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

Shan Aman-Rana† Leonard Wantchekon‡ Lazare Kovo§

August 9, 2026

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 8.7 percent of the control mean. These improvements are broad-based: deliberation reduces financial irregularities, moves procurement away from less competitive procedures, strengthens project implementation and reallocates investment toward citizen-facing services rather than the administration's own overhead. While performance improves, interpersonal trust declines. These results highlight the potential for low-cost institutional changes to improve bureaucratic performance by changing how existing organizational performance information is used within bureaucracies, 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, H83, O1

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

*We are grateful to Michael Best, Nick Bloom, Anujit Chakraborty, Gaurav Chiplunkar, Guido Friebel, Paul Gertler, Sam Kapon, Virginia Minni, Clement Minaudier, Priyama Majumdar, Diana Moreira, Arman Rezaee, John Van Reenen, Raffaella Sadun, Megha Patnaik, Sandip Sukhtankar, Silvia Vannutell, Gerard Padro i Miquel, Paola Sapienza, Ashish Shenoy, Sheetal Seekhri, Ivo Welch, Arthur Zeyang Yu, Shaoda Wang, Yanhui Wu and participants at seminars and workshops for comments and suggestions. Sanjog Rajiv, Saad Hassan, Camila Janada, and Leyao Wang provided excellent research assistance. Financial support from J-PAL Governance Initiative is gratefully acknowledged. This trial is registered retrospectively in the AEA RCT Registry (AEARCTR-0019263) and was approved by the Princeton University IRB (#7220).

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

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

§Department of Political Science, 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. A complementary policy lever could be to modify institutional processes, or the "rules of the game," within organizations to improve performance.

One such rule is the organization's information structure. Existing work has focused on two margins: information about individual bureaucrats' performance available to their principals, and information provided to workers about their performance relative to their peers.[2] Both rely on measures of individual performance, which are often difficult to construct in bureaucracies, where officials perform multiple tasks and many outputs are hard to measure (Dixit, 2002; Finan et al., 2017). Information on the performance of the organization as a whole is more widely available—governments routinely produce audit reports, inspection reports, budget execution reviews, and citizen surveys—but is typically produced for reporting and external oversight rather than for use within the bureaucracy. Because it describes how the organization as a whole is functioning, such information can affect beliefs about how others behave and can shape each bureaucrat's own behavior.

In this paper, we conduct a field experiment with municipal bureaucracies in Benin to test whether changing beliefs about aggregate organizational 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.

[1]There 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; George and Mattsson, 2025), selection (Moreira and Pérez, 2024; Dal Bó et al., 2013; Hanna and Wang, 2017; 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).

[2]On information about individual bureaucrats available to principals, see Dhaliwal and Hanna (2017); Callen et al. (2020); Muralidharan et al. (2021); Dal Bó et al. (2021); Raffler (2022); Mattsson (2025); Debnath et al. (2023). On workers' relative-performance information within organizations, see Blanes i Vidal and Nossol (2011); Barankay (2012); Blader et al. (2020).

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 revise their beliefs about the organization, coming to believe that corruption is more prevalent and rule enforcement weaker than they had previously thought; they attend fewer meetings, though discussions increasingly center on 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 tension: gains in performance can come alongside a decline in interpersonal trust.

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, 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.

With formal incentives weak, informal relationships and the interpersonal trust that underpins them play an important role in shaping bureaucratic behavior. In our setting, however, this trust does not appear to support performance: among the control municipalities, higher interpersonal trust is associated with lower municipal performance (-0.55, p-value < 0.1) and worse citizen experiences with government (-0.53, p-value $= 0.14$), consistent with relationships that sustain collusion and low-effort equilibria rather than accountability (Tirole, 1986; Milgrom and Roberts, 1988).[3]

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.[4]

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

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

[4]Implementing 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).

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 performance.

In the treated municipalities, three deliberation meetings were held at monthly intervals, beginning in September 2016, and 58% of all bureaucrats attended at least one of the three sessions. Each meeting focused on a distinct source of performance information. The first presented the municipality's own audit report—the observations of the central government auditors, which were already available within the administrative system but were not typically discussed collectively across the bureaucracy. The second and third presented results from a bureaucrat survey and a citizen survey, respectively, both conducted by the research team one month before the intervention to complement the audit report. Participants were shown averages from the survey data and the auditors' highlights of overall municipal performance, rather than anecdotal complaints or accusations directed at particular individuals. The audit report and the citizen survey were specific to the municipality, whereas the bureaucrat survey was presented in aggregate across all study municipalities, to reduce social-desirability concerns.

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.

We measure bureaucratic performance using municipality-level administrative audit data from 2014 to 2019, conducted by the central government oversight body, the National Commission for Local Finance (CoNaFiL).[5] These audits are the primary instrument through which municipal performance is evaluated, and they assess core dimensions of bureaucratic functioning, including record keeping, transparency in public procurement, and overall administrative management. Such administrative capabilities have long been viewed as foundational to effective bureaucracy and state capacity (Weber, 1922; Mann, 1984; Wilson, 1989; Fukuyama, 2013; Sánchez De La Sierra, 2020; Cantoni et al., 2024).[6]

[5]CoNaFiL's oversight role has been described as transparent and effective in prior work (Bedasso, 2021a; KfW Development Bank, 2021; Riedl and Dickovick, 2014; Behanzin et al., 2018).

[6]These administrative capabilities are also emphasized by policymakers and international organizations as core components of effective government. For example, the World Bank's *Government Effectiveness 2025* framework highlights administrative systems, public financial management, procurement, and organizational capability as central dimensions of government effectiveness (World Bank, 2025). They are also closely related to the management practices measured by the World Management Survey, which emphasizes systematic documentation, performance monitoring, and the review of organizational performance (Bloom et al., 2013, 2019).

Our primary outcome is a municipality-year panel of annual audit scores assigned to each municipality, expressed as a percentage of the maximum attainable score. 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 improved municipal performance. After deliberation, municipal audit scores increase by 6.8 percentage points (8.7 percent relative to the control mean; randomization inference p-value = 0.026) in treated municipalities relative to control municipalities. The improvements are broadly distributed across treated municipalities rather than driven by a small number of outliers, and are larger among municipalities with lower baseline performance.

