describe each of these steps below.

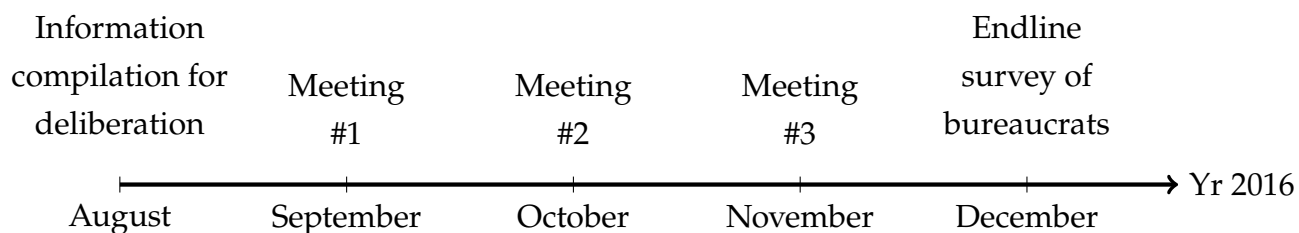


Figure I: Experiment timeline

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%.¹⁰ 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.

¹⁰Among 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 aggregate performance information for deliberation sessions. The audit reports, which were already produced at the municipality level through CoNaFiL's existing audit system, were presented alongside citizen survey results aggregated at the municipality level. In contrast, bureaucrat survey responses were aggregated further and presented only at the level of the full experimental sample rather than at the level of individual municipalities or officials. Bureaucrats were informed of this at the time of data collection in order to mitigate concerns about social desirability and encourage candid responses.

The presentation materials consisted of slide-based summaries of key indicators and findings, rather than selective discussions of failures or misconduct. 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). Across all meetings, the emphasis was on presenting factual and organization-wide information rather than individualized assessments or prescriptive recommendations. Examples of slides from the deliberation meetings are provided in Appendix C.

For the survey of bureaucrats, a representative survey of 30 bureaucrats was drawn from each municipality.¹¹ This data included questions on their knowledge of rules and regulations, municipal administration and ongoing projects, perceptions of the administration's role in improving education and health services, markets, road maintenance, and water and sanitation provision, as well as interactions with colleagues within the bureaucracy.

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, demand for improved public services, perceptions of local governance and corruption, trust in local institutions, and satisfaction with municipal performance. Following the presentation of performance information, 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'

¹¹In five municipalities we were not able to survey 30 bureaucrats. These include: Gogounou (26), Houeyogbe (24), and Pehunco (29) from the control municipalities and Grand-Popo (25) and Toucountouna (19) from the treatment municipalities.

recommendations. On average, 31 participants attended, of whom 22 were bureaucrats (see Table 7). The General Secretary was present in all municipalities except one. Figure B.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 B.3 and B.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 7; Figure B.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 B.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 B.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 7; Figure B.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 B.3 and B.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.

3.3 Implementation and Internal Validity

Balance. Figure 2 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.¹²

Figure 3 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.

Compliance. 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.¹³ Given this and the fact that 58% of bureaucrats in treated municipalities attended at least one of the three meetings, all survey-based analyses report intention-to-treat estimates. Owing to perfect compliance at the municipality level, this distinction is not relevant for our main results on performance based on the municipal audit panel data described below.

Hawthorne effects. While our main analysis does not rely on data collected during the deliberation sessions, we nevertheless took steps to minimize observer-induced distortions in behavior, which may have been particularly relevant in treated municipalities where deliberation took place. Discussions were therefore not recorded, allowing interactions to proceed as naturally as possible in the field (Levitt and List, 2007; Spradley,

¹²Only 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.

¹³Three 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.

1980). Enumerators attended meetings unobtrusively, only taking attendance and details before the start of the sessions, and completed observation protocols after sessions had concluded. Participants were also informed that there would be no real-time monitoring of discussions.

Social desirability bias and experimenter demand. A potential concern is that deliberation changed how bureaucrats reported their beliefs and trust rather than the beliefs themselves. Several features of the design mitigate this concern. First, the endline survey was conducted approximately one month after the conclusion of the intervention, reducing the likelihood that responses simply reflect immediate reactive behavior during the meetings. Second, bureaucrats were assured that responses would remain confidential and would only be reported in aggregate form. To further reduce reporting concerns, respondents completed survey forms individually rather than reporting answers publicly to enumerators. Finally, the direction of the estimated effects is not obviously consistent with simple reporting incentives or experimenter demand effects. Deliberation leads bureaucrats to report lower trust, weaker rule enforcement, and higher perceived corruption within their own municipality, responses that potentially reflect negatively on the administration itself rather than presenting it in a more favorable light.

4 Data

Our analysis draws on two main data sources. The first is the municipal audit scores reported by CoNaFiL auditors, which we use to measure municipal performance. The second is an endline survey of bureaucrats conducted in December 2016, one month after the intervention, which we use to examine mechanisms.

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. Appendix D provides details for each evaluated component of performance along with our construction of the performance measure.

The record keeping component focuses on municipal documentation, which is vital for accountability. Specific metrics include 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 in public procurement component assesses compliance with procurement regulations, the functioning of procurement oversight bodies, the completeness of published reports, and the traceability of fund transfers.

Finally, the overall administrative management component focuses on the General Secretary, the highest-ranking bureaucrat in the municipality and the chief coordinator of local administrative services. This component measures internal administrative capacity along three operational dimensions. First, auditors verify whether the General Secretary holds regular meetings with department heads, with maximum points awarded only if signed meeting minutes are available for at least one meeting per month. Second, auditors review the content of these minutes to assess whether the General Secretary monitors the implementation of previously assigned tasks and issues new operational directives. Third, the component captures internal oversight by verifying whether the General Secretary formally reviews administrative workflows, including financial vouchers, land transactions, and incoming administrative files, as evidenced by physical initials on these 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. Figure B.5 plots baseline municipal audit-based performance averaged over the pre-treatment period (2014–2016) separately for treated and control municipalities. The figure shows substantial variation in baseline performance across municipalities, with broadly comparable distributions between the two groups prior to the intervention. Figures B.6 to B.8 present baseline averages for the three underlying audit components, record keeping, transparency in public procurement, and overall administrative management, showing substantial cross municipality variation but broadly comparable distributions across treated and control municipalities prior to the intervention.

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 ($p = 0.01$). This pattern is consistent with the audit-based measure capturing meaningful variation in administrative perfor-

mance and service delivery. This pattern suggests that the audit-based measure captures meaningful variation in administrative performance and service delivery. It is also consistent with existing evidence documenting institutional 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 ensure that the sample was representative of the municipal administration, we randomly selected approximately 30 bureaucrats from the personnel roster of each municipality, irrespective of participation in the deliberation meetings. The final sample consists of 595 bureaucrats, including 299 from treated municipalities and 296 from control municipalities. Appendix Figure B.9 reports the number of bureaucrats surveyed in each municipality, and Appendix Section E provides details on the survey questions and the construction of the indices described below.

The survey collected information on bureaucrats' beliefs, interpersonal trust, and internal meeting practices within the administration. We measure beliefs about rule enforcement using responses to questions on whether respondents believe that procedures governing the acquisition of office equipment, public procurement, and staff recruitment are followed. Beliefs about corruption are measured using questions that ask whether respondents think the municipal council, city hall private vendors, and their colleagues engage in corruption. Following Anderson (2008a), these measures are combined into a standardized index, which we refer to as the belief about rule of law index, with the sign of the corruption measures reversed so that higher values correspond to stronger perceived rule enforcement and lower perceived corruption.

Combining multiple related outcomes into a single summary index improves statistical power when treatment effects move in the same direction within a domain (O'Brien, 1984; Anderson, 2008a; Kling et al., 2007) and reduces the risk of false positives arising from multiple hypothesis testing (Anderson, 2008a).

Trust is measured using responses to questions asking whether respondents trust the municipal council, their office colleagues, colleagues within the same department, and other staff in the town hall.¹⁴ As with the belief measures, these responses are aggregated

¹⁴Anderson et al. (2004) show that survey measures of trust are significant predictors of contribution levels in a canonical public goods experiment. While Glaeser et al. (2000) critique broad attitudinal measures of trust, such as agreement with the statement "most people can be trusted," they find that more specific survey-based trust measures are predictive of behavior in experimental trust games.

into a standardized trust index, where higher values correspond to greater trust.

While we do not directly observe the specific checks undertaken by managers, we use reported meeting practices within the administration as proxies for monitoring and accountability related interactions. Respondents report both the frequency of internal meetings and whether these meetings involved discussions of the City Hall FADeC audit report and corruption in municipal affairs, the municipality’s financial and taxation issues, challenges and perspectives facing city hall, the functioning of local government, the performance of the municipal administration and council, and good governance within city hall.

5 Effects on Municipal Performance

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

5.1 Estimation Strategy

Using municipality-year panel data on audit scores, we employ a difference-in-differences (DiD) strategy that compares the evolution of performance in treated municipalities relative to control municipalities before and after the intervention. Specifically, for municipality m in year t , we estimate:

$$y_{mt} = \alpha_m + \alpha_t + \beta(T_m \times \text{Post}_t) + \varepsilon_{mt} \quad (1)$$

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. α_m and α_t denote municipality and year fixed effects, respectively, absorbing time-invariant municipal characteristics and common temporal shocks to all municipalities, respectively. 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 randomization mechanism and provides exact finite-sample inference under the sharp null of no treatment effect, offering a reliable alternative to asymptotic cluster-robust inference when the number of clusters is small (Fisher, 1935; MacKinnon and Webb, 2020; Young, 2019).

5.2 Results

Table 2 reports the results. We find that when performance relevant information about the bureaucracy as a whole is made public and discussed collectively municipal performance improves.

Results in Column (2) show that 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). The effects are both economically and statically significant, representing 8.7% of the control mean (conventional p -value <0.05 , randomization inference p -value <0.01). These findings highlight the potential for low-cost institutional changes to improve performance, even in the absence of changes to formal bureaucratic incentives.

Appendix Figure B.10 presents treatment effects separately for each treated municipality. The improvements are not driven by a single municipality, but instead reflect gains distributed 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. Appendix Table A.3 presents the results. 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 initial levels of performance. This suggests that such interventions may represent a low-hanging fruit for improving bureaucratic performance in weaker-performing administrations.

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 helps us understand which dimensions of administrative performance are most affected.

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; randomization inference p -value = 0.22), reflecting improvements in accounting and documentation practices, including the maintenance of accounting records, municipal archives, and financial execution statements. Procurement transparency increases by 6 percentage points (7.4% of the control mean; randomization inference p -value = 0.18), reflecting greater compliance with procurement regulations and improved traceability of public expenditures and fund transfers. The largest point estimate is observed for overall administrative management, which increases by 14 percentage points (17.2% of the control mean; randomization inference p -value = 0.22), suggesting improvements in internal bureaucratic coordination and oversight, including more regular monitoring of administrative activities and implementation of operational directives by the General Secretary.

Finally, we also examine whether improvements in administrative performance translate into changes in citizens' experiences with local government using Afrobarometer survey data, a cross-country survey of citizen attitudes and experiences in African countries.¹⁵ Because Afrobarometer data are not available annually and do not cover all municipalities in our sample, these estimates are based on a shorter panel and a smaller set of municipalities than the audit data. We find positive but imprecise effects overall, with larger improvements among municipalities with lower baseline performance, mirroring the patterns observed in the audit-based measures. Together with the strong baseline correlation between audit-based performance and citizen experiences (Table A.2), these findings are consistent with improvements in administrative performance translating into better service delivery outcomes. Full results and details on the construction of the citizen experience index are reported in Appendix Section F.

Persistence of effects. We examine the temporal dynamics of the treatment effect to assess whether the observed improvements in performance reflect short-term salience

¹⁵Afrobarometer is designed to be nationally representative and is not intended to provide representative estimates at the municipality level. We therefore interpret these estimates as suggestive evidence on downstream citizen experiences.

or more persistent changes in behavior. Figure B.12 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), differences between treated and control municipalities are economically and statistically 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 owing to limited statistical power, the absence of pre-trends and the persistence of the post-treatment coefficients suggest that the observed gains are unlikely to reflect only short-term salience or transitory responses to the intervention.

6 Effects on the Organizational Environment

Guided by the conceptual framework, we examine how deliberation reshapes the organizational environment within the bureaucracy. We use this term to refer to the beliefs, accountability relationships, managerial practices, and informal relational arrangements that shape bureaucratic behavior within the organization. The framework suggests that deliberation can improve performance by making organizational failures collectively visible, weakening the relational arrangements that sustain low accountability, and changing expectations about accountability and rule enforcement within the organization. We therefore examine three sets of outcomes: bureaucrats’ beliefs about corruption and rule enforcement, internal meetings and accountability-related managerial practices, and interpersonal trust within the bureaucracy.

6.1 Bureaucratic beliefs about rule of law

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

$$y_{im} = \gamma + \phi T_m + u_{im}. \quad (2)$$

The outcome y_{im} includes measures of bureaucrats’ beliefs about corruption among councilors, other bureaucrats, and city hall private vendors, as well as beliefs about the enforcement of rules governing the acquisition of office equipment, public procurement, and recruitment.

As described in the data section, we construct a summary index of beliefs about the rule of law, such that coefficients can be interpreted in standard deviation units relative to the control group. 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. Constructing a summary index improves statistical power and addresses concerns related to multiple hypothesis testing (O'Brien, 1984; Anderson, 2008a; Kline and Santos, 2012).

Table 3 reports the results. Deliberation leads to a large and statistically significant negative shift in beliefs about the rule of law. The aggregate index decreases by 0.40 standard deviations in treated municipalities relative to the control group (p -value = 0.006; randomization inference p -value = 0.006), indicating a substantial decline in perceived rule compliance and integrity within the bureaucracy. These patterns suggest that deliberation made organizational failures and weaknesses harder to dismiss as isolated incidents within the bureaucracy.

The sub-components show a similar 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 perceive corruption among vendors relative to the control group (p -value < 0.01; randomization inference p -value < 0.01). Estimated effects for perceived 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 -value < 0.10; randomization inference p -value < 0.10). Estimated effects for equipment acquisition procedures and hiring rules are similar in magnitude, though not statistically significant at conventional levels.

Attrition A potential concern with these estimates is differential attrition between treatment and control municipalities. Table A.5 shows that assignment to treatment does not systematically predict missingness for the aggregate beliefs index or most sub-components, with the exception of a marginally significant effect for beliefs about corruption among other bureaucrats ($p = 0.096$). We follow de Mel et al. (2008) and construct Lee (2009) bounds to formally assess robustness. Table A.6 shows that the Lee bounds for the aggregate beliefs index are tight and entirely negative, ranging from -0.43 to -0.35 standard deviations, with both bounds significant at the 1 percent level. This indicates that the

estimated decline in beliefs about the rule of law is robust to any plausible pattern of selection into the sample.

Quantile estimates. We also examine whether the treatment effects are concentrated in particular parts of the belief distribution or observed more broadly across it. Figure 5 presents the results graphically, and Appendix Table A.9 reports the corresponding estimates along with randomization inference p -values. The effects are not confined to the mean but are present across the belief distribution, with negative effects from the 10th to the 70th percentile (randomization inference p -value < 0.05 throughout). This suggests that the negative updating in beliefs is broad-based and extends across multiple parts of the belief distribution.

Figure 6 further shows that bureaucrats in treated municipalities who did not attend the deliberation meetings also exhibit more negative beliefs about rule enforcement and corruption relative to bureaucrats in control municipalities. While attendance was voluntary and these comparisons should not be interpreted causally, the figure suggests that belief changes may not have been mechanically confined to direct participants in the deliberation sessions.

Convergence in beliefs within municipalities. We next examine whether deliberation affected the dispersion of beliefs within municipalities. If deliberation generates coordinated updating and more aligned interpretations of performance-relevant information, we would expect beliefs to become less dispersed in treated municipalities, particularly among bureaucrats who attended the meetings. For this analysis, we collapse the data to the municipality level and compute the standard deviation of beliefs across bureaucrats within each municipality.

Table A.10 reports municipality-level standard deviations of beliefs, computed separately for bureaucrats who attended the deliberation meetings and those who did not. Columns (1)–(2) report results for the standard deviation of overall beliefs, columns (3)–(4) for beliefs about corruption, and columns (5)–(6) for beliefs about the enforcement of procurement and hiring rules. Lower values for treated municipalities indicate that beliefs become less dispersed following the deliberation exercise.

The results in Table A.10 support this prediction. Dispersion declines in treated municipalities, with the strongest effects among bureaucrats who attended the meetings, particularly for beliefs about corruption. While dispersion also falls among non-attendees, these effects are smaller in magnitude and statistically insignificant.

Taken together, these findings show that deliberation not only shifts beliefs on average and across the belief distribution, but also reduces their dispersion within municipalities. Among attendees, beliefs become significantly more aligned, particularly regarding corruption, suggesting that deliberation generated more similar interpretations of organizational performance within the bureaucracy.