Beyond the aggregate audit score, we digitize the detailed records in the audit reports on each municipality's financial management, procurement, project implementation, and investment allocation. Aggregating these into a single index, deliberation raised treated municipalities' public financial management by 1.68 standard deviations relative to the control group (randomization inference p-value = 0.000). These effects emerge only after the intervention and remain positive throughout the three-year post-treatment period that we observe. Treated municipalities were 40 percentage points less likely to exhibit any financial irregularity (randomization inference p-value = 0.003) and 37 percentage points less likely to undertake ineligible projects (randomization inference p-value = 0.034); they moved procurement away from less competitive procedures; they strengthened project implementation, cutting work stoppages on stalled projects and advancing new ones; and they reallocated investment toward service-delivery sectors and away from their own administration. Together, these findings indicate that providing bureaucrats with aggregate organizational performance information and a structured forum for collective deliberation generated broad-based improvements across multiple core administrative functions. These gains did not come from changes in formal incentives or from additional resources: treated municipalities received and committed no more funding than the control group.

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 bureaucrats to revise their beliefs about the organization downward. On a combined index of beliefs about corruption and rule enforcement—for which lower values denote more perceived corruption and

weaker enforcement—bureaucrats in treated municipalities score 0.4 standard deviations lower than the control group (randomization-inference p-value=0.006). 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, like workers, revised their beliefs about the organization downward, coming to perceive more corruption and weaker rule enforcement. We next examine whether these changes were accompanied by shifts in internal organizational interactions. Excluding the deliberation sessions themselves, bureaucrats in treated municipalities report attending fewer routine meetings. At the same time, meetings become substantially more likely to include discussions of audit findings and corruption, with no comparable changes for other topics. Together, these results suggest that deliberation altered the focus of internal organizational interactions, shifting attention towards accountability-related issues.

We next examine the effects of deliberation on interpersonal trust among bureaucrats. We find that deliberation erodes trust within the bureaucracy, with trust falling by 0.35 standard deviations in the treated group relative to control (randomization inference $p = 0.042$). 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.

Overall, our findings imply that information about aggregate organizational performance is an important lever for improving bureaucratic effectiveness. Such information is produced routinely—through audits, inspections, and budget reviews—but typically for external reporting and oversight rather than for use within the organization itself. We show that channeling this existing information into collective deliberation improves performance across multiple core administrative functions, at low cost and without changes to formal incentives. At the same time, greater awareness of organizational shortcomings can erode interpersonal trust, indicating that performance and trust need not move together.

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

We contribute most directly to work on state capacity and bureaucratic performance

(Besley and Persson, 2009, 2010; Bardhan, 2016; Acemoglu et al., 2015; Pepinsky et al., 2017; Finan et al., 2017; Page and Pande, 2018; Besley et al., 2022; Muralidharan, 2024). While much of this work has focused on incentives, selection, monitoring, and political accountability, we instead focus on changing beliefs about aggregate organizational performance.[7] Relative to the two informational margins discussed above, we show that information about the performance of the organization as a whole, when made public and discussed collectively within the bureaucracy, can change beliefs about how the organization functions and improve performance.

Within the state capacity literature, work on audits and accountability has shown that providing audit information publicly can 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 within the bureaucracy and discussed collectively, it can reshape 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; Mastrorocco and Teso, 2023; Fenske et al., 2023; Angelucci et al., 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 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). Existing work largely views trust as supporting organizational performance and state capacity (Brown et al., 2015; Keefer and Vlaicu, 2024; Nguyen, 2026; Lowes and Montero, 2021; León-Ciliotta et al., 2025; Besley and Dray, 2024), and work linking information flows to trust also finds a positive relationship (Fisman and Khanna, 1999). We instead show that deliberation can improve organizational performance while reducing trust among bureaucrats, consistent with theoretical accounts of how collusion can sustain low-effort equilibria (Tirole, 1986; Milgrom and Roberts, 1988).[8]

[7] Within 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).

[8] Theoretical models in economics illustrate how informal arrangements can sustain suboptimal equilibria (Basu, 1986; Dasgupta, 2005; Bloch et al., 2007; Alesina and Giuliano, 2014). This aligns with foundational

Our paper also 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; Enikolopov and Zhuravskaya, 2007; 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 beliefs about aggregate performance 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).

Finally, the paper 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; Beath et al., 2017; Arora et al., 2023). 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. Recent work studies deliberation as part of broader interventions designed to shift low-effort norms among public school teachers (Ferraz et al., 2023). Our intervention instead focuses on deliberation over objective information about the organization's own performance, and shows that it improves bureaucratic performance while substantially reshaping the organizational environment within the bureaucracy.

work in sociology and political science highlighting the potential downsides of social capital and informal relationships (Banfield, 1958; Gambetta, 1993; Portes and Landolt, 1996). Empirically, recent evidence from outside organizational settings demonstrates that trust is not unconditionally welfare-improving: Butler et al. (2016) find that excessive individual trust increases exposure to exploitation, while Marx et al. (2021) show that an electoral mobilization campaign successfully increased turnout but unexpectedly degraded institutional trust. We complement this literature by documenting these dynamics within the internal organization of a state bureaucracy. We show that sharing aggregate performance information improves organizational outcomes but erodes interpersonal trust, demonstrating that in weakly institutionalized settings, performance gains can require disrupting informal relational arrangements.

2 Context and research design

2.1 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 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.

Alongside the formal administrative hierarchy, bureaucrats develop interpersonal relationships with councilors and colleagues within and across departments through repeated workplace interactions. These relationships form an important part of the organizational environment in which bureaucrats operate. However, across the control municipalities, higher interpersonal trust—a proxy for relational arrangements (Macchiavello and Morjaria, 2022)—is associated with lower municipal audit scores (-0.55, p-value < 0.1) and poorer citizen experiences with government (-0.53, p-value $= 0.14$), suggesting that, in this setting, stronger interpersonal trust may coexist with weaker accountability and lower organizational performance.

Municipal audits. The National Commission for Local Finance (CoNaFiL), in collaboration with the Office of the Auditor General, conducts yearly audits across all municipalities. It is widely regarded as an important oversight institution with substantial technical capacity and authority (Riedl and Dickovick, 2014; Behanzin et al., 2018; Bedasso, 2021a). 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).

Audits are based on a municipality's utilization of central government's FADeC funds disbursed annually and are therefore conducted at the end of each administrative year (January–December).[9] Auditors evaluate the management of municipal funds, focusing on record-keeping, transparency in public procurement, and the quality of administrative management; these components form the municipal performance score that is our primary outcome. Details of the criteria used are provided in Appendix Subsection D.4. Beyond this summary score, the reports document in detail how each municipality managed its FADeC funds—whether spending was flagged as a financial irregularity or directed to an ineligible project, how competitively contracts were procured through open rather than restricted procedures, whether funded projects were completed or left stalled and abandoned, and how investment was allocated across sectors such as education, health, and the municipality's own administration.[10]

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 allocations in subsequent FADeC transfers.[11]

Yet this oversight architecture has not eliminated persistent weaknesses in local governance. Existing evidence points to limited enforcement of rules governing access to public information and public procurement procedures (Imbert et al., 2018; Ch et al., 2019; Canen

[9]To 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).

[10]The reports also assess council-level rule compliance, such as adherence to budget-adoption and financial-reporting requirements, which do not directly pertain to municipal bureaucratic performance.

[11]Transfers from FADeC are categorized into two types: general-purpose grants and investment grants. The general-purpose grant is allocated to cover recurrent expenditures, while the investment grant is designated for funding local public goods and services. 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. Municipalities that perform well receive larger transfers in subsequent years, as outlined in the FADeC Procedural Manual. This indicates that municipalities face meaningful financial incentives to perform well on the evaluated criteria.

et al., 2023), the influence of private vendors in local procurement, and broader concerns about state capture (Ch et al., 2019; Canen et al., 2023). Municipalities also face substantial coordination and information frictions across layers of the bureaucracy, contributing to low budget execution: between 2015 and 2018, municipal budget execution averaged only 60 percent, falling to 35 percent for capital expenditures (Bedasso, 2021a).

These frictions are compounded by the absence of forums to act on the information the audits produce. Despite the availability of detailed audit reports and recommendations, municipalities lack standardized or institutionalized processes for collectively interpreting audit findings or coordinating responses across levels of the bureaucracy. While some municipalities establish ad hoc committees or internal processes to follow up on audit recommendations (KfW Development Bank, 2021; Bedasso, 2021a), these practices are uneven and not embedded within a broader deliberative process.

2.2 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 1).

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

Figure I describes the timeline of the experiment. In August 2016, we compiled the information 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.

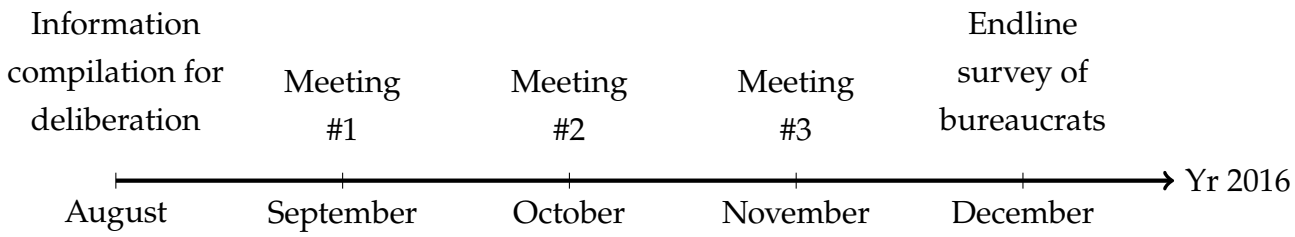


Figure I: Experiment timeline

Generating aggregate performance information for deliberation sessions. Performance information for the deliberation sessions was generated from three sources: existing municipal audit reports, a representative survey of bureaucrats, and a representative survey of citizens.

For the survey of bureaucrats, a representative sample of 30 bureaucrats was drawn from each municipality.[12] The survey included questions on knowledge of rules and

[12]In five municipalities we were not able to survey 30 bureaucrats. These include: Gogounou (26),

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.

Survey responses were then aggregated to generate organization-level performance information for the deliberation sessions. Citizen survey responses were aggregated at the municipality level. Bureaucrat survey responses, by contrast, were aggregated at the level of the full experimental sample rather than individual municipalities or officials. Bureaucrats were informed of this at the time of data collection to mitigate concerns about social desirability and encourage candid responses. Audit reports, which were already produced at the municipality level through CoNaFiL's existing audit system, required no additional aggregation.

The resulting information was presented as slide-based summaries of key indicators and findings rather than as individualized assessments or selective discussions of failures or misconduct. In the audit-based meetings, information was organized along administrative and financial dimensions and presented as observations and associated risks identified by auditors. In the survey-based meetings, results were presented as summary statistics (e.g., the share of respondents reporting specific knowledge, perceptions, or experiences). Across all meetings, the emphasis was on factual, organization-wide information rather than evaluations of individual officials. Examples of the presentation slides are provided in Appendix C.

Following the presentation of the performance information, participants engaged in an open discussion.

Meeting #1: Deliberation on audit reports and recommendations. The first meeting, held in September 2016, focused on observations from the audit reports and the auditors' recommendations. On average, 31 participants attended, of whom 22 were bureaucrats

Houeyogbe (24), and Pehunco (29) from the control municipalities and Grand-Popo (25) and Toucountouna (19) from the treatment municipalities.

(see Table 1). 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 eight contributions or suggestions. The number of criticisms and questions was also higher than in the second meeting, which focused on bureaucrat survey results. Enumerators assessed the meeting as both interactive and informative, and reported no tensions.

Meeting #2: Deliberation on the experiences of bureaucrats. The second meeting, held in October 2016, focused on results from the bureaucrat survey conducted across the 20 municipalities. On average, 32 participants attended, of whom 24 were bureaucrats (see Table 1; Figure 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 1; 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 Figure B.3, presentation focused primarily on citizens' trust in the state and corruption, citizens' participation and demand for governance. Figure B.4 shows that the discussion centered largely on citizens' trust, corruption and government performance.