6.2 Managers' beliefs, internal meetings and accountability

The conceptual framework suggests that the deliberation intervention acts as an organization-wide information shock, leading managers to update their beliefs about organizational performance and accountability within the bureaucracy and potentially shift managerial attention and accountability-related practices. In this subsection, we examine this further.

Given the institutional setting, we classify managers as individuals serving as General Secretary or those holding director- or chief-level positions. A natural limitation is that managers constitute a substantially smaller share of the bureaucracy, reducing statistical power for detecting effects within this subgroup.

We first examine managerial exposure to the deliberation meetings. Appendix Figure B.11 regresses bureaucrats' socioeconomic, demographic, and professional characteristics on attendance at the deliberation meetings. The figure shows that while there are no systematic differences across most socioeconomic and demographic characteristics, managers were substantially more likely to attend the deliberation meetings than non-managerial bureaucrats. In addition, the General Secretary, the highest-ranking bureaucrat in the municipality, was present in all but one municipality across the deliberation sessions (Table 7). This is notable given that the largest improvements in audit-based performance are observed for overall administrative management, a component specifically centered on internal administrative coordination and oversight (Table A.4).

Consistent with these patterns, Figures B.14–B.15 show that managers exhibit similar negative updates in beliefs about the rule of law as workers. The estimated effects for managers are economically and statistically significant for the aggregate beliefs index. Together, these patterns suggest that managers were exposed to the deliberation process and negatively updated their beliefs in response to it.

We next examine whether these changes in beliefs were accompanied by changes in accountability-related interactions within the administration. We use bureaucrats' reported meeting attendance and meeting content as proxies for managerial attention and accountability-related practices within the bureaucracy. This focus on meetings reflects the

institutional setting we study, where internal oversight is often exercised through regular administrative meetings. Consistent with this, whether such meetings are regularly held forms a core component of the auditors' performance evaluation criteria, as discussed in data Subsection 4.1.

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 in the treated municipalities relative to the control. When deliberation meetings are excluded, meeting attendance is lower in treated municipalities. Figure 10 shows that this decline is driven primarily by managers rather than workers. Taken together, these patterns suggest that deliberation substituted for routine managerial oversight interactions rather than increasing the overall quantity of oversight within the bureaucracy.

We next examine whether deliberation affected the content of meetings or what got discussed in the meetings that did take place. While the overall number of meetings did not increase, Figure 11 shows that their content shifted. In treated municipalities, discussions of audit reports and corruption increase relative to the control group, with no corresponding increase in discussions of taxes and finances, broader administrative challenges, or general governance issues (Figure 12). This pattern suggests that deliberation shifted managerial attention toward accountability-related concerns and organizational failures highlighted by the audit information rather than generating a broad-based expansion in oversight activities.

Taken together, these findings suggest that deliberation reshaped the focus of managerial interactions within the bureaucracy toward accountability concerns and organizational failures highlighted by the audit information rather than simply increasing the quantity of oversight activities. We next examine whether the broader changes in beliefs and managerial interactions altered interpersonal trust and relational arrangements within the bureaucracy.

6.3 Interpersonal trust

The preceding results show that deliberation led bureaucrats and managers to negatively update their beliefs about rule enforcement and corruption and shifted managerial interactions toward accountability-related concerns. We next examine whether these broader changes in beliefs and accountability relationships altered interpersonal trust within the

bureaucracy.

We interpret interpersonal trust as capturing the strength and stability of informal organizational relationships within the bureaucracy (Macchiavello and Morjaria, 2022). In environments where formal incentives and monitoring are weak, bureaucratic performance may depend partly on relational arrangements and informal understandings among bureaucrats regarding effort, rule enforcement, and accountability. Such arrangements can facilitate coordination and cooperation, but they can also sustain tolerance of low effort or weak enforcement of rules.

Table 4 reports the effects of deliberation on bureaucratic trust. The aggregate trust index declines by 0.35 standard deviations in treated municipalities relative to the control group (p -value = 0.039; randomization inference p -value = 0.060), indicating a substantial reduction in interpersonal trust following deliberation.

The decline is observed across multiple relational categories. Trust in councilors falls by 0.38 standard deviations in treated municipalities relative to the control group (p -value = 0.031; randomization inference p -value = 0.020), representing the largest effect. Trust in departmental colleagues declines by 0.28 standard deviations (p -value = 0.064; randomization inference p -value = 0.060). Estimated effects for trust in office colleagues and city hall colleagues more broadly are similar in magnitude and direction, although less precisely estimated.

Attrition We test for differential attrition in responses to the trust measures across treated and control municipalities. For trust, Table A.7 shows no differential attrition on any sub-component, and all 595 endline respondents provide a non-missing value for the aggregate trust index. The estimated decline of 0.35 standard deviations is therefore not subject to attrition concerns. We, however, report Lee bounds for the trust sub-components in Table A.8 for completeness.

Quantile estimates. Consistent with the results for beliefs, the effects on trust are not confined to the mean. Figure 8 presents the quantile treatment effects graphically, while Appendix Table A.11 reports the corresponding estimates and randomization inference p -values. The results show negative effects across much of the trust distribution, particularly between the 20th and 80th percentiles. This suggests that the decline in interpersonal trust reflects a broad-based shift within the bureaucracy rather than changes concentrated among a narrow subset of bureaucrats.

Figure 9 further indicates that declines in trust extend beyond direct participants in the deliberation meetings. Bureaucrats in treated municipalities who did not attend the meetings also exhibit lower trust relative to bureaucrats in control municipalities, suggesting that the intervention altered relational dynamics more broadly within the organization.

Managers. Managers exhibit similar directional declines in trust, although the estimates are more muted and less precisely estimated (Figures B.16–B.17). Overall, we cannot reject that the effects on managers are statistically similar to those for workers. More broadly, the decline in trust among workers is observed consistently across multiple categories of actors within the bureaucracy, including councilors, departmental colleagues, and city hall colleagues. This pattern suggests that deliberation changed how bureaucrats viewed and related to others throughout the organization, rather than simply reducing trust in one particular group.

The joint pattern of results—improvements in performance, negative updating of beliefs about rule enforcement and corruption, changes in managerial attention toward accountability concerns, and declines in interpersonal trust—is notable because trust and social cohesion are often viewed as supporting organizational performance (Brown et al., 2015; Keefer and Vlaicu, 2024; Nguyen, 2026). One interpretation of these findings, however, is that in weakly institutionalized settings interpersonal trust and informal relational arrangements may instead sustain low-effort organizational equilibria. This interpretation relates to a broader literature emphasizing the downsides of relationships and interpersonal trust and how it can also 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; Tirole, 1986; Milgrom and Roberts, 1988).

Taken together, these findings suggest that deliberation improved performance not by strengthening interpersonal cohesion within the bureaucracy, but by disrupting relational arrangements and shared expectations that previously sustained weak accountability and low effort within the organization.¹⁶

¹⁶As discussed in the introduction, in this setting there is a negative relationship between trust and performance at baseline using municipal audit-based performance and a citizen experience index constructed from the Afrobarometer (Figures B.13a and B.13b).

6.4 Discussion

This paper shows that low-cost institutional changes within bureaucracies can generate meaningful improvements in performance by reshaping organizational beliefs and expectations. Deliberation changed how bureaucrats interpreted organizational functioning, reducing trust and shifting attention toward accountability-related concerns.

A natural question is why such practices do not emerge organically within bureaucracies. One possibility is that organizational environments characterized by high interpersonal trust and stable relational arrangements may resist processes that publicly reveal collective failures or weaken existing norms. Even when such changes improve performance, initiating them may be individually costly for bureaucrats whose day-to-day interactions depend on maintaining cooperative workplace relationships. In this sense, institutional change may require coordination that existing organizational equilibria make difficult to achieve internally.

More broadly, the findings suggest that institutional reforms need not operate solely through changes in formal incentives, monitoring technologies, or personnel selection. Modifying how information is collectively interpreted within organizations may itself be an important lever for improving state performance.

7 Conclusion

We study whether changing organizational norms through deliberation can improve bureaucratic performance. In a field experiment with municipal administrations in Benin, collective deliberation over performance-relevant information improved audit-based measures of bureaucratic performance. These improvements were accompanied by shifts in beliefs about corruption and rule enforcement, declines in interpersonal trust, and changes in the content of internal discussions within the bureaucracy.

The findings suggest that bureaucratic underperformance may persist not only because of weak formal incentives, but also because of organizational environments sustained by shared expectations and informal relational arrangements. Deliberation altered these organizational environments by making performance-relevant information collectively visible and open to discussion.

More broadly, the results highlight the potential for low-cost “institutional edits” within organizations to improve state performance. In weakly institutionalized settings,

improving accountability may require disrupting organizational norms and relationships that sustain low-effort equilibria. An important direction for future work is to understand how organizations can sustain accountability while rebuilding forms of trust and cooperation that support high performance.

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.
- Acemoglu, Daron, Simon Johnson, and James A Robinson**, “The colonial origins of comparative development: An empirical investigation,” *American economic review*, 2001, 91 (5), 1369–1401.
- Agte, Patrick and Mariel Bedoya**, “The making of a public sector worker: The causal effects of temporary work assignments to poor areas,” Technical Report, Working Paper 2023.
- 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**, “Meritocracy in a Bureaucracy,” *Journal of Development Economics*, 2025, 175, 103428.
- **and Clement Minaudier**, “Spillovers in State Capacity Building: Evidence from the Digitization of Land Records in Pakistan,” *American Economic Review*, 2026.
- , – , **and Sandip Sukhtankar**, “Informal fiscal systems in developing countries,” *Journal of Development Economics*, 2025, p. 103712.
- Anderson, Lisa R, Jennifer M Mellor, and Jeffrey Milyo**, “Social capital and contributions in a public-goods experiment,” *American Economic Review*, May 2004, 94 (2), 373–376.
- 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.

- Angelucci, Charles, Simone Meraglia, and Nico Voigtländer**, “Organizing a kingdom,” Technical Report, National Bureau of Economic Research 2024.
- 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.
- Avis, Eric, Claudio Ferraz, and Frederico Finan**, “Do government audits reduce corruption? Estimating the impacts of exposing corrupt politicians,” *Journal of Political Economy*, 2018, 126 (5), 1912–1964.
- Baker, George, Robert Gibbons, and Kevin J Murphy**, “Relational Contracts and the Theory of the Firm,” *The Quarterly Journal of Economics*, 2002, 117 (1), 39–84.
- 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.
- , **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, Raghavendra 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.
- Banfield, Edward C.**, *The Moral Basis of a Backward Society*, New York: Free Press, 1958.
- Bardhan, Pranab**, “State and development: The need for a reappraisal of the current literature,” *Journal of Economic Literature*, 2016, 54 (3), 862–892.
- 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.
- Bergeron, Augustin, Pedro Bessone, John Kabeya Kabeya, Gabriel Z Tourek, and Jonathan L Weigel**, “Optimal assignment of bureaucrats: Evidence from randomly assigned tax collectors in the DRC,” Technical Report, National Bureau of Economic Research 2022.
- 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.
- **, Robin Burgess, Adnan Khan, and Guo Xu**, “Bureaucracy and development,” *Annual Review of Economics*, 2022, 14 (1), 397–424.
- Bidwell, Kelly, Katherine Casey, and Rachel Glennerster**, “Debates: Voting and expenditure responses to political communication,” *Journal of Political Economy*, 2020, 128 (8), 2880–2924.
- Blader, Steven, Claudine Gartenberg, and Andrea Prat**, “The contingent effect of management practices,” *The Review of Economic Studies*, 2020, 87 (2), 721–749.
- 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.

- 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.
- , – , **Nicholas Y Li, and Laura Schechter**, “Information technology and government decentralization: Experimental evidence from Paraguay,” *Econometrica*, 2021, 89 (2), 677–701.
- Bobonis, Gustavo J, Luis R Cámara Fuertes, and Rainer Schwabe**, “Monitoring corruptible politicians,” *American Economic Review*, 2016, 106 (8), 2371–2405.
- Brown, Sarah, Daniel Gray, Jolian McHardy, and Karl Taylor**, “Employee trust and workplace performance,” *Journal of Economic Behavior & Organization*, 2015, 116, 361–378.
- Brown, Tahir Andrabi Christina**, “Subjective versus objective incentives and employee productivity,” Technical Report, Working Paper 2022.
- 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, Jonathan L Weigel, and Noam Yuchtman**, “Experiments about institutions,” *Annual Review of Economics*, 2024, 16 (1), 105–131.
- , **Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan, and Arman Rezaee**, “Data and Policy Decisions: Experimental Evidence from Pakistan,” *Journal of Development Economics*, 2020, 146, 102523.
- , – , – , – , – , **and –**, “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), 766–99.
- de Mel, Suresh, David McKenzie, and Christopher Woodruff,** “Returns to Capital in Microenterprises: Evidence from a Field Experiment,” *The Quarterly Journal of Economics*, November 2008, 123 (4).
- 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.*
- Dell, Melissa,** “The Persistent Effects of Peru’s Mining Mita,” *Econometrica*, 2010, 78 (6), 1863–1903.
- , **Nathan Lane, and Pablo Querubin,** “The Historical State, Local Collective Action, and Economic Development in Vietnam,” *Econometrica*, 2018, 86 (6), 2083–2121.
- Deserranno, Erika, A Stefano Caria, Philipp Kastrau, and Gianmarco León-Ciliotta,** “The allocation of incentives in multilayered organizations: evidence from a community health program in Sierra Leone,” *Journal of Political Economy*, 2025, 133 (8), 2506–2562.
- , **Miri Stryjan, and Munshi Sulaiman,** “Leader selection and service delivery in community groups: Experimental evidence from uganda,” *American Economic Journal: Applied Economics*, 2019, 11 (4), 240–267.

- , **Philipp Kastrau, and Gianmarco León-Ciliotta**, “Promotions and productivity: the role of meritocracy and pay progression in the public sector,” *American Economic Review: Insights*, 2025, 7 (1), 71–89.
- 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.
- Dhillon, Amrita, Antonio Nicolò, and Fei Xu**, “Corruption, intrinsic motivation, and the love of praise,” *Journal of Public Economic Theory*, 2017, 19 (6), 1077–1098.
- Dipoppa, Gemma and Saad Gulzar**, “Administrative Incentives Impact Crop-Residue Burning and Health in South Asia,” *Working Paper*, 2023.
- 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.
- Fahn, Matthias, W Bentley MacLeod, and Gerd Muehlheusser**, “Symposium on the Economics of Relational Contracts: Past and Future Developments,” *Journal of Institutional and Theoretical Economics*, 2023, 179 (3/4), 413–440.
- Fenzia, Alessandra**, “Managers and productivity in the public sector,” *Econometrica*, 2022, 90 (3), 1063–1084.
- Fenske, James, Muhammad Haseeb, and Namrata Kala**, “How rules and compliance impact organizational outcomes: Evidence from delegation in environmental regulation,” *Technical Report, National Bureau of Economic Research* 2023.
- Ferraz, Claudio and Frederico Finan**, “Exposing corrupt politicians: the effects of Brazil’s publicly released audits on electoral outcomes,” *The Quarterly journal of economics*, 2008, 123 (2), 703–745.
- 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.
- Fisman, Raymond and Tarun Khanna**, “Is trust a historical residue? Information flows and trust levels,” *Journal of Economic Behavior & Organization*, 1999, 38 (1), 79–92.

- Fujiwara, Thomas and Leonard Wantchekon**, “Can informed public deliberation overcome clientelism? Experimental evidence from Benin,” *American Economic Journal: Applied Economics*, 2013, 5 (4), 241–255.
- Gambetta, Diego**, “*The Mafia: A ruinous rationality*,” 1993.
- Garfias, Francisco and Emily A Sellars**, “Fiscal legibility and state development: Theory and evidence from colonial Mexico,” *American Journal of Political Science*, 2021.
- Gibbons, Robert and Rebecca Henderson**, “Relational contracts and organizational capabilities,” *Organization science*, 2012, 23 (5), 1350–1364.
- Glaeser, Edward L, David I Laibson, Jose A Scheinkman, and Christine L Soutter**, “Measuring trust,” *The Quarterly Journal of Economics*, August 2000, 115 (3), 811–846.
- 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.
- Henn, Soeren J**, “Complements or substitutes? how institutional arrangements bind chiefs and the state in africa,” URL: <http://soerenhenn.com/files/HennChiefs.pdf>, 2020.
- 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.
- 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*,” *American Economic Review*, 2025, 115 (7), 2343–2375.
- Khan, Mushtaq H**, “*Political settlements and the analysis of institutions*,” *African affairs*, 2018, 117 (469), 636–655.
- Kline, Patrick and Andres Santos**, “*A score based approach to wild bootstrap inference*,” *Journal of Econometric Methods*, 2012, 1 (1), 23–41.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz**, “*Experimental Analysis of Neighborhood Effects*,” *Econometrica*, 2007, 75 (1), 83–119.
- 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 Treatment Effects*,” *Review of Economic Studies*, 2009, 76 (3), 1071–1102.
- Levin, Jonathan**, “*Relational Incentive Contracts*,” *American Economic Review*, 2003, 93 (3), 835–857.
- 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.
- Macchiavello, Rocco and Ameet Morjaria**, “*Empirical Approaches to Trust and Relational Contracts*,” *International Journal of Industrial Organization*, 2022, 84, 102849.
- 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.