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.

2.3 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. Standard errors are clustered at the municipality level. Overall, covariates are well balanced across treated and control municipalities at conventional significance levels.[13]

Compliance and take-up. Deliberation meetings were held only in treated municipalities; no such activities took place in control municipalities. Bureaucrats are rarely transferred across municipalities, and cross-municipality attendance was minimal: 5 of 296 surveyed control bureaucrats (1.7 percent) reported attending at least one meeting in a treated municipality.[14] In treated municipalities, 173 of 299 surveyed bureaucrats (57.9 percent) attended at least one of the three meetings.[15] Because take-up was partial in treated municipalities and a few control bureaucrats attended, all survey-based analyses in Section 5 report intention-to-treat (ITT) estimates; these capture the effect of assignment to deliberation at the observed take-up of roughly 58 percent. The ITT is the relevant policy parameter, since participation in deliberation is inherently voluntary. This distinction does not arise for our main performance results, which are based on the municipality-level audit panel (Section 4), where compliance is complete at the level of treatment assignment.

Researcher interference. Our main analysis does not rely on observations collected during the deliberation sessions themselves. Nevertheless, to minimize researcher interference with the deliberative process, discussions were not audio- or video-recorded, allowing interactions to proceed as naturally as possible in the field (Levitt and List, 2007;

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

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

[15]Of the 299, 116 (38.8 percent), 118 (39.5 percent), and 120 (40.1 percent) attended the first, second, and third meetings, respectively.

Spradley, 1980). Enumerators attended meetings only to record attendance and basic meeting information before the sessions began and completed structured observation protocols after the meetings had concluded. They did not participate in or monitor the discussions, and participants were informed that their interactions would not be evaluated in real time.

Social desirability bias and experimenter demand. A potential concern is that deliberation changed how bureaucrats reported their beliefs and trust rather than any real changes to these variables. 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.

3 Data

Our analysis draws on two main data sources. The first is the set of annual audit reports (2014-2019) produced by CoNaFiL auditors, which provide both an overall performance score and detailed records of each municipality's financial management. The second is an endline survey of bureaucrats conducted in December 2016, one month after the intervention, which we use to examine changes in the organizational environment.

3.1 Audit Data

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, such as record keeping, transparency in public procurement, and overall administrative management, which capture key aspects of bureaucratic performance within the municipal administration.

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

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.1 shows that at baseline (2014), the correlation is 0.62 ($p = 0.01$). This pattern suggests that the audit scores capture meaningful variation in administrative performance and service delivery.

Public financial management. The same audit reports document, in detail, how each municipality managed its FADeC funds, allowing us to look beyond the aggregate score at the specific administrative functions. We digitize from the reports municipality-year measures in four domains—financial irregularities, procurement, project implementation, and investment composition—each summarized by a standardized index following Anderson (2008) and normalized to the control group. Aggregating the outcomes within a domain into a single index increases statistical power and mitigates the multiple-testing concerns that arise when many outcomes are examined separately (O'Brien, 1984; Anderson, 2008; Kline and Santos, 2012). We define the measures in each domain, and the construction of each index, in turn.

Financial irregularities. The auditors flag two kinds of misuse of FADeC funds—financial irregularities (expenditures or payments they find to contravene the rules governing FADeC funds) and spending on projects ineligible for FADeC financing. From these we construct three measures: an indicator for whether any financial irregularity is flagged; the flagged value as a share of the FADeC transfers the municipality received; and an indicator for whether the municipality undertook any ineligible project. The *irregularity index* combines these three measures in their natural direction, with none reverse-scored, so that *higher* values denote *more* irregularities and a reduction in irregularities appears as a fall in the index.

Procurement. The reports record the procedure used to award each contract. We distinguish *open tender* (*appel d'offres ouvert*), which is fully competitive, from two more restrictive procedures: *restricted tender* (*appel d'offres restreint*), open only to firms the administration has pre-selected, and *restricted consultation* (*consultation de fournisseurs*), in which it solicits quotes from a small, self-chosen set of suppliers without open advertisement. We combine the latter two into a single *restricted* category and, for each municipality-year, measure the number and share of contracts awarded through open tender and through restricted procedures, taking all contracts the municipality awards as the denominator. The two shares do not sum to one: the remaining contracts are awarded under procedures fixed by the contract's value—outside the municipality's discretion—and so are excluded from

both. The *procurement openness index* combines the open-tender share (entering positively) and the restricted share (reverse-scored), so that higher values denote more competitive procurement.

Project implementation. For each municipality-year the reports record the share of the municipality's works projects that are halted mid-execution (a work stoppage), abandoned, and under active construction (ongoing), each as a share of all such projects. The *implementation index* combines them, with work stoppages and abandonment reverse-scored and ongoing entering positively, so that higher values denote better implementation. The reports additionally distinguish contracts carried over from earlier years (*marchés non soldés*—signed previously and still unsettled at the prior year-end) from new contracts signed in the current year (*marchés conclus*), and record the same stoppage, abandonment, and ongoing shares separately for each group. For this split the denominator is the number of carried-over (or new) contracts.

Investment composition. The reports give the amount of FADeC investment the municipality allocates to each expenditure sector, which we express as a share of its total FADeC investment. We distinguish the municipality's *own administration*—the *Administration Locale* sector, comprising spending on its own town hall, offices, and administrative equipment—from *service-delivery* ("social") sectors, which comprise every other sector (education, health, water, sanitation, markets, rural and urban roads, agriculture, culture, and youth), apart from a small community-managed and unclassified residual (the PS-DCC line). The residual stays in the denominator but is assigned to neither group, so the service-delivery and own-administration shares do not sum to one. The *social allocation index* combines the service-delivery share (entering positively) and the own-administration share (reverse-scored), so that higher values denote more investment directed to citizen-facing sectors and less to the municipality's own overhead.

We combine the four domain indices into an *overall index* of public financial management, oriented so that higher values denote better performance. Because the irregularity index runs in the opposite direction (higher = more irregularities), it enters the overall index with its sign reversed, while the other three enter directly.