- MacLeod, W Bentley**, “Reputations, relationships, and contract enforcement,” *Journal of Economic Literature*, 2007, 45 (3), 595–628.
- Mahadevan, Meera**, “The price of power: Costs of political corruption in Indian electricity,” *American Economic Review*, 2024, 114 (10), 3314–3344.
- Majumdar, Priyama and Elaheh Fatemipour**, “Government by Discussion: Deliberation, Media, and Policy-Making,” 2024. Working paper.
- Mansuri, Ghazala and Vijayendra Rao**, “Localizing development: Does participation work?,” 2013.
- Mastrococco, Nicola and Edoardo Teso**, “State capacity as an organizational problem. Evidence from the growth of the US state over 100 years,” Technical Report, National Bureau of Economic Research 2023.
- Mattsson, Martin**, “Information Systems, Service Delivery, and Corruption: Evidence From the Bangladesh Civil Service,” Working Paper, 2022.
- 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.
- Moreira, Diana and Santiago Pérez**, “Civil Service Exams and Organizational Performance: Evidence from the Pendleton Act,” *American Economic Journal: Applied Economics*, 2024, 16 (3), 250–291.
- Muralidharan, Karthik**, *Accelerating India’s Development: A State-Led Roadmap for Effective Governance*, Penguin Random House India Private Limited, 2024.
- **and Paul Niehaus**, “Experimentation at scale,” *Journal of Economic Perspectives*, November 2017, 31 (4), 103–124.
- , – , **Sandip Sukhtankar, and Jeffrey Weaver**, “Improving last-mile service delivery using phone-based monitoring,” *American Economic Journal: Applied Economics*, April 2021, 13 (2), 52–82.
- Myerson, Roger B**, “Analysis of democratic institutions: Structure, conduct and performance,” *Journal of economic perspectives*, 1995, 9 (1), 77–89.

- Nguyen, Kieu-Trang**, “Trust and Innovation Within the Firm: Evidence from Matched CEO-Firm Data,” *The Quarterly Journal of Economics*, May 2026, 141 (2). <https://doi.org/10.1093/qje/qjaf002>.
- Niehaus, Paul and Sandip Sukhtankar**, “The marginal rate of corruption in public programs: Evidence from India,” *Journal of Public Economics*, 2013, 104, 52–64.
- North, Douglass C**, *Institutions, institutional change and economic performance*, Cambridge university press, 1990.
- Nunn, Nathan**, “The Long-term Effects of Africa’s Slave Trades,” *The Quarterly Journal of Economics*, 2008, 123 (1), 139–176.
- **and Leonard Wantchekon**, “The Slave Trade and the Origins of Mistrust in Africa,” *American Economic Review*, 2011, 101 (7), 3221–3252.
- O’Brien, Peter C.**, “Procedures for Comparing Samples with Multiple Endpoints,” *Biometrics*, 1984, 40 (4), 1079–1087.
- Olken, Benjamin A.**, “Monitoring Corruption: Evidence from a Field Experiment in Indonesia,” *Journal of Political Economy*, 2007, 115 (2), 200–249.
- Olken, Benjamin A.**, “Direct democracy and local public goods: Evidence from a field experiment in Indonesia,” *American Political Science Review*, 2010, 104 (2), 243–267.
- Ostrom, Elinor**, *Governing the commons: The evolution of institutions for collective action*, Cambridge university press, 1990.
- Overbeck, Daniel and Eliya Lungu**, “Bargaining Over Taxes,” *Technical Report, CESifo Working Paper 2025*.
- Page, Lucy and Rohini Pande**, “Ending global poverty: Why money isn’t enough,” *Journal of Economic Perspectives*, 2018, 32 (4), 173–200.
- Parkinson, John and Jane Mansbridge**, *Deliberative systems: Deliberative democracy at the large scale*, Cambridge University Press, 2012.
- Pepinsky, Thomas B, Jan H Pierskalla, and Audrey Sacks**, “Bureaucracy and service delivery,” *Annual Review of Political Science*, 2017, 20, 249–268.
- 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.
- Raffler, Pia J**, “Does political oversight of the bureaucracy increase accountability? Field experimental evidence from a dominant party regime,” *American Political Science Review*, 2022, 116 (4), 1443–1459.
- 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.
- Tirole, Jean**, “Hierarchies and Bureaucracies: On the Role of Collusion in Organizations,” *Journal of Law, Economics, and Organization*, 1986, 2 (2), 181–214.
- Vannutelli, Silvia**, “From lapdogs to watchdogs: Random auditor assignment and municipal fiscal performance,” *Technical Report, National Bureau of Economic Research* 2022.
- Wantchekon, Leonard and Jenny Guardado**, “Deliberation and Ethical Voting Behavior: Evidence from a Campaign Experiment in Benin,” *American Political Science Review*, Forthcoming.
- Wantchékon, Léonard and Omar García-Ponce**, “Critical junctures: independence movements and democracy in Africa,” 2013.
- Weber, M**, *Economy and Society*, 4 ed., Tübingen: Mohr Siebeck, 1922.
- Xu, Guo**, “The Costs of Patronage: Evidence from the British Empire,” *American Economic Review*, November 2018, 108 (11), 3170–98.
- Young, Alwyn**, “Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results,” *Quarterly Journal of Economics*, 2019, 134 (2), 557–598.

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 if 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 (1)	DID Estimate (2)	Standard Error (3)	P-Value (4)	Randomization Inference P-Value (5)	Obs. (6)
Municipal Audit-Based Performance	77.835	6.8286**	2.9555	0.032	0.026	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 (1)	Standard P-Value Error (2)	P-Value (3)	RI P-Value (4)	Obs. (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. It 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 (1)	Standard Error (2)	P-Value (3)	RI P-Value (4)	Obs. (5)
Trust Index	-0.3511**	0.1585	0.039	0.042	595
Bureaucrats' Trust in					
Councilors	-0.3769**	0.1613	0.031	0.035	581
Office Colleagues	-0.2732	0.1589	0.102	0.120	538
Colleagues in Same Dept	-0.2788*	0.1415	0.064	0.073	574
City Hall Colleagues	-0.2170	0.1476	0.158	0.172	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: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

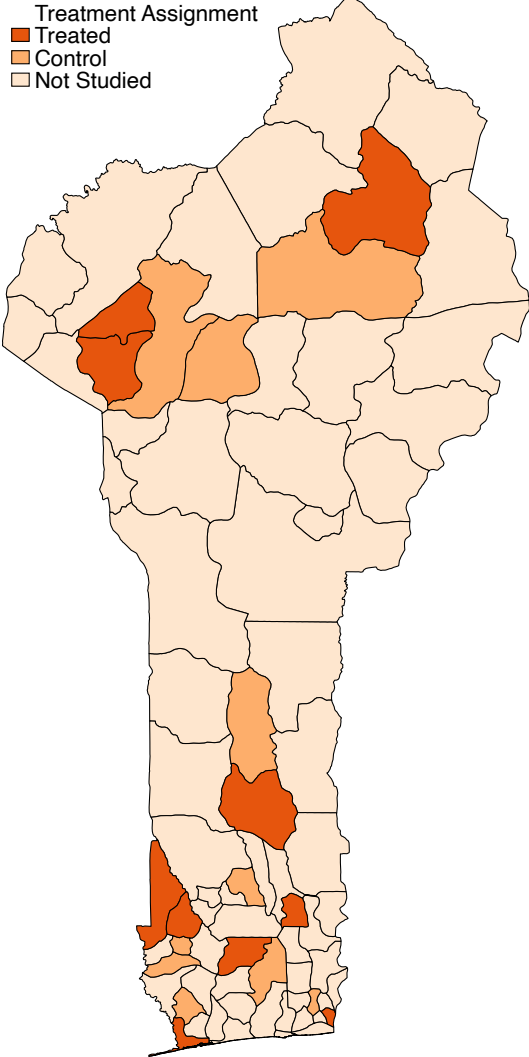
Table 5: Number of meetings held

	Number of City-hall Meetings attended in the last 3 months	
	(1)	(2)
	Overall	Excluding Deliberation
Treatment	-0.460 (0.896) {0.598}	-1.610* (0.847) {0.0790}
Control Mean	6.132 (0.625) {0.303}	6.098** (0.622) {0.0420}
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: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

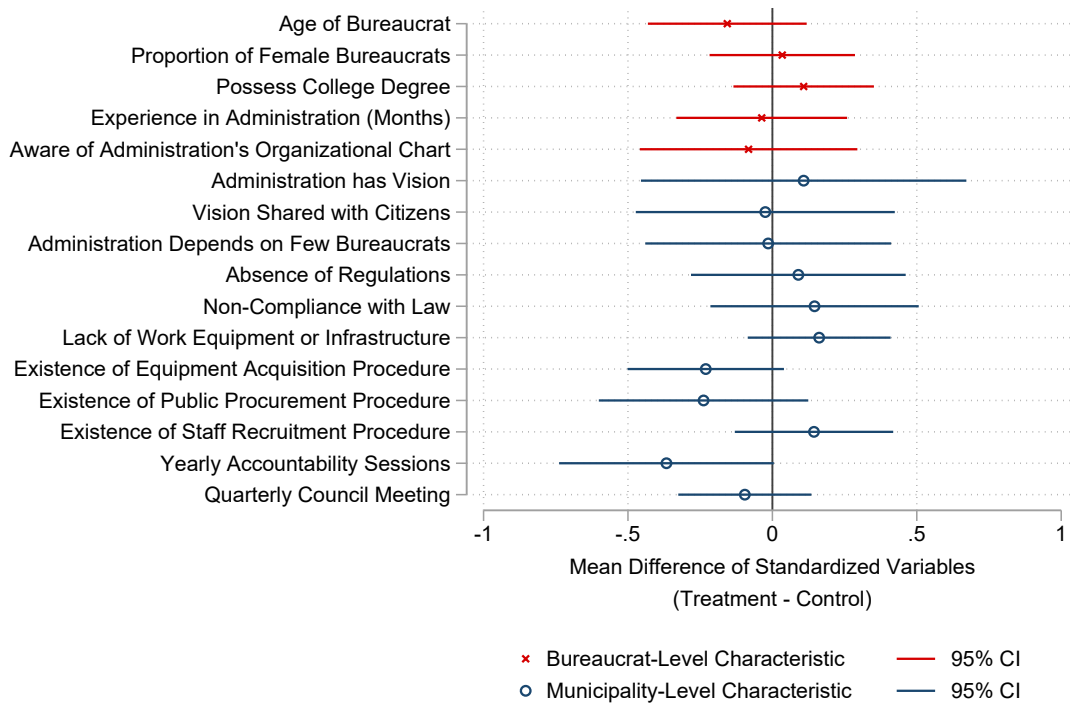
Figures

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.

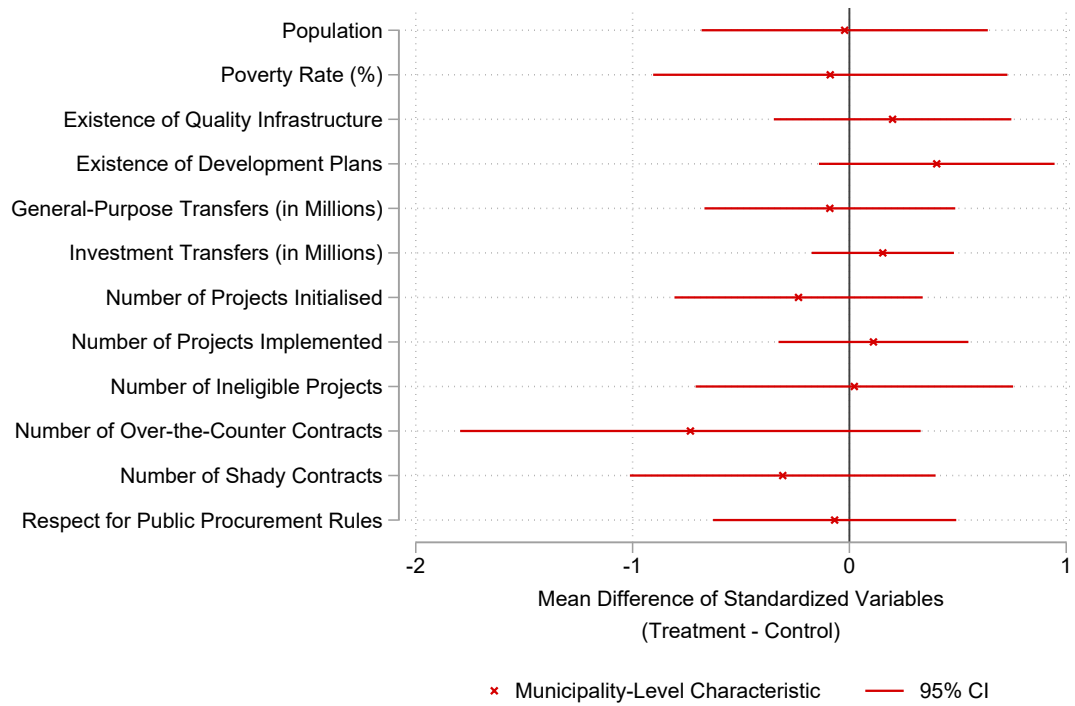
Figure 2: Balance test on characteristics of bureaucrats



Notes.

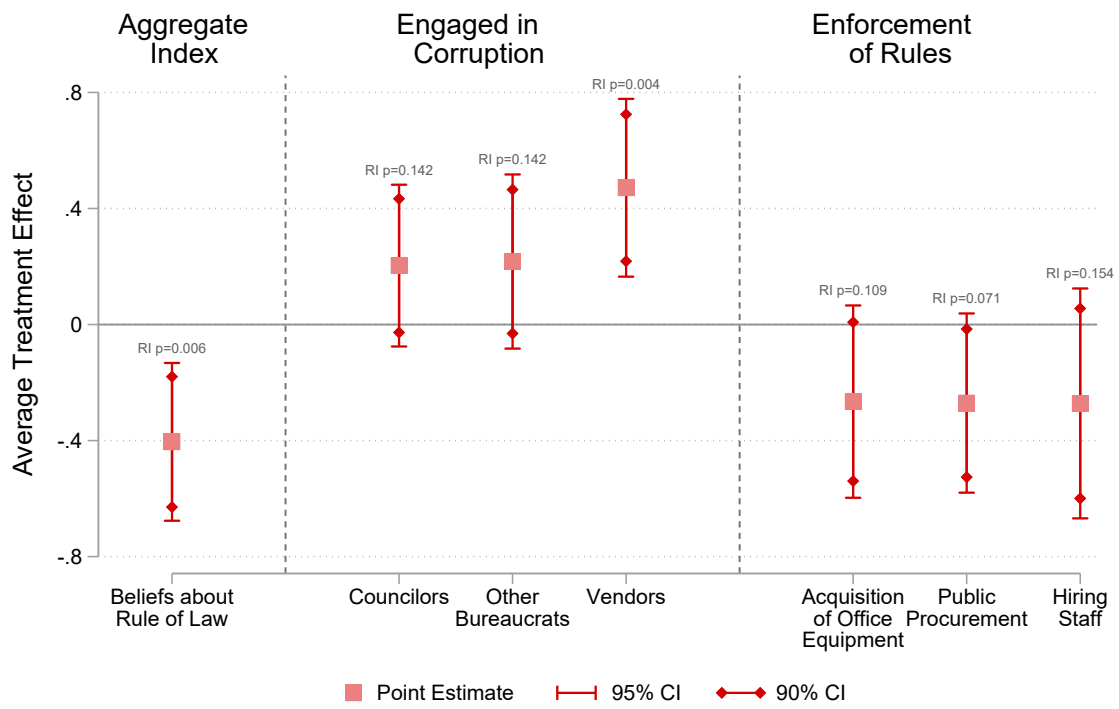
The variables have been standardized and their mean difference between treatment and control group is plotted along with 95% CIs. The standard errors are clustered at the municipality level. The source of the data is bureaucrat survey.

Figure 3: Balance test on municipal characteristics



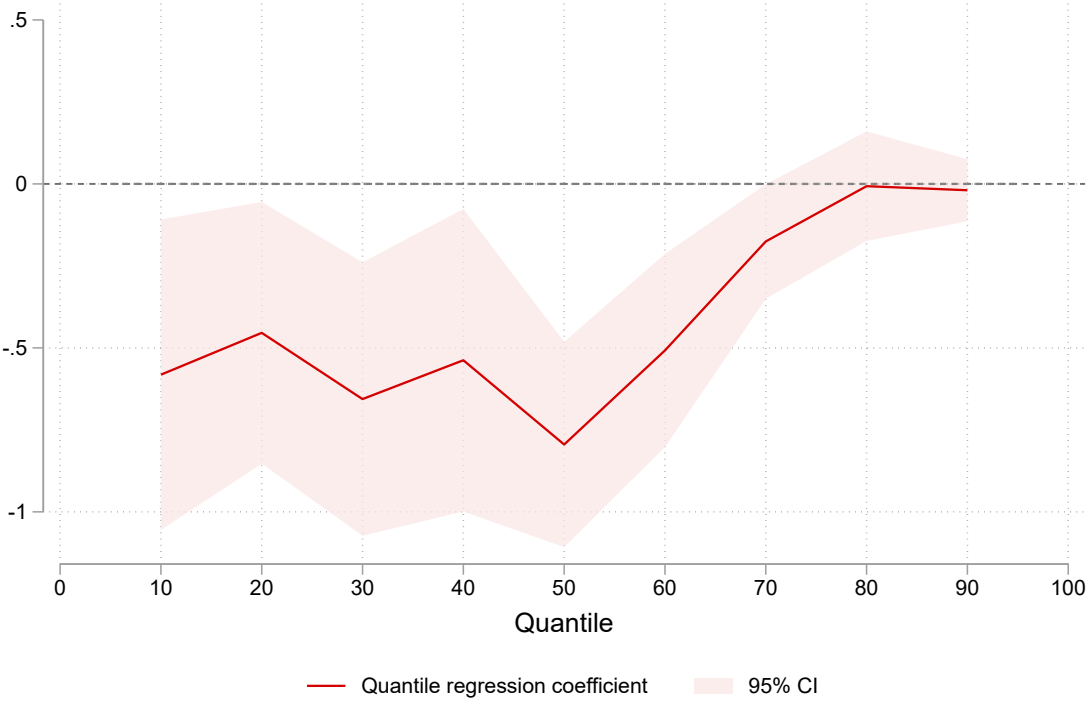
Notes. The variables have been standardized and their mean difference between treatment and control group is plotted along with 95% CIs. The standard errors are clustered at the municipality level. The source of the data are the municipal audit reports, which are available on the National Commission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

Figure 4: Did Deliberation Affect Bureaucratic Beliefs about Corruption and Enforcement of Rules



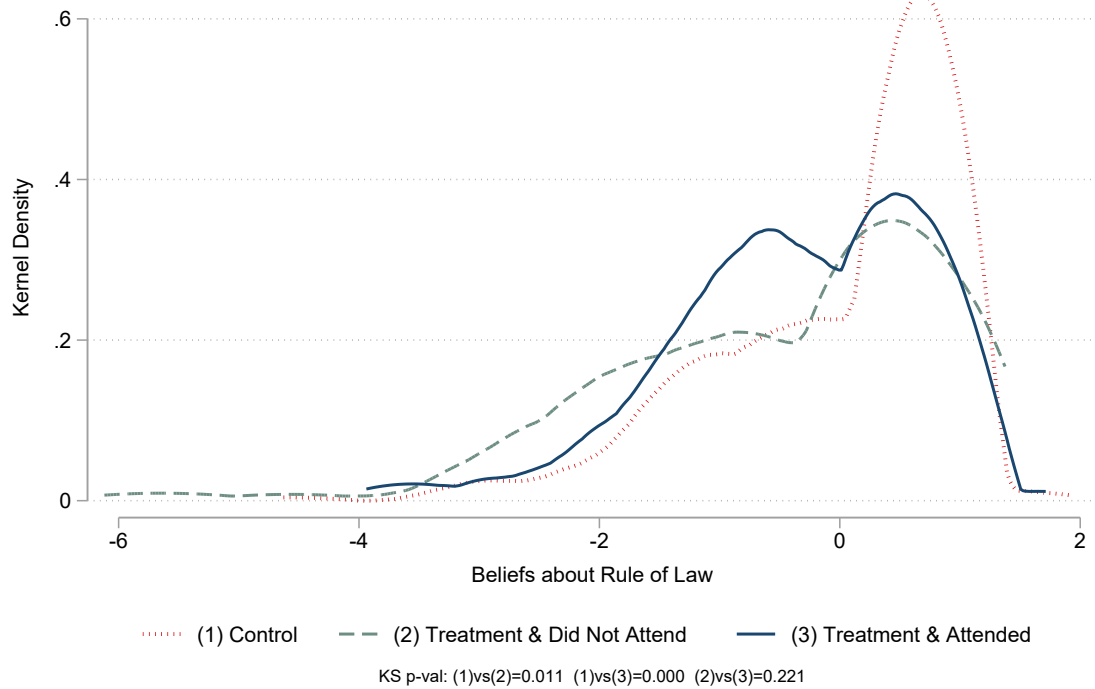
Notes: The figure reports estimates of the effect of deliberation on bureaucrats' beliefs about corruption and enforcement of rules, with 90% and 95% confidence intervals. Randomization inference p-values (1,000 repetitions) are shown above each estimate. The unit of observation is a bureaucrat. Standard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

Figure 5: Quantile Regression Estimates: Did Deliberation Affect Bureaucratic Beliefs about Corruption and Enforcement of Rules



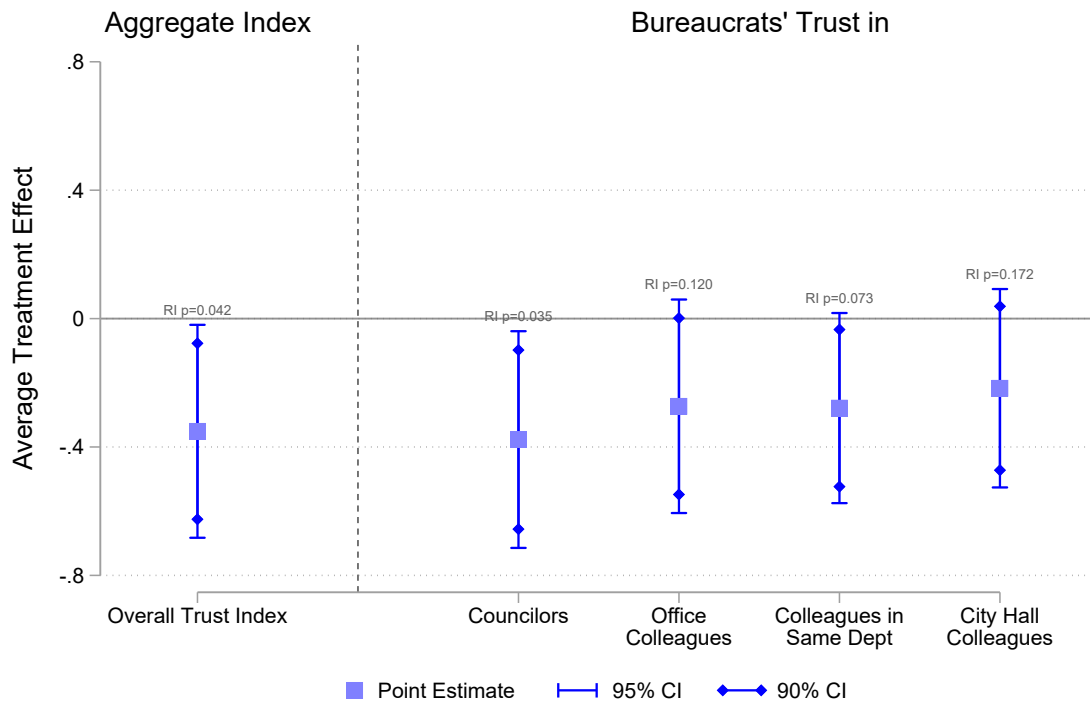
Notes: The figure reports quantile regression estimates of the treatment effect on the index of beliefs about rule of law across quantiles 10 to 90. The shaded area represents the 95% confidence interval. Standard errors are clustered at the municipality level.

Figure 6: Kernel Density of Bureaucratic Beliefs about Corruption and Enforcement



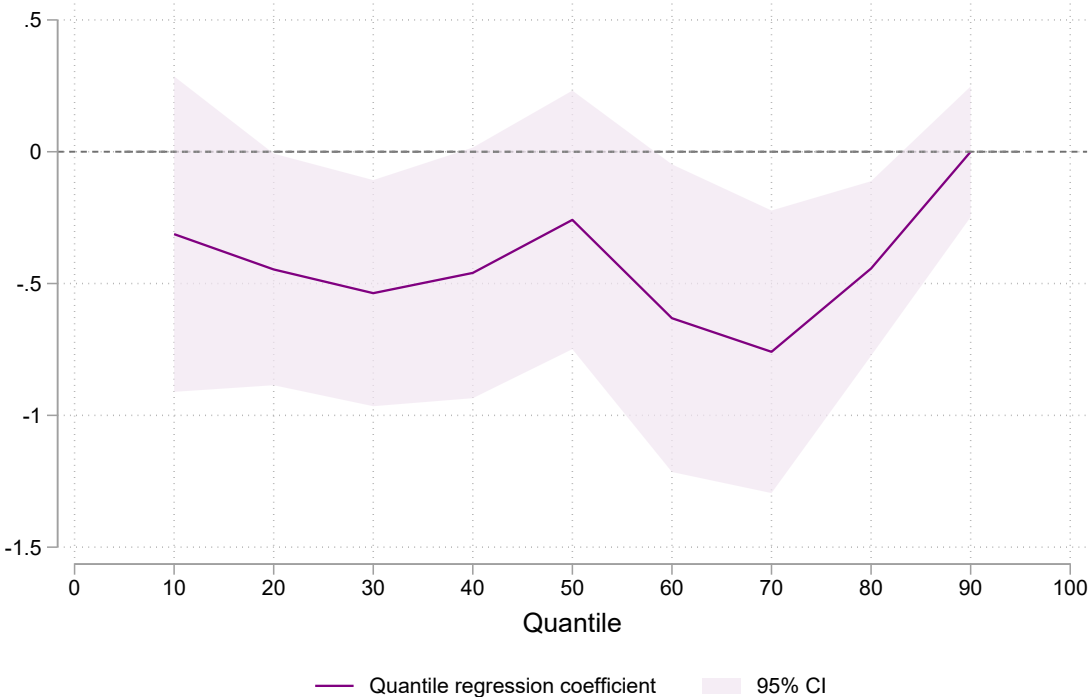
Notes: The figure shows kernel density estimates of the Beliefs about Rule of Law index by treatment and attendance status. Kolmogorov-Smirnov test p -values for pairwise comparisons are reported below the figure.

Figure 7: Did Deliberation Affect Bureaucrats' Trust in Others?



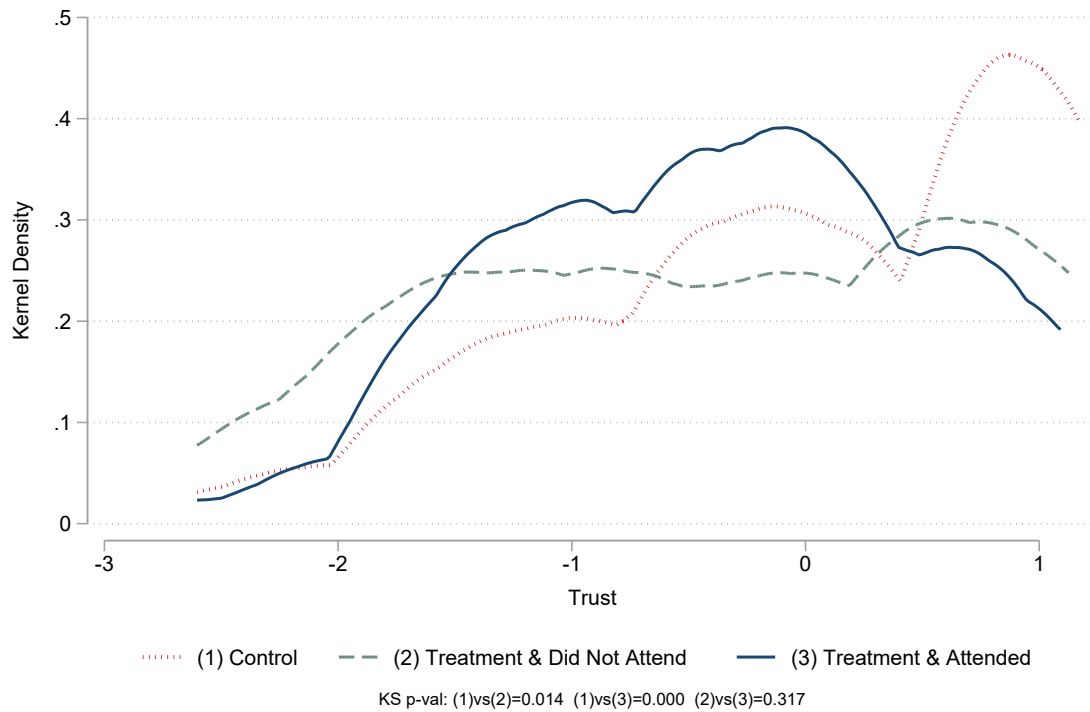
Notes: The figure reports estimates of the effect of deliberation on bureaucrats' trust, with 90% and 95% confidence intervals. Randomization inference p-values (1,000 repetitions) are shown above each estimate. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality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

Figure 8: 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 9: 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 10: Number of city-hall meetings attended by managers and workers

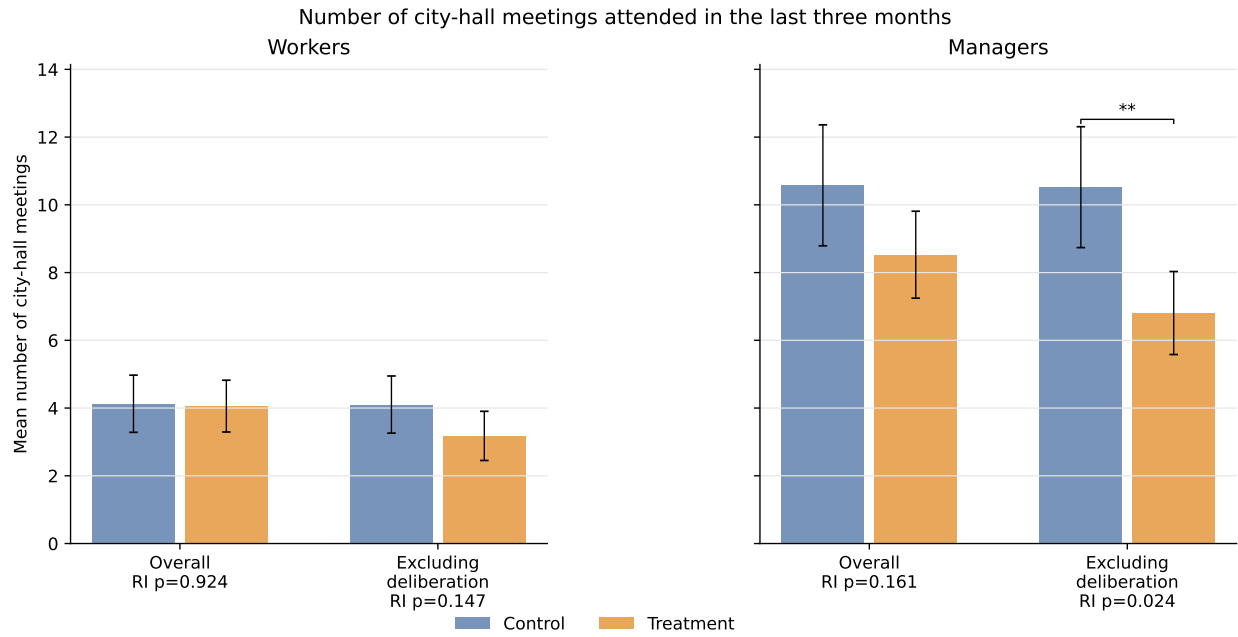
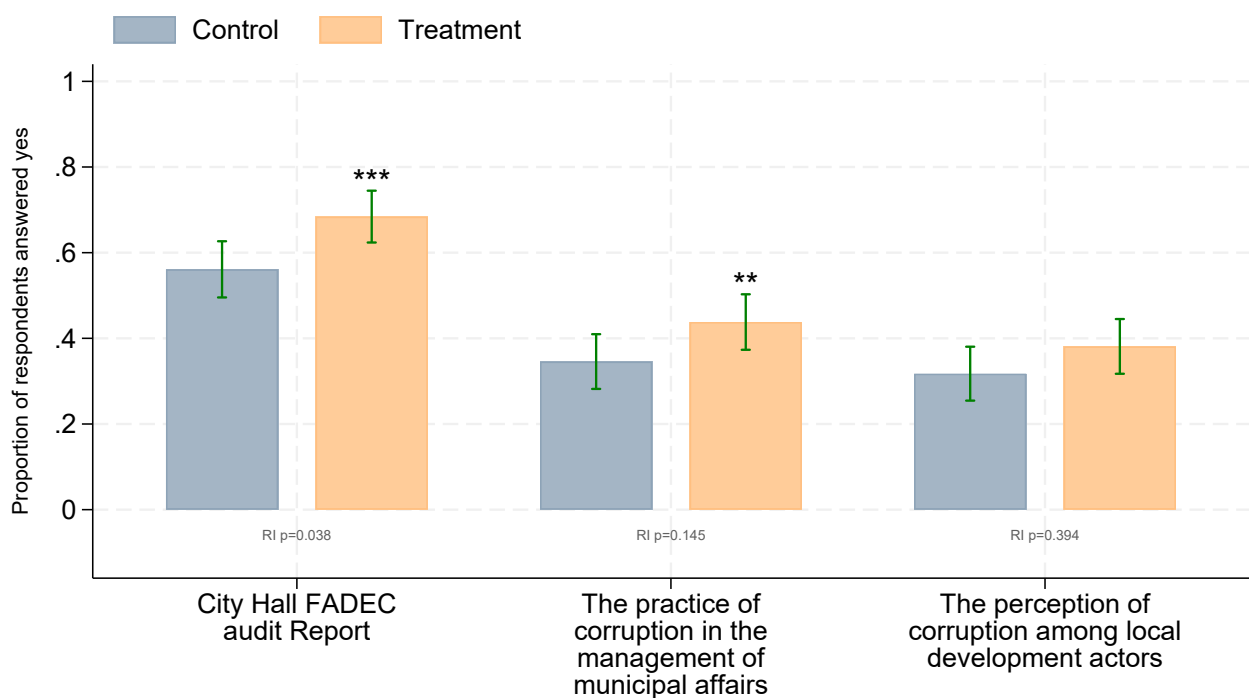
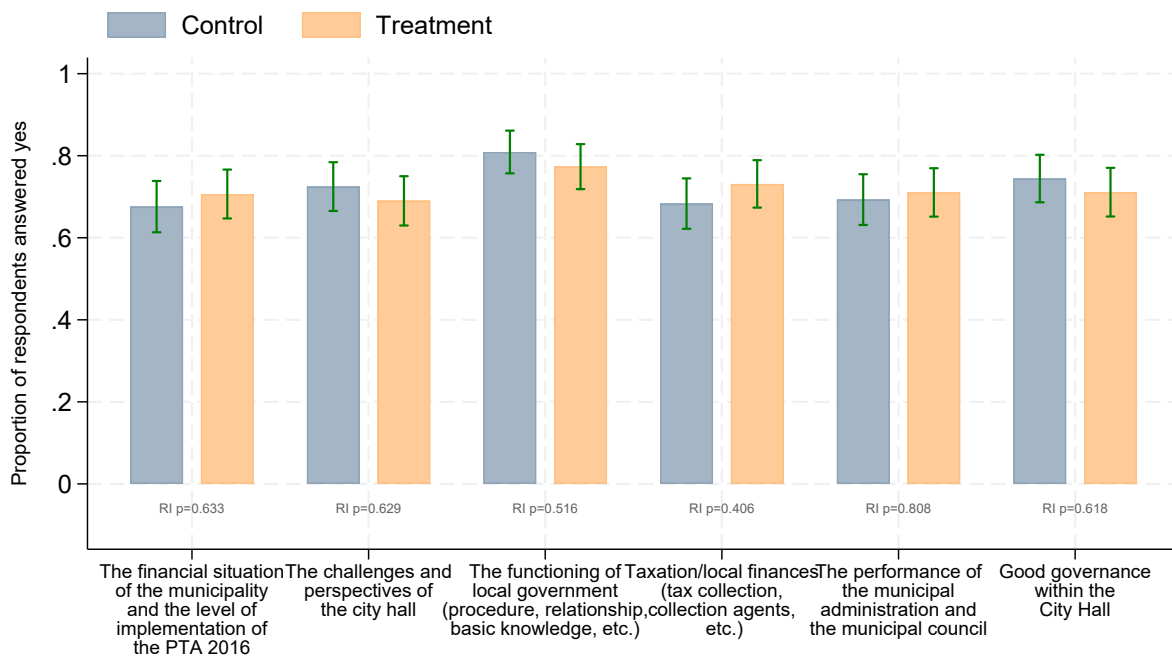