From the transfer tables of the same reports we additionally digitize the municipality's funding and its budget execution. *Total FADeC received* is the total FADeC investment transfer the municipality receives, allocated by the central government; we take its log. The *commitment rate* (*engagement*) is the share of the available FADeC funds the municipality legally commits during the year—the first stage of budget execution, undertaken by the mayor as authorizing officer—and the *payment rate* (*mandatement*) is the share for which

a payment order is issued and the treasury disburses the funds—the final stage, funds that actually reach projects. Both rates capture the municipality's own budget-execution capacity; the central government's role is upstream, in transferring the funds. We use them, together with total transfers, to assess whether any performance gains reflect changes in resources or in aggregate budget execution rather than in how the municipality manages and allocates its funds.

The panel covers all 20 municipalities over 2014–2019 (120 municipality-years). A few line items are not tabulated in every report—most notably, the work-stoppage line is absent from the 2014 reports and the carried-over/new split is reported only from 2016—so the corresponding outcomes are observed for fewer municipality-years. We show that any attrition is unrelated to treatment (Table A.2).

3.2 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 (2008), 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.

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.[16] As with the belief measures, these responses are aggregated

[16] Anderson et al. (2004) show that survey measures of trust are significant predictors of contribution levels

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

We also use data on reported meeting practices within the administration. 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.

4 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.

4.1 Estimation Strategy

Using municipality-year panel data on audit scores for municipality m in year t, we estimate the following two-way fixed effects regression:

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

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 is β.

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

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.

each permutation, the treatment indicator is reassigned while preserving the number of treated units, and the specification in Equation 1 is re-estimated. This generates a placebo distribution of the coefficient on the treatment × 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).

4.2 Main 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 after deliberation, the average municipal audit score in treated municipalities increases by 6.8 percentage points relative to the control group. The effects are both economically and statistically significant, representing 8.7% of the control mean (randomization inference p-value=0.026). These findings highlight the potential for low-cost institutional changes to improve performance, even in the absence of changes to formal bureaucratic incentives.[17]

Appendix Figure B.10 presents treatment effects separately for each treated municipality. The improvements are broadly distributed across treated municipalities rather than driven by a small number of outliers, with positive and statistically significant point estimates in 7 of the 10 treated municipalities. We also examine whether treatment effects differ by baseline municipal performance. Appendix Table A.3 shows larger treatment effects among municipalities with lower pre-treatment audit scores, consistent with greater scope for improvement at lower initial levels of performance. This pattern suggests deliberation may represent low-hanging fruit specifically for weaker-performing administrations, where organizational expectations are further from the performance frontier and collective information about shortcomings is more actionable.

[17]Point estimates are positive across all three underlying audit components — record keeping (9.3 pp), transparency in public procurement (5.8 pp), and overall administrative management (14.4 pp) — consistent with broad-based improvement, though individual components are estimated with less precision than the composite index (Appendix Table A.4).

4.3 Effects on Public Financial Management

To understand how deliberation improved municipal performance, we turn to the detailed audit reports. While our primary outcome is the municipality's audit-based performance score, the reports also document the underlying dimensions of administration—financial irregularities, procurement, project implementation, and the sectoral allocation of investment (all defined in Section 3.1). We estimate Equation 1 for each outcome and summarize each domain with its standardized index; Table 3 reports the results.

Deliberation improved public financial management broadly. The *overall index* of public financial management, which aggregates the four domain indices, increases by 1.68 standard deviations in treated municipalities relative to the control group (randomization-inference p-value = 0.000). The improvement is not confined to a single dimension: all four domain indices move in the direction of better performance, three of them significantly. We discuss each domain in turn.

Financial management. Deliberation sharply reduced financial irregularities (Panel A). The *irregularity index*—oriented so that higher values denote more irregularities—falls by 1.27 standard deviations (randomization-inference p-value = 0.000), a large and precisely estimated improvement. Treated municipalities are 40 percentage points less likely to exhibit any financial-management irregularity (randomization-inference p-value = 0.003), and the value of the irregularities auditors flag falls by 4.2% of FADeC transfers received (randomization-inference p-value = 0.008); both effects are large relative to their low control means and statistically significant. Treated municipalities are also 36.7 percentage points—85 percent of the pre-period control mean—less likely to spend FADeC funds on projects ineligible for FADeC financing (randomization-inference p-value = 0.034). Deliberation thus reduced both the incidence and value of financial irregularities and the use of funds on ineligible projects.

Procurement. Deliberation reduced municipalities' reliance on less competitive procurement procedures (Panel B). The share of contracts awarded through restricted procedures falls by 11.3 percentage points (randomization-inference p-value = 0.030), and the number of such contracts falls as well (randomization-inference p-value = 0.017). The share awarded through open tender moves in the same direction—up 6.4 percentage points, or 14 percent of the control mean—but is not statistically significant (randomization-inference p-value = 0.495). The *procurement openness index*, which combines these margins,

rises by a sizable 0.54 standard deviations, though the aggregate index is not significant at conventional levels (randomization-inference p-value = 0.103).

Project implementation. Deliberation strengthened the implementation of development projects (Panel C). The *implementation index* rises by 0.81 standard deviations (randomization-inference p-value = 0.011). The share of projects halted mid-execution falls by 2.2 percentage points (randomization-inference p-value = 0.030). The shares of abandoned and of ongoing projects move in the same direction—fewer abandoned and more under active construction, consistent with the index—though neither is individually significant (randomization-inference p-value = 0.118 and 0.147, respectively).

Disaggregating by contract type (Table A.5) shows the gains concentrated where the backlog was worst: among carried-over contracts, the share halted mid-execution falls by 5.2 percentage points (randomization-inference p-value = 0.044), while among new contracts the share under active construction rises by 26.9 percentage points (randomization-inference p-value = 0.002). Deliberation therefore both unblocked stalled inherited contracts and advanced new ones. Given that implementation failures and stalled public investment projects remain a central constraint on government effectiveness and state capability in many developing countries (Pritchett et al., 2013; Andrews et al., 2017; Kapur, 2020; Dasgupta and Kapur, 2020; Muralidharan, 2024), these findings indicate that deliberation strengthened municipalities' capacity to execute public investment.