Figure 11: Topics Discussed During General Meetings in the Municipality



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 12: 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.

Appendices

A Appendix Tables	58
B Appendix Figures	67
C Deliberation Meeting Materials	84
D Audit-Based Performance Measures	86
D.1 Data Source and Digitization	86
D.2 Performance Components	86
D.3 Construction of the Performance Measure	86
D.4 Details of the criteria used by auditors	87
E Bureaucrat Survey Measures	90
E.1 Belief Measures	90
E.2 Trust	91
E.3 Workplace Interactions	92
E.4 Meeting Practices	92
E.5 Meeting Content	92
E.6 Index Construction	93
F Downstream effects on citizen experience	93

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** (735236) [0.0406]	1564939* (754951.5) [0.0520]	1546852* (763319.3) [0.0570]	2029669** (905340.7) [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: Heterogeneity based on baseline audit-based performance - Did Deliberation Affect Municipal Performance?

Dependent Variables	Below Mean Municipality		Treatment	Treatment × Obs.	P-Value: T + T ×
	Control Mean		Above Mean	Above Mean = 0	
	(1)	(2)	(3)	(4)	(5)
Municipal Audit-Based Performance	63.138	10.779*** (2.605) {0.056}	-7.572*** (2.225) {0.942}	112	0.312
Sub-Components of Municipal Audit-Based Performance					
Record Keeping	63.800	18.583* (9.860) {0.208}	-18.562** (8.692) {0.642}	112	0.997
Transparency of Public Procurement	68.238	11.061* (5.623) {0.148}	-10.110 (6.057) {0.717}	112	0.848
Overall Management	65.000	39.935** (17.788) {0.129}	-42.201** (17.554) {0.249}	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 repetitions) 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.4: Sub-components of Performance: Did Collective Deliberation Affect Municipal Audit-based Performance?

Dependent Variables	Pre-Period Control Mean (1)	DID Estimate (2)	Standard Error (3)	P-Value (4)	Randomization Inference P-Value (5)	Obs. (6)
Record Keeping	78.493	9.3016	7.7031	0.242	0.216	112
Transparency of Public Procurement	78.297	5.7866	4.6223	0.226	0.181	112
Overall Management	83.500	14.4186	13.4617	0.298	0.227	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.5: 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.6: Lee Bounds: Bureaucrats' Beliefs About Corruption and Enforcement of Rules

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

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.7: 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.8: Lee Bounds - Trust

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

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.9: Quantile Regression Estimates: Did Deliberation Affect Bureaucratic Beliefs about Corruption and Enforcement of Rules

Quantile	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
Treatment	-0.582** (0.241) {0.0570}	-0.454** (0.203) {0.0280}	-0.656*** (0.212) {0.00400}	-0.538** (0.235) {0.00700}	-0.795*** (0.159) {0.0150}	-0.508*** (0.150) {0.0140}	-0.176** (0.0889) {0.0250}	-0.00687 (0.0850) {0.0490}	-0.0193 (0.0483) {0.110}
Observations	580	580	580	580	580	580	580	580	580

Note: The unit of observation is a bureaucrat. The table reports estimates from quantile regressions of bureaucrats' Beliefs about Rule of Law index on treatment, at the 10th through 90th percentiles of the outcome distribution — corresponding to the curves plotted in Figure 5. Each column reports the treatment effect at the indicated quantile. Standard errors clustered at the municipality level are in parentheses; randomization inference p -values (1,000 reps) are in braces. Significance levels are denoted as: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Municipality-level dispersion in bureaucratic beliefs by participation in deliberation meetings

	All Beliefs		Corruption		Enforcement	
	(1)	(2)	(3)	(4)	(5)	(6)
	Attended	Not-Attended	Attended	Not-Attended	Attended	Not-Attended
Treatment	-0.393** (0.154) [0.015]	-0.177 (0.202) [0.386]	-0.574** (0.241) [0.026]	-0.074 (0.355) [0.838]	-0.247 (0.308) [0.444]	0.006 (0.251) [0.980]
Constant	1.712*** (0.121) [0.008]	1.712*** (0.121) [0.177]	2.703*** (0.175) [0.007]	2.703*** (0.175) [0.399]	2.237*** (0.144) [0.238]	2.237*** (0.144) [0.521]
Observations	20	20	20	20	20	20

Note: The unit of observation is a municipality. For each bureaucrat, we construct belief measures about corruption within the organization and about whether rules are followed (enforcement). We then compute the within-municipality standard deviation of these beliefs. Columns (1)–(2) report dispersion in overall beliefs, separately for bureaucrats who attended at least one deliberation meeting and those who did not. Columns (3)–(4) and (5)–(6) repeat this exercise for beliefs about corruption and enforcement, respectively. Standard errors clustered at the municipality level are reported in parentheses. P -values from randomization inference with 1,000 permutations are reported in square brackets. Significance levels are denoted as $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Table A.11: Quantile Regression Estimates: Did Deliberation Affect Bureaucrats' Trust in Others

Quantile	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
Treatment	-0.312 (0.305) {0.243}	-0.446** (0.224) {0.135}	-0.537** (0.218) {0.0910}	-0.460* (0.242) {0.0860}	-0.258 (0.250) {0.127}	-0.631** (0.297) {0.0480}	-0.759*** (0.273) {0.0570}	-0.443*** (0.169) {0.0250}	0 (0.126) {1}
Observations	595	595	595	595	595	595	595	595	595

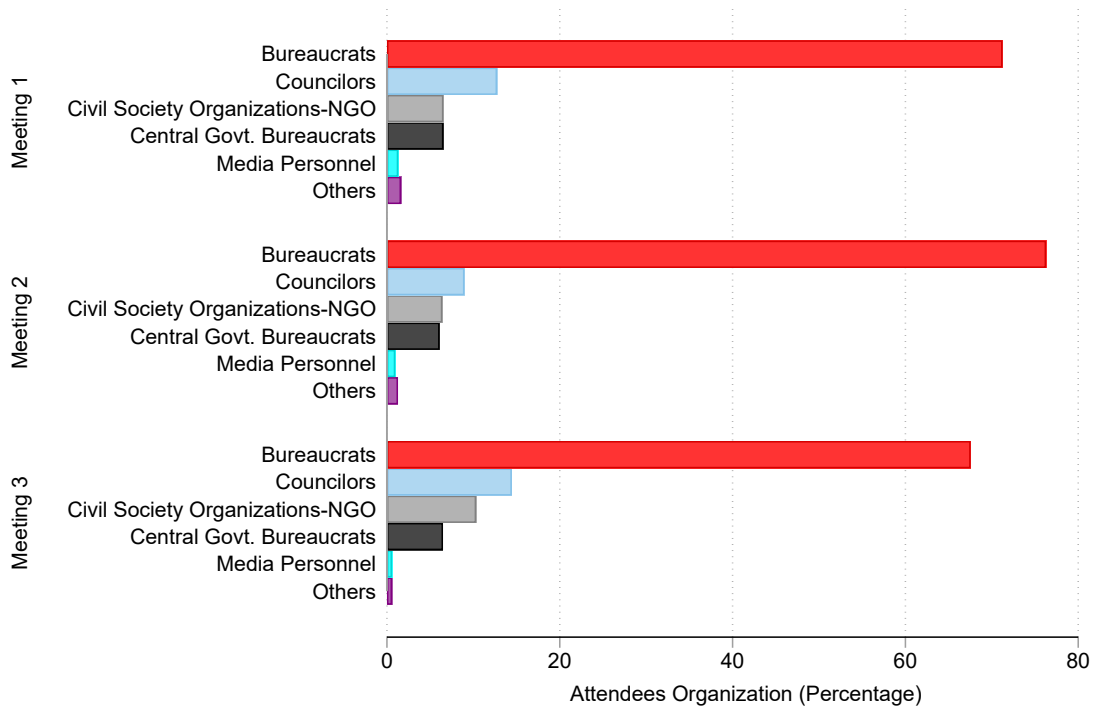
Notes: The unit of observation is a bureaucrat. The table reports estimates from quantile regressions of bureaucrats' Trust index on treatment, at the 10th through 90th percentiles of the outcome distribution — corresponding to the curves plotted in Figure 8. Each column reports the treatment effect at the indicated quantile. Standard errors clustered at the municipality level are in parentheses; randomization inference p -values (1,000 reps) are in braces. Significance levels are denoted as: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B Appendix Figures

Figure B.1: Collective Deliberation Meetings

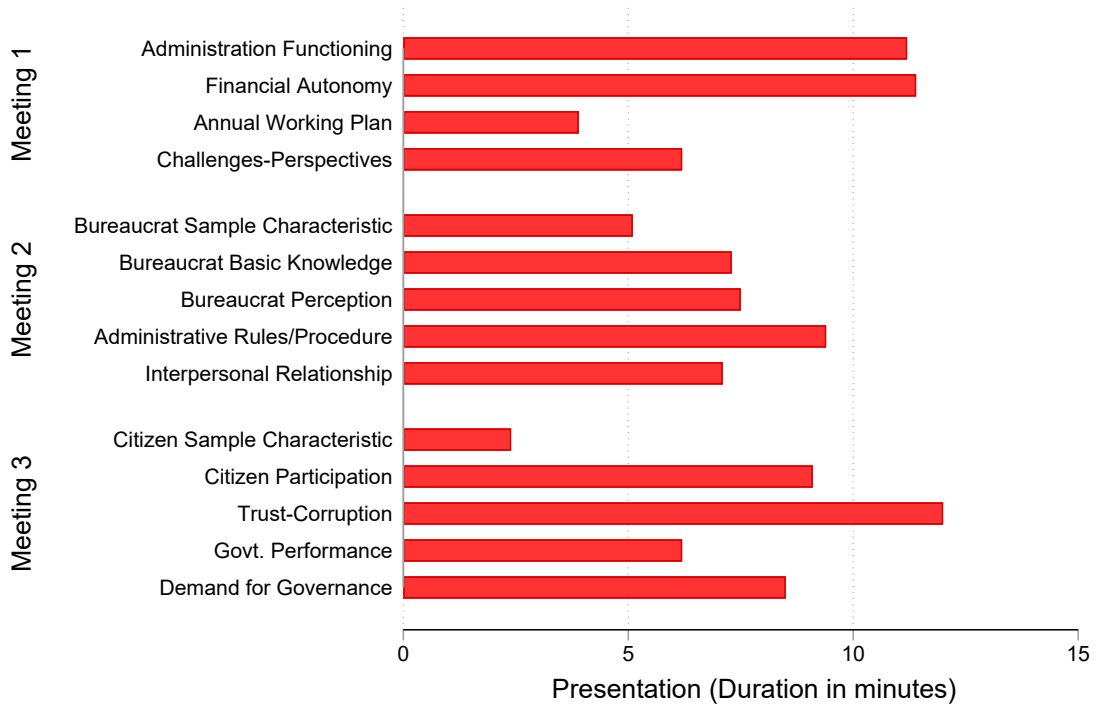


Figure B.2: Attendance by Types of Actors in the Collective Deliberation Meetings



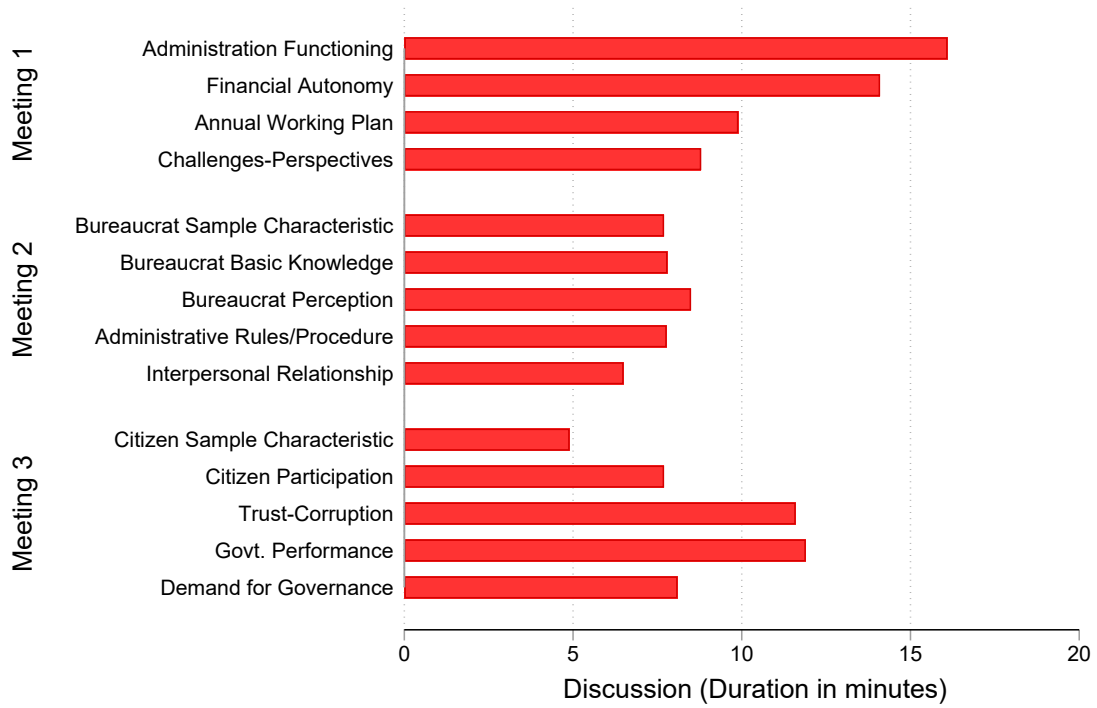
Notes: Data are collected by enumerators before the start of the meeting.