Investment composition. Finally, deliberation changed *what* municipalities spent on (Panel D). The *social allocation index* rises by 0.74 standard deviations (randomization-inference p-value = 0.004). Treated municipalities direct 11.7 percentage points more of their FADeC investment to service-delivery sectors—15 percent of the pre-period control mean (randomization-inference p-value = 0.015)—and 10.8 percentage points less to their own administration, a reduction equal to 77 percent of the control mean (randomization-inference p-value = 0.005). The sector-by-sector decomposition (Table A.6) locates the increase in education, which rises by 14.4 percentage points, or 53 percent of the control mean (randomization-inference p-value = 0.040), with no significant movement in the remaining sectors. Deliberation thus tilted investment away from the administration's own overhead and toward citizen-facing services.

Funding and budget execution. These gains did not come from additional resources (Table A.7). *Total FADeC investment received* is unchanged (randomization-inference p-

value = 0.722), and treated municipalities commit no larger a share of their available budget than the control group—the *commitment rate* is flat (randomization-inference p-value = 0.627). The *payment rate*—the share of the available budget actually disbursed—is higher by 6.9 percentage points, the direction one would expect as stalled projects resume and are paid for, but it is imprecisely estimated (randomization-inference p-value = 0.224). The improvement therefore did not come from a larger budget or from committing more of it; it came from managing and allocating the same resources better—fewer irregularities, more competitive procurement, stalled projects unblocked, and investment redirected toward the social sector.

Taken together, the audit reports show that the improvement in overall performance was broad-based: deliberation sharply reduced financial irregularities, moved procurement away from restricted procedures, unblocked stalled projects while advancing new ones, and reallocated investment toward citizen-facing services. Notably, these gains were achieved without additional financial resources, changes to formal incentives, or technical assistance—simply by making existing aggregate performance information public within the bureaucracy and providing a structured forum in which it could be collectively discussed.

4.4 Persistence of Effects

To assess whether the improvements reflect short-term salience or a more persistent change in behavior, we estimate event-study specifications with a separate treatment effect for each year from 2014 to 2019:

$$y_{mt} = \alpha_m + \alpha_t + \sum_{k \neq 2016} \beta_k \left(T_m \times \mathbf{1}[t = k] \right) + \varepsilon_{mt}, \tag{2}$$

with municipality and year fixed effects and standard errors clustered on municipality level. The intervention was implemented in the final quarter of 2016 (three monthly meetings between September and November 2016), so the 2016 administrative year that auditors assess is almost entirely pre-intervention; we take 2016 as the last pre-treatment year and as the omitted reference, so each β_k measures the treated–control difference in year k relative to 2016, and the first full post-intervention year is 2017.

Figure 3 plots these estimates for the aggregate municipal audit-based performance score—the CoNaFiL score of Section 3.1, expressed as a percentage of the maximum attainable score. In the pre-treatment period (2014–2016), treated and control municipali-

ties are economically and statistically indistinguishable. Following the intervention, audit scores in treated municipalities rise in 2017 and remain elevated through 2018 and 2019; the year-by-year estimates are positive but imprecise, each identified from comparisons across 10 treated and 10 control municipalities.

Figure 4 repeats the exercise for the overall index of public financial management. The pre-treatment leads are jointly indistinguishable from zero (joint p-value = 0.59), and the effect emerges only after 2016 and persists through 2019. Each post-treatment coefficient on the overall index is positive and individually significant (2017–2019), indicating that deliberation produced a lasting rather than a transitory improvement in public financial management.

5 Effects on the Organizational Environment

We next examine how deliberation changes the organizational environment within the bureaucracy. Specifically, we study three dimensions of organizational functioning: bureaucrats' beliefs about corruption and rule enforcement, internal meetings and accountability-related discussions, and interpersonal trust within the bureaucracy. Together, these outcomes help characterize how deliberation changes the internal environment in which bureaucrats operate.

5.1 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:

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

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.

Table 4 reports the results, and Figure 5 presents them graphically. 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 (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, the private firms that win the municipality's public procurement contracts. Bureaucrats in treated municipalities are 0.47 standard deviations more likely to perceive corruption among vendors relative to the control group (randomization inference p-value=0.004). 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 (randomization inference p-value=0.071). 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.8 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.9 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 6 presents the results graphically, and Appendix Table A.10 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 7 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.11 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.11 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.

5.2 Managers' beliefs, internal meetings and accountability

We next examine whether deliberation changed managers' beliefs about organizational performance and accountability within the bureaucracy, as well as internal meetings and accountability-related discussions.[18] We estimate the same bureaucrat-level specification as in Equation 3, comparing treated and control municipalities; for beliefs, we estimate it separately for managers and non-managerial staff.

We begin by examining managerial exposure to the deliberation meetings. We classify managers as individuals serving as General Secretary or those holding director- or chief-level positions.[19] 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 1), and notably the largest improvements in audit-based performance are observed in the administrative management component of the audit, which captures internal administrative coordination and oversight by the General Secretary (Table A.4).

We then turn to managers' beliefs about organizational performance and accountability. Figures B.12–B.13 show that managers revise their beliefs downward, coming to perceive more corruption and weaker rule enforcement. The estimated effects for managers are economically and statistically significant for the aggregate beliefs index.

Finally, we examine whether these changes in beliefs were accompanied by changes in interactions within the administration. We focus on bureaucrats' reported meeting attendance and the topics discussed during meetings as measures of internal coordination and accountability within the bureaucracy. This focus reflects the institutional setting we study, where regular administrative meetings are an important mechanism for coordination and oversight. Consistent with this, whether such meetings are regularly convened

[18] Existing studies document the important role of managers in shaping organizational performance (Fenizia, 2022; Minni, 2026).

[19] A natural limitation is that managers constitute a substantially smaller share of the bureaucracy, reducing statistical power for detecting effects within this subgroup.

by the General Secretary forms a core component of the auditors' evaluation of municipal administrative performance, as discussed in Subsection 3.1.