Figure B.3: Presentation topics and their duration



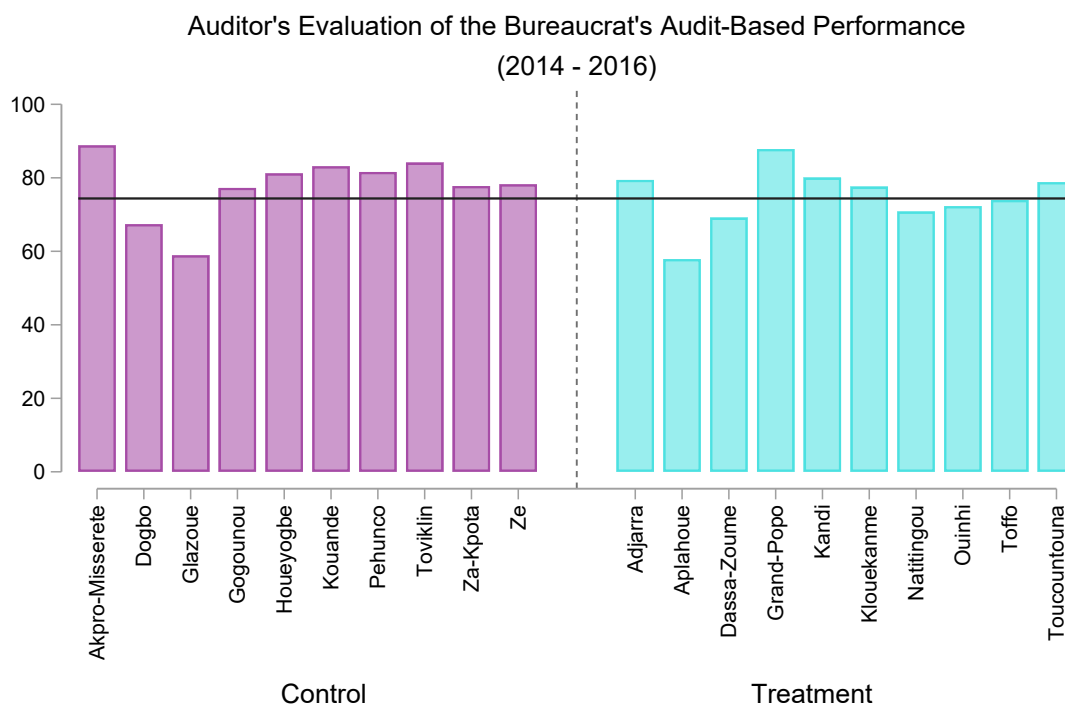
Notes: Data are from meeting notes taken by enumerators.

Figure B.4: Discussion topics and their duration



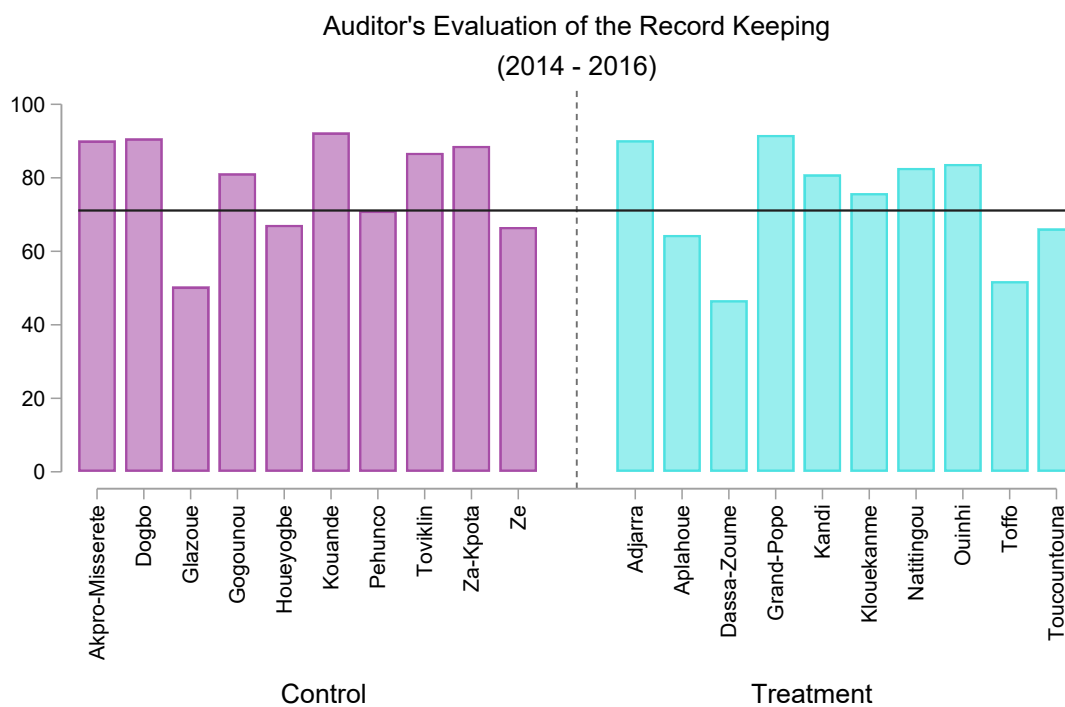
Notes: Data are from meeting notes taken by enumerators.

Figure B.5: Baseline Aggregate Average of Municipal Audit-Based Performance



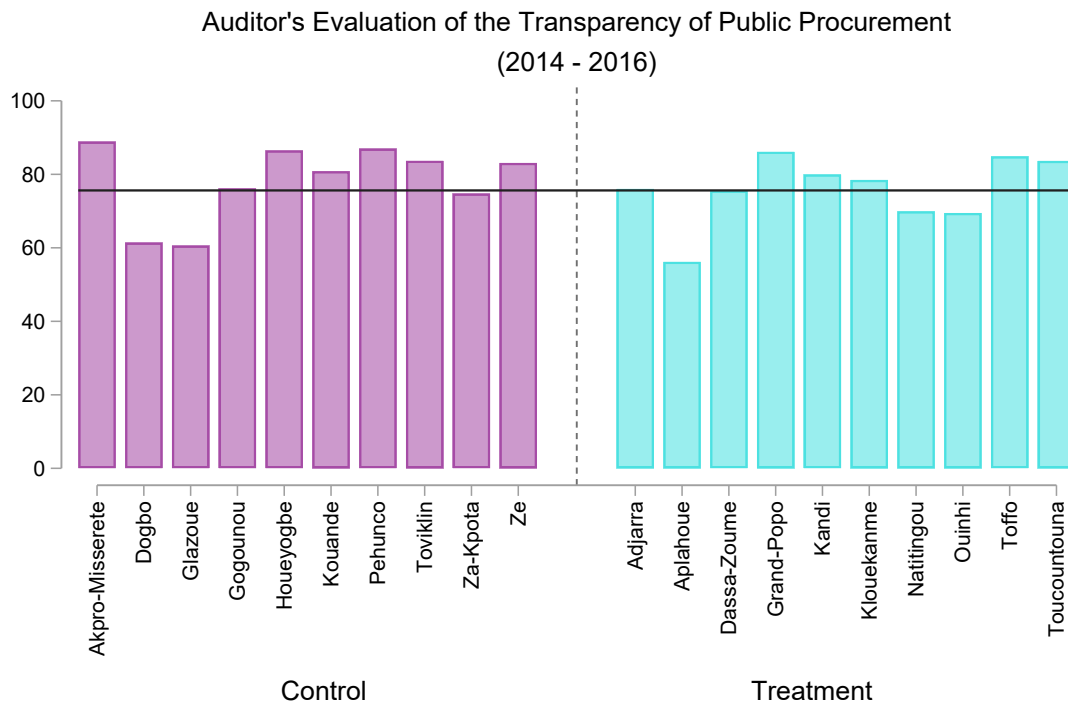
Notes. Municipal Audit-Based Performance is a score constructed from three sub-components: (i) Record Keeping (ii) Transparency of Public Procurement (iii) Overall Management. Record Keeping is constructed from scores on: maintenance of accounting records, existence of documentation room/archive accessible 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.6: Baseline Average of Auditor’s Evaluation of Record Keeping by Bureaucrats



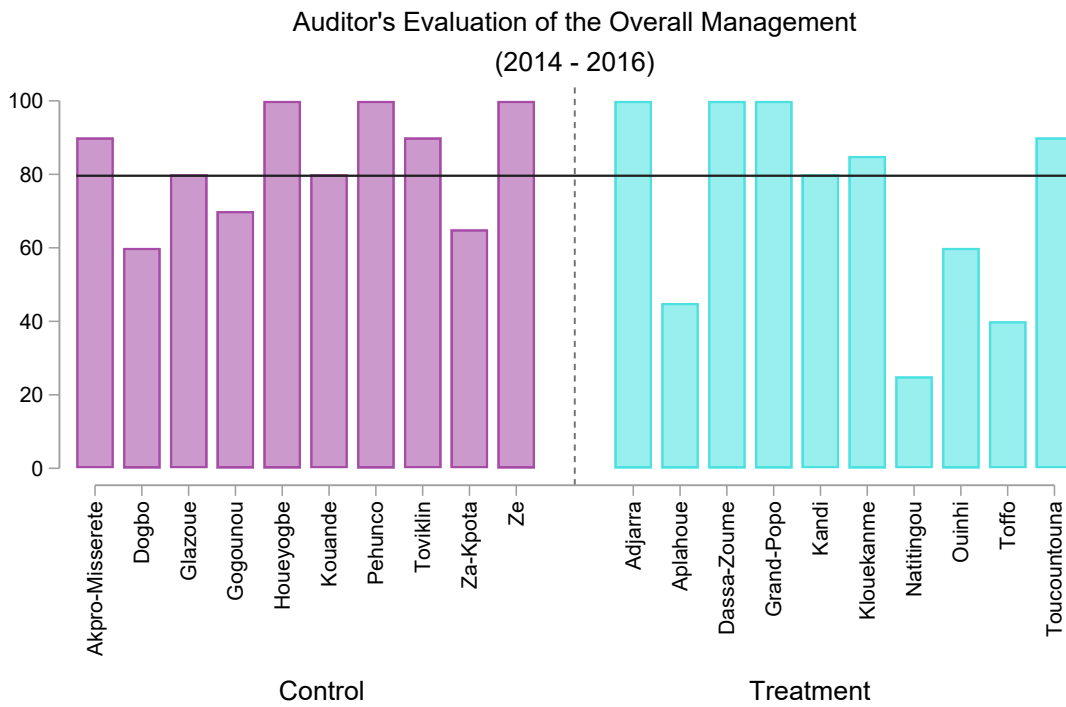
Notes. Record Keeping is constructed from scores on: maintenance of accounting records, existence of documentation room/archive accessible to the public and equipped with storage furniture, statements of transferred resources, updated physical registers and stock-taking by the authorized officers.

Figure B.7: Baseline Average of Auditor’s Evaluation of Transparency of Public Procurement by Bureaucrats



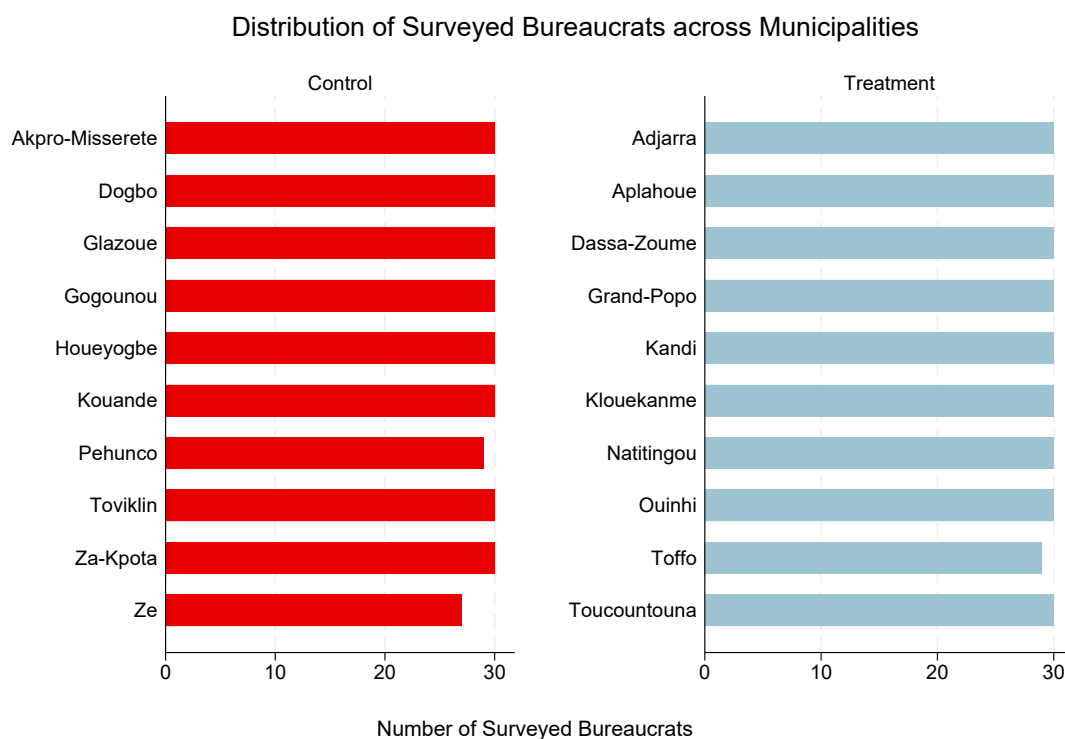
Notes. Transparency of Public Procurement is constructed from scores given on: Secretariat of the organizations responsible for public procurement is functional, the number and exhaustiveness of the reports published by them, details on FaDeC funds and spending, compliance with public procurement procedures,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

Figure B.8: Baseline Average of Auditor’s Evaluation of Overall Management by Bureaucrats



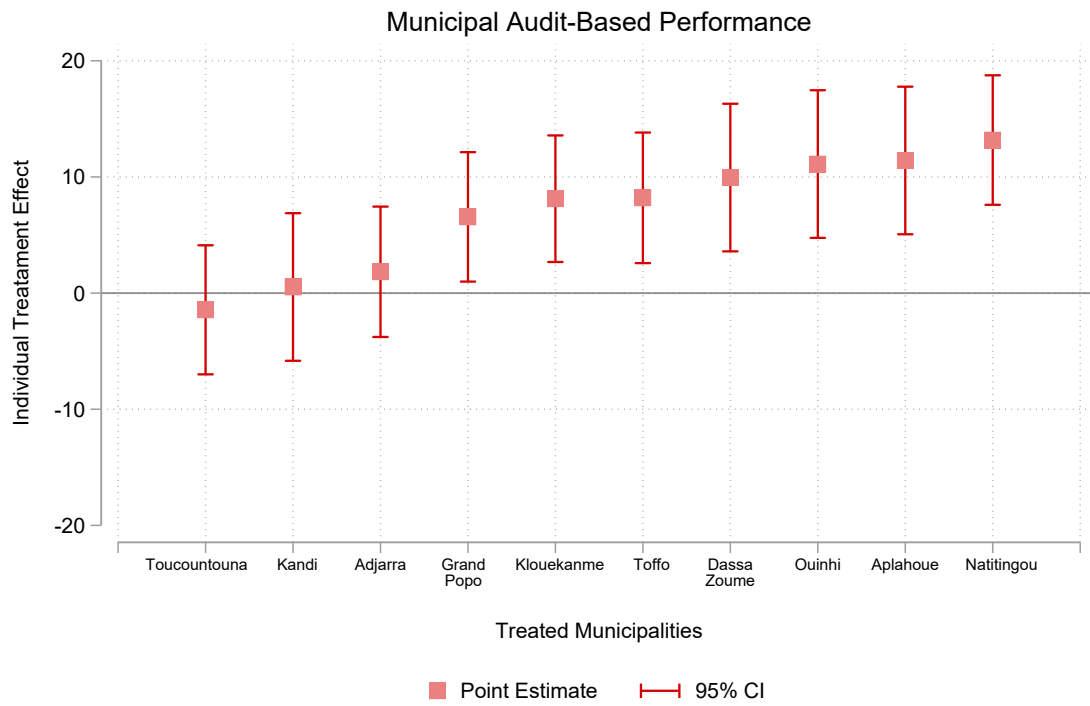
Notes. Overall Management is constructed from scores given on effectiveness of the role of coordinator of local services by the Secretary General, and management of certificates of civil status (certificates of birth, death, marriage and divorce)

Figure B.9: Number of bureaucrats surveyed across treatment and control municipalities in December 2016