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 difference in overall meeting attendance between treated and control municipalities. Excluding the deliberation meetings, however, bureaucrats in treated municipalities attended fewer meetings. Taken together, these findings suggest that deliberation largely substituted for routine administrative meetings rather than increasing overall meeting attendance.

We then examine whether deliberation changed what was discussed during internal meetings. While the overall number of meetings did not increase, Figure 8 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 other administrative matters (Figure 9). This pattern suggests that deliberation shifted attention toward accountability-related concerns and organizational failures.

We next examine whether deliberation also changed interpersonal trust and relational arrangements within the bureaucracy.

5.3 Interpersonal trust

In environments where formal incentives and monitoring are weak, bureaucratic performance may depend partly on informal relational arrangements among bureaucrats. Such arrangements can facilitate coordination and cooperation, but they may also sustain tolerance of low effort or weak rule enforcement. We interpret interpersonal trust as reflecting, in part, these informal relational arrangements within the bureaucracy (Macchiavello and Morjaria, 2022), and examine whether deliberation changed it, estimating the same bureaucrat-level specification as in Equation 3.

Table 6 reports the effects of deliberation on bureaucratic trust, and Figure 10 presents them graphically. The aggregate trust index declines by 0.35 standard deviations in treated municipalities relative to the control group (randomization inference p-value = 0.042), 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

(randomization inference p-value = 0.035), representing the largest effect. Trust in departmental colleagues declines by 0.28 standard deviations (randomization inference p-value = 0.073). Estimated effects for trust in office colleagues and city hall colleagues more broadly are similar in magnitude and direction, although less precisely estimated.

The joint pattern of results—improvements in organizational performance alongside 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 is that, in weakly institutionalized settings, interpersonal trust and informal relational arrangements may instead sustain low-effort organizational equilibria. Consistent with this interpretation, among control municipalities, interpersonal trust is negatively correlated with both municipal audit-based performance and citizen-reported experiences with local government, as measured using the Afrobarometer (Figures B.14a and B.14b). This perspective is closely related to work in political science and sociology highlighting the potential downsides of social capital and informal relationships (Banfield, 1958; Gambetta, 1993; Portes and Landolt, 1996), as well as to theoretical work in economics showing how informal relational arrangements can sustain collusion, weak incentives, or low-effort organizational equilibria (Basu, 1986; Dasgupta, 2005; Bloch et al., 2007; Tirole, 1986; Milgrom and Roberts, 1988).

Attrition We test for differential attrition in responses to the trust measures across treated and control municipalities. Table A.12 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.13 for completeness.

Quantile estimates. Consistent with the results for beliefs, the effects on trust are not confined to the mean. Figure 11 presents the quantile treatment effects graphically, while Appendix Table A.14 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 12 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.15–B.16). 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. 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.

5.4 Discussion

This paper shows that low-cost institutional changes within bureaucracies can generate meaningful improvements in performance, while also reducing interpersonal trust and changing bureaucrats' beliefs about organizational functioning.

We also examine whether these administrative improvements translate into better experiences for citizens, using Afrobarometer data.[20] These data do not cover all sample municipalities, and those they do cover are not observed in every year, so the analysis relies on a shorter, unbalanced panel of municipalities and we treat it as complementary evidence. Using a specification analogous to Equation 1, we estimate that deliberation improved citizens' experiences of government services by 0.17 standard deviations, a positive but imprecise estimate (randomization-inference p-value > 0.1). The gains are concentrated in the municipalities with the lowest baseline performance, where citizen experience rises by 0.29 standard deviations—the same pattern of larger effects in initially worse-performing municipalities that we find in the audit data. Despite low power, this alignment with the audit results, together with the strong baseline correlation between audit performance and citizens' experiences, is consistent with the administrative gains reaching citizens (see Appendix Section F).

[20] Afrobarometer is nationally representative but not designed to produce municipality-level estimates.

A related question is whether similar effects would have emerged from sharing aggregate organizational information only with managers, from exposing workers to the information in the absence of managerial belief updating, or from the deliberative discussion itself. Our experiment was not designed to distinguish among these channels, which we leave to future work. The experiment establishes that the intervention as a whole—aggregate organizational information made public within the bureaucracy and discussed collectively—improves performance across multiple core administrative functions while reshaping beliefs and interpersonal trust. This margin may be particularly valuable where individual performance is difficult to observe: when bureaucrats perform multiple tasks and many outputs are hard to measure, aggregate organizational information may provide a more informative and actionable basis for improvement than information on individual officials alone.

If such practices improve bureaucratic performance, why do they not arise organically? Our findings suggest one possible explanation. Deliberation improved organizational performance while simultaneously reducing interpersonal trust and increasing awareness of organizational shortcomings. If confronting collective failures is individually costly for bureaucrats because it disrupts existing workplace relationships or shared understandings, organizations may have limited incentives to adopt such practices voluntarily even when they improve collective performance.

More broadly, the findings suggest that institutional reforms need not operate solely through changes in formal incentives, monitoring technologies, or personnel selection. They also suggest that how organizations share and collectively interpret information may itself be an important determinant of state performance.

6 Conclusion

We study whether changing beliefs about aggregate organizational performance 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 gains reflected broad-based improvements in financial management, procurement practices, the implementation of development projects, and the allocation of investment toward service-delivery sectors, and 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 aggregate 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 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, Camilo García-Jimeno, and James A Robinson, "State capacity and economic development: A network approach," *American economic review*, 2015, *105* (8), 2364–2409.

_ , 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.

Andrews, Matt, Lant Pritchett, and Michael Woolcock, "Building State Capability: Evidence, Analysis, Action," *Oxford University Press*, 2017.

Angelucci, Charles, Simone Meraglia, and Nico Voigtländer, "Organizing a kingdom," Technical Report, National Bureau of Economic Research 2024.

Arora, Abhishek, Siddharth George, Vijayendra Rao, and MR Sharan, "The added value of local democracy: evidence from a natural experiment in India," Technical Report, The World Bank 2023.

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

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

Barankay, Iwan, "Rank incentives: Evidence from a randomized workplace experiment," 2012.

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.