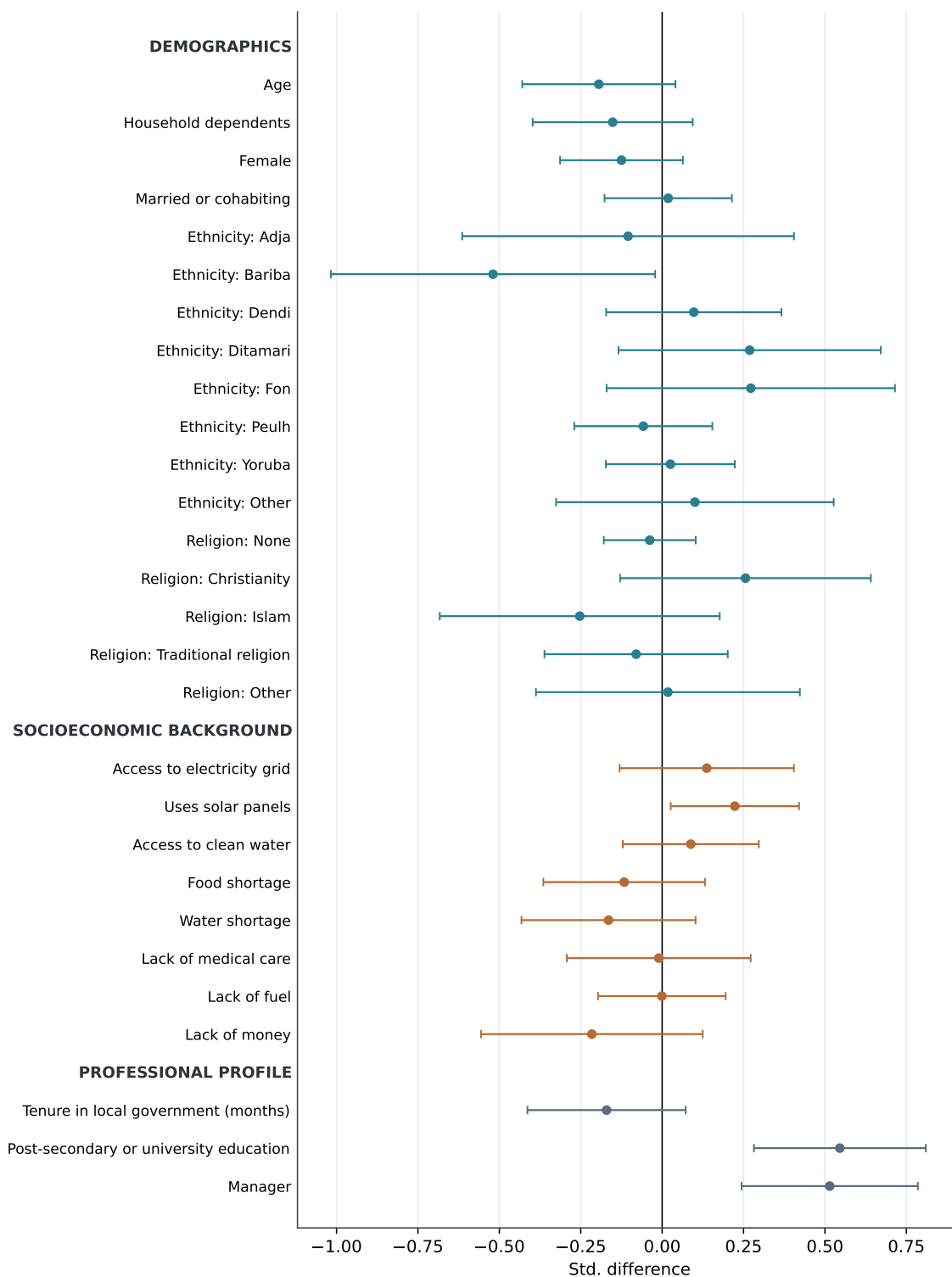
Notes. For the endline survey conducted in December 2016, 30 bureaucrats were randomly selected within each of the 20 municipalities. However, there were some exceptions: 27 in Ze and 29 each in Pehunco & Toffo, bringing the total number of observations to 595.

Figure B.10: 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.

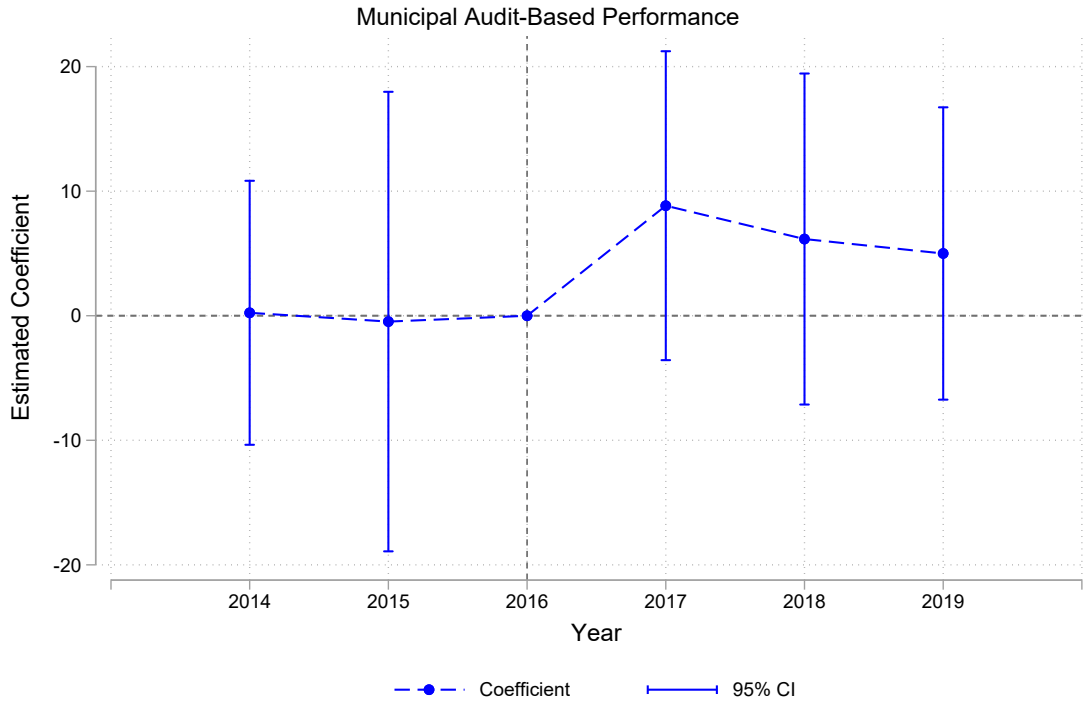
Figure B.11: Characteristics of those bureaucrats that attended any of the deliberation



Notes. Each point is the coefficient from a separate regression of a standardized respondent characteristic on an indicator for attending at least one deliberation meeting. Bars show 95% confidence intervals using

standard errors clustered by municipality.

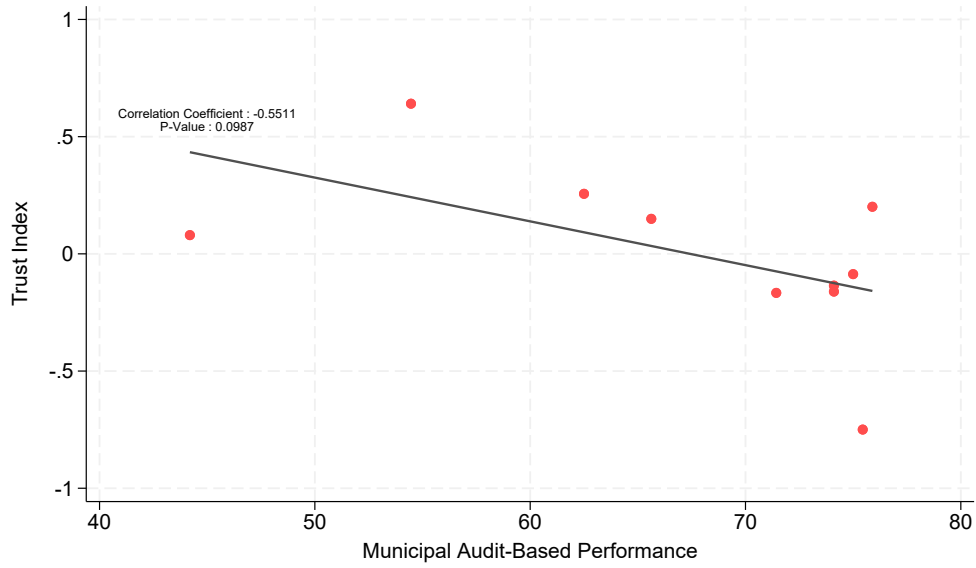
Figure B.12: Event Study Plot of Municipal Audit-Based Performance



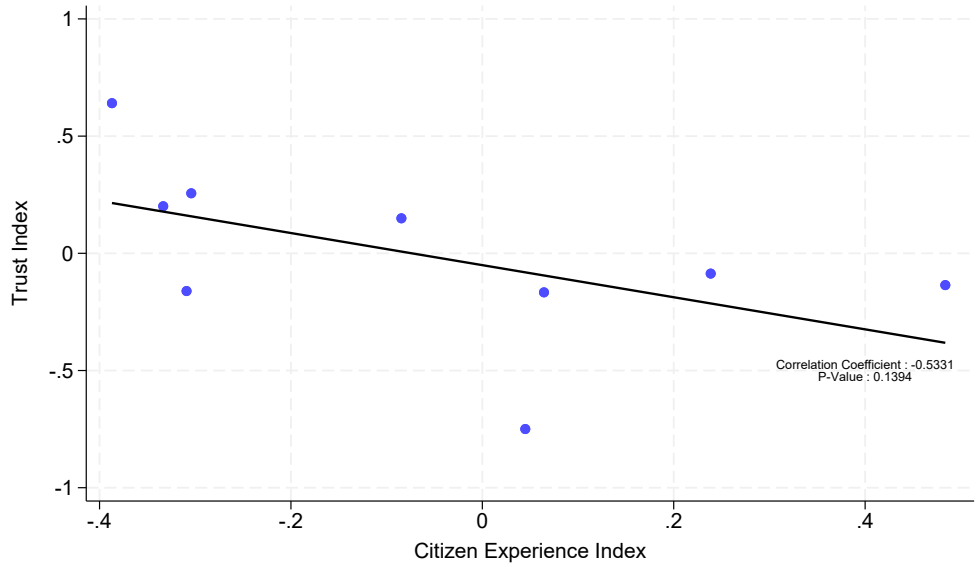
Notes. Municipal Audit-Based Performance is a score built from the audit reports of FADeC. It score is constructed from three sub-components: (i) Record Keeping (ii) Transparency of Public Procurement (iii) Overall Management. Record Keeping is constructed from scores given based on availability of quarterly 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.13: Baseline Relationship between Trust and Performance Outcomes

(a) Trust and Audit-Based Performance

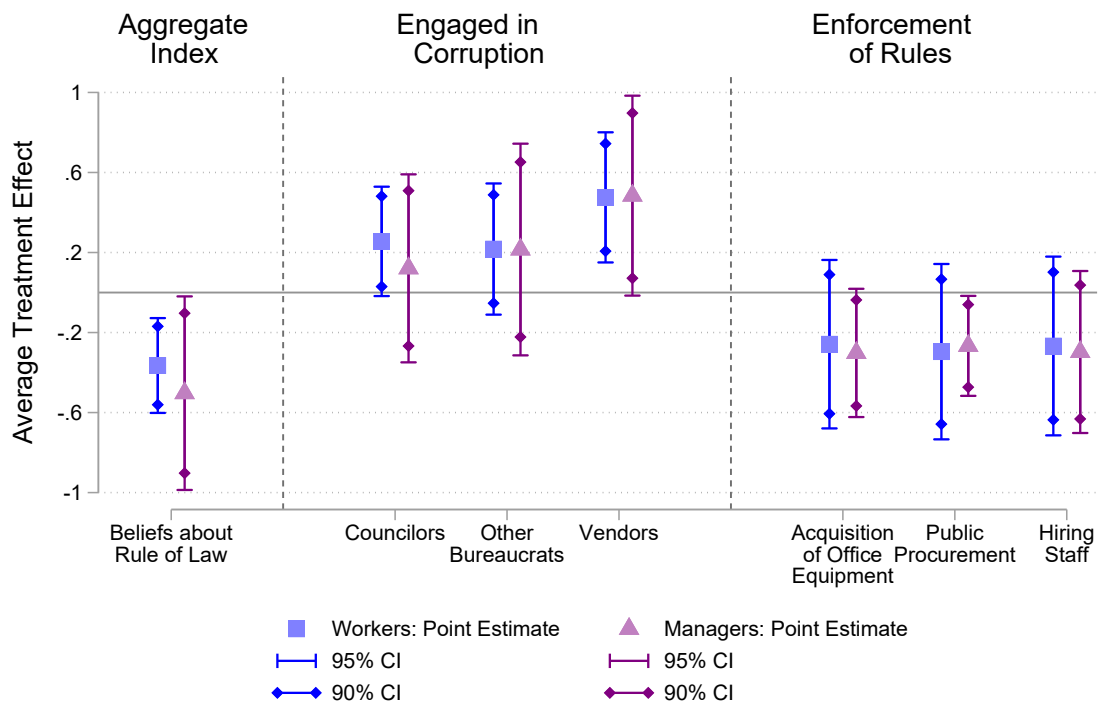


(b) Trust and Citizen Experience



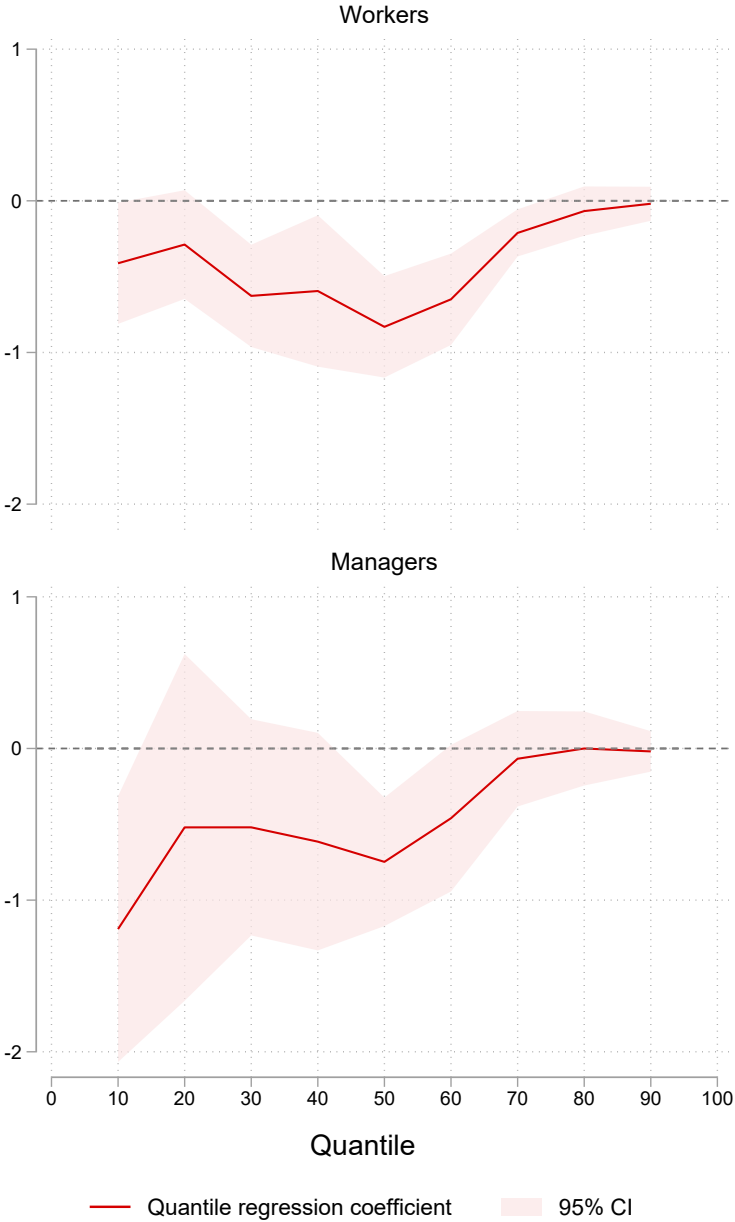
Notes: The top panel plots the relationship between the trust index and municipal audit-based performance. The bottom panel plots the relationship between the trust index and the citizen experience index.

Figure B.14: Workers and managers: Effects of deliberation on bureaucratic beliefs about rule of law



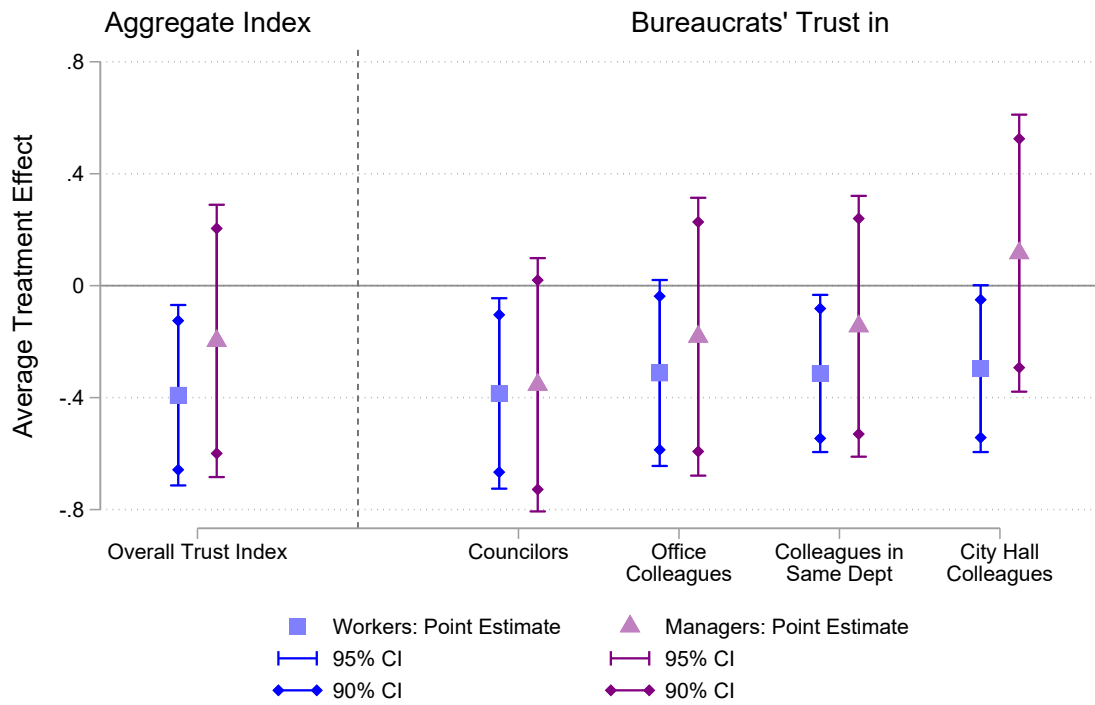
Notes: The figure reports estimates of the effect of deliberation on beliefs about corruption and enforcement of rules, separately for workers and managers, with 90% and 95% confidence intervals. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality level.

Figure B.15: Workers and managers: Quantile estimates of the effects of deliberation on bureaucratic beliefs about rule of law



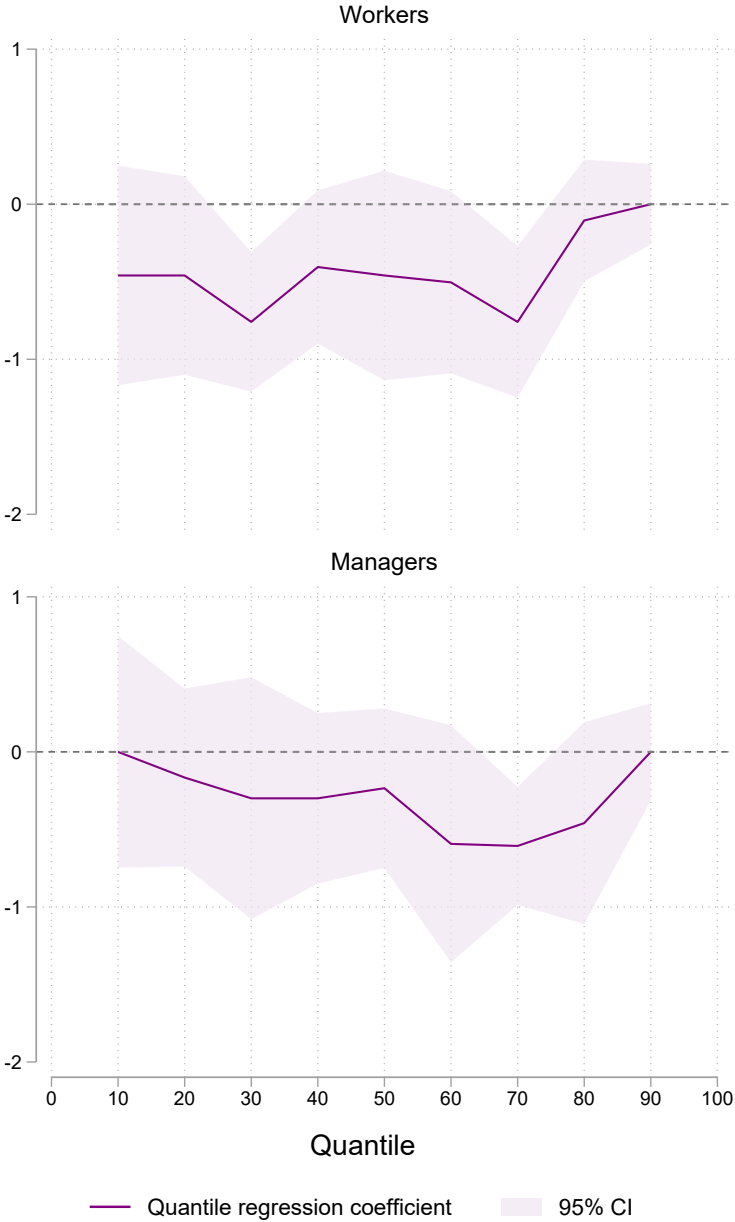
Notes: The figure reports quantile regression estimates of the treatment effect on the Beliefs about Enforcement index, separately for workers and managers, across quantiles 10 to 90. The shaded area represents the 95% confidence interval. Standard errors are clustered at the municipality level.

Figure B.16: 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.17: 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.

C Deliberation Meeting Materials

Figure C.18: Meeting 1: Audit-based presentation

Presentation Outline

1. **Administrative and Organizational Functioning of the Municipality**
2. **Municipality's Financial Statement**
3. **Level of Implementation of the Annual Working Plan in 2016**
4. **Challenges and Prospects**



Resource management and financial statement of the Municipality

Resource management at the level of authorizing officer

Information about resource availability and accounting

✓ **Observation**

- The local treasury has informed the municipality of resource availability;
- Very long effective time taken to make BTRs available compared to the date of issue;
- Auxiliary account records are updated to reflect the availability of resources to commit to spending needs.

✓ **Risks**

- Delays in budget execution;
- Low consumption of credits transferred;
- Lack of Traceability in the Management of FADeC Resources.



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.19: Meetings 2 and 3: Survey-based presentations

Some Insights

Obstacles to the smooth running of the administration



Obstacles to the smooth running of the administration:

- A. Dependency on certain individuals
- B. Absence of failure of texts
- C. Non-compliance with texts and rules
- D. Lack of equipment and infrastructure

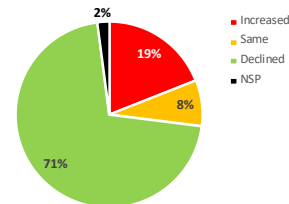
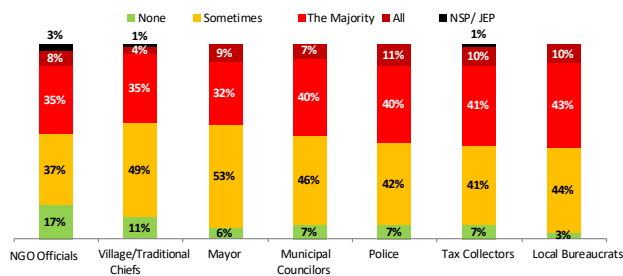
Meeting-2: Survey outcome -Staff of local administration

Institutional Trust and Corruption

Corruption Perception

- Corruption is perceived at all levels of local affairs, but to varying degrees

- 19% of sample report that corruption has increased the last 12 months



Meeting--3: Results of the survey -- populations of the Adjarra

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.1, D.2, and D.3 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.

$$\text{Municipal Audit - Based Performance}_{mt} = \frac{\text{RawScore}_{mt}}{\text{MaxScore}_t} \times 100$$

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.4 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.1: 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.2: 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.3: 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.4: 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 \(2008b\)](#).

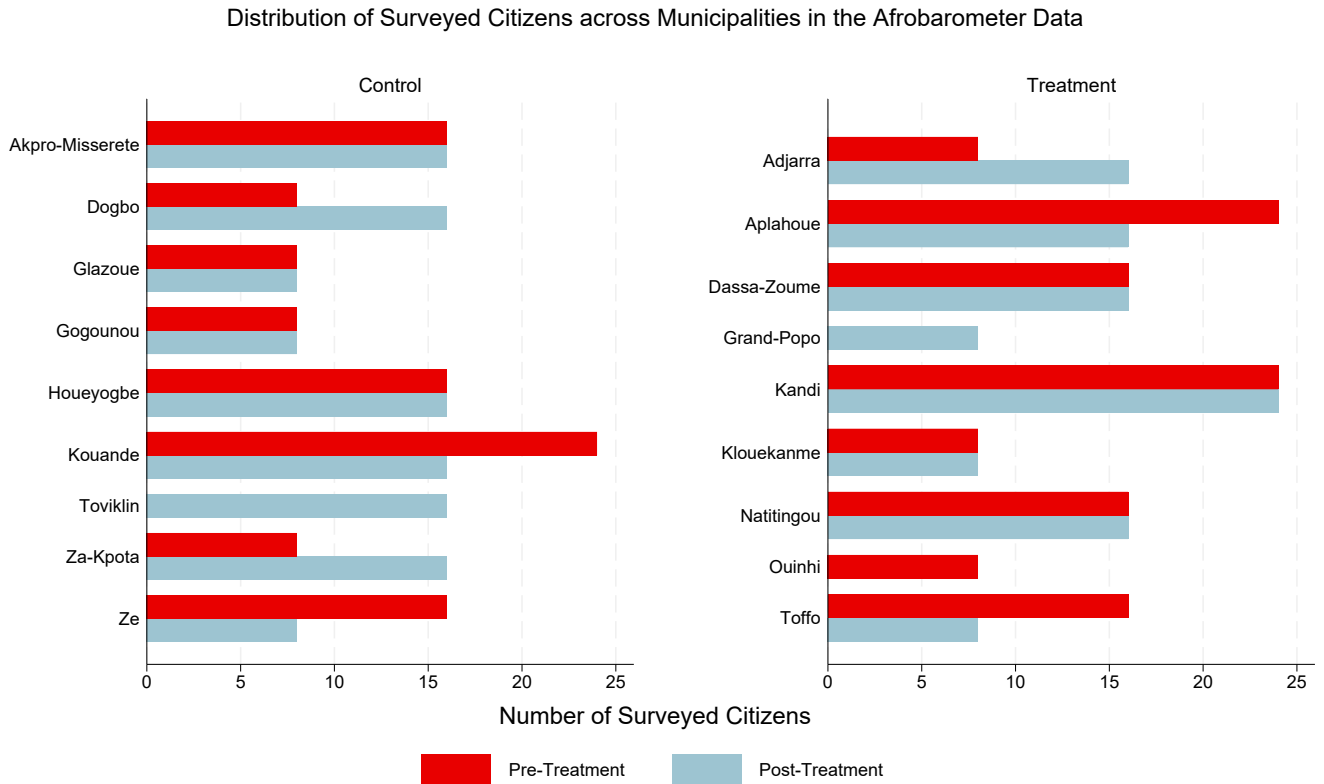
F Downstream effects on citizen experience

While our primary outcome is administrative performance as measured through municipal audits, an important question is whether these improvements translate into better citizen experiences with local government. To examine downstream effects on public service delivery, we use data from the Afrobarometer, a nationally representative cross-sectional survey. The survey provides independent measures of citizens' interactions with local government, including access to services and exposure to corruption.

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, allowing us to examine short-run changes in citizen experiences. Moreover, a few municipalities that appear in both the pre- and post-treatment audit data are missing observations in one Afrobarometer round (Toviklin, Grand-Popo, and Ouinhi), while Pehunco (control) and Toucountouna (treatment) are not covered in the Afrobarometer data altogether. 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 respondents paid bribes. Variables reflecting worse outcomes 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 1](#), replacing the outcome with citizen experience.

Table [F.1](#) 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, the effects are larger 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 [F.2](#)). Estimates for individual components of the index (Table [F.3](#)) follow a similar pattern. While estimates are imprecise, the direction of the results, particularly in lower-performing municipalities, together with 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.

Figure F.1: 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.

Table F.1: Did Bureaucratic Deliberation Affect Citizens' Outcomes (2014, 2017)?

Dependent Variables	Pre-Period Control Mean (1)	DID Estimate (2)	Standard Error (3)	P-Value (4)	Randomization Inference P-Value (5)	Obs. (6)
Aggregate Index						
Citizen Experience Index	.	0.1668	0.1498	0.281	0.375	456
Panel A : Corruption						
Corrupt among Govt Officials	0.960	0.0005	0.0407	0.990	0.992	444
Officials Go Unpunished	0.892	-0.0196	0.0472	0.683	0.716	449
Panel B : Present Living Conditions						
Are Good	0.279	0.0516	0.1127	0.653	0.679	456
Better Compared to Others	0.233	0.0961	0.0705	0.191	0.227	455
Panel C : Difficulty in Access to						
Identity Document Services	0.652	-0.0675	0.1595	0.678	0.669	206
Household Services	0.857	0.0159	0.1434	0.913	0.917	88
Panel D : Paid Bribes for						
Identity Document Services	0.304	-0.0141	0.1250	0.912	0.907	206
Household Services	0.143	-0.4373	0.2576	0.110	0.129	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 F.2: Heterogeneity based on baseline citizen experience

Dependent Variables	Treatment (1)	Treatment \times Above Mean (2)	Obs. (3)	P-Value: T + T \times Above Mean = 0 (4)
Aggregate Index				
Citizen Experience Index	0.285* (0.149) {0.201}	-0.256*** (0.0790) {0.730}	432	0.845

Notes: The unit of observation is a citizen-year (Afrobarometer 2014 and 2017). The outcome is the citizen experience index, a standardized aggregate of perceptions of corruption, living conditions, and difficulty / bribery in accessing public services and identity documents (sub-components reported in Table A.13). Above- vs. below-mean municipalities are classified by the baseline (2014) citizen experience index. The estimating equation is $y = \beta_1(T \times Post) + \beta_2(T \times Post \times Above) + \gamma_t + \delta_c + \varepsilon$ with year and commune fixed effects. Column (1) reports the control mean for below-mean municipalities at baseline; Columns (2) and (3) report the treatment effect for below-mean municipalities and the differential effect for above-mean municipalities; Column (5) reports the p -value for the sum of the two (the treatment effect for above-mean municipalities). Standard errors clustered at the commune level in parentheses; randomization inference p -values (1,000 reps) in braces. Significance levels: $p < 0.1$, $p < 0.05$, $p < 0.01$.

Table F.3: Heterogeneity for sub-components of citizen experience index based on baseline citizen experience

Dependent Variables	Below Mean Municipality	Treatment	Treatment ×	Obs.	P-Value: T + T ×
	Control Mean		Above Mean		
	(1)	(2)	(3)	(4)	(5)
Panel A: Corruption					
Corrupt among Govt Officials	1	-0.0293 (0.0385) {0.636}	0.0646* (0.0335) {0.650}	420	0.427
Officials Go Unpunished	0.894	-0.0381 (0.0545) {0.637}	0.0402 (0.0566) {0.897}	425	0.970
Panel B : Present Living Conditions					
Are Good	0.0833	0.146 (0.102) {0.435}	-0.206* (0.114) {0.705}	432	0.674
Better Compared to Others	0.125	0.163** (0.0744) {0.0670}	-0.145* (0.0845) {0.202}	431	0.832
Panel C : Difficulty in Access to					
Identity Document Services	0.680	-0.0119 (0.188) {0.964}	-0.121 (0.119) {0.757}	195	0.375
Household Services	0.778	0.0118 (0.178) {0.942}	0.0105 (0.200) {0.916}	86	0.896
Panel D : Paid Bribes for					
Identity Document Services	0.200	0.0858 (0.114) {0.534}	-0.218* (0.114) {0.282}	195	0.340
Household Services	0.222	-0.380 (0.288) {0.160}	-0.147 (0.210) {0.743}	86	0.0520

Note: The unit of observation is a citizen-year. The analysis is based on Afrobarometer rounds 6 (2014) and 7 (2017) and reports treatment effects on the sub-components of the citizen experience index in Table A.12. Panel A consists of indicators of whether citizens perceive officials to be engaged in corruption and to go unpunished; Panel B, whether their present living conditions are good and whether they are better compared to others; Panels C and D, whether citizens faced difficulties in accessing or had to pay bribes for (i) identity documents and (ii) household services. Each row reports estimates from a single regression of the form $y = \beta_1, (T \times Post) + \beta_2, (T \times Post \times Above) + \gamma_t + \delta_c + \varepsilon$, with year and commune fixed effects; above- vs. below-mean municipalities are classified by the baseline (2014) citizen experience index. Columns (2) and (3) report β_1 — the treatment effect for below-mean municipalities — and β_2 — the differential effect for above-mean municipalities; Column (5) reports the p -value for $\beta_1 + \beta_2 = 0$. Standard errors clustered at the commune level are in parentheses; randomization inference p -values (1,000 reps) are in braces. Significance levels: $p < 0.1$, $p < 0.05$, $p < 0.01$.