Beath, Andrew, Fotini Christia, and Ruben Enikolopov, "Direct democracy and resource allocation: Experimental evidence from Afghanistan," *Journal of Development Economics*, 2017, *124*, 199–213.

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 Sacha Dray**, "Trust and State Effectiveness: The Political Economy of Compliance," *The Economic Journal*, 2024, *134* (662), 2225–2251.

_ **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.

Blanes i Vidal, Jordi and Mareike Nossol, "Tournaments without prizes: Evidence from personnel records," *Management science*, 2011, *57* (10), 1721–1736.

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

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

_ , **Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen**, "The New Empirical Economics of Management," *Journal of the European Economic Association*, 2019, *17* (2), 355–403.

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.

Butler, Jeffrey V., Paola Giuliano, and Luigi Guiso, "The Right Amount of Trust," *Journal of the European Economic Association*, 2016, *14* (5), 1155–1180.

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.

Cantoni, Davide, Cathrin Mohr, and Matthias Weigand, "The rise of fiscal capacity: Administration and state consolidation in the Holy Roman Empire," *Econometrica*, 2024, *92* (5), 1439–1472.

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.

Dal Bó, Ernesto, 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.

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.

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.

Dixit, Avinash, "Incentives and organizations in the public sector: An interpretative review," *Journal of human resources*, 2002, pp. 696–727.

Enikolopov, Ruben and Ekaterina Zhuravskaya, "Decentralization and political institutions," *Journal of Public Economics*, 2007, *91* (11-12), 2261–2290.

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.

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

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.

Fenizia, 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.

_ , James Habyarimana, and Stuti Khemani, "Is Absenteeism of Public Service Providers a Low-Effort Norm? Evidence from Teachers in Uganda," 2023. Working paper, draft dated September 8, 2023.

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.

Fukuyama, Francis, "What Is Governance?," *Governance*, 2013, *26* (3), 347–368.

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.

George, Siddharth and Martin Mattsson, "Public Sector Salaries and the Quality of Governance: Evidence from Frontline Bureaucrats in India," *Available at SSRN 4955273*, 2025.

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.

Hanna, Rema and Shing-Yi Wang, "Dishonesty and selection into public service: Evidence from India," *American Economic Journal: Economic Policy*, August 2017, *9* (3), 262–90.

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.

Kapur, Devesh, "Why Does the Indian State Both Fail and Succeed?," *Journal of Economic Perspectives*, 2020, *34* (1), 31–54.

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

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

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.

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

La Porta, Rafael, 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.

Lee, David S., "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 2009, *76* (3), 1071–1102.

León-Ciliotta, Gianmarco, Dijana Zejcirovic, and Fernando Fernandez, "Policymaking, Trust, and the Demand for Public Services: Evidence from a Mass Sterilization Campaign," *American Economic Journal: Economic Policy*, 2025, *17* (1), 181–215.

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.

Lowes, Sara and Eduardo Montero, "The Legacy of Colonial Medicine in Central Africa," *American Economic Review*, 2021, *111* (4), 1284–1314.

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.

Mann, Michael, "The autonomous power of the state: its origins, mechanisms and results," *European Journal of Sociology/Archives européennes de sociologie*, 1984, *25* (2), 185–213.

Mansuri, Ghazala and Vijayendra Rao, "Localizing development: Does participation work?," 2013.

Marx, Benjamin, Vincent Pons, and Tavneet Suri, "Voter Mobilisation and Trust in Electoral Institutions: Evidence from Kenya," *The Economic Journal*, 2021, *131* (638), 2585–2612.

Mastrorocco, 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," *American Economic Journal: Applied Economics*, 2025, *17* (3), 414–445.

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.

Minni, Virginia, "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs," *The Quarterly Journal of Economics*, 2026, *141* (3), 1871–1920.

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, 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.

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

Pritchett, Lant, Michael Woolcock, and Matt Andrews, "Looking Like a State: Techniques of Persistent Failure in State Capability for Implementation," *Journal of Development Studies*, 2013, *49* (1), 1–18.

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.

Sánchez De La Sierra, Raúl, "On the origins of the state: Stationary bandits and taxation in eastern congo," *Journal of Political Economy*, 2020, *128* (1), 000–000.

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., Tubingen: Mohr Siebeck, 1922.

Wilson, James Q., *Bureaucracy: What Government Agencies Do and Why They Do It*, New York: Basic Books, 1989.

World Bank, "The Worldwide Governance Indicators: 2025 Methodology Revision," Worldwide Governance Indicators 2025.

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

Notes: 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	DID Estimate	Standard Error	P-Value	Randomization Inference P-Value	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Municipal Audit-Based Performance	77.835	6.8286**	2.9555	0.032	0.026	112

Notes: 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 Improve Public Financial Management?

Dependent Variable	Pre-Period Control Mean	DID Estimate	Standard Error	p-Value	RI p-Value	Obs.
Overall Index		1.682***	0.347	0.000	0.000	120
Panel A: Financial Irregularities						
Irregularity Index		-1.273***	0.297	0.000	0.000	120
Any Financial Irregularity	0.100	-0.400***	0.105	0.001	0.003	120
Fin. Irregularity (% of FADeC)	0.905	-4.211**	1.785	0.029	0.008	120
Any Ineligible Project	0.433	-0.367**	0.148	0.023	0.034	120
Panel B: Procurement Openness						
Procurement Openness Index		0.541*	0.302	0.089	0.103	119
# Restricted	1.733	-2.877*	1.447	0.061	0.017	119
% Restricted	5.668	-11.270**	5.009	0.037	0.030	119
# Open	11.833	1.395	2.493	0.582	0.605	119
% Open	45.597	6.373	8.883	0.482	0.495	119
Panel C: Development Projects						
Implementation Index		0.812**	0.333	0.025	0.011	120
% Stoppage	0.579	-2.242**	1.040	0.044	0.030	100
% Abandoned	0.980	-1.437	0.863	0.112	0.118	120
% Ongoing	19.167	6.868	4.569	0.149	0.147	120
Panel D: Investment Composition						
Social Allocation Index		0.742**	0.267	0.012	0.004	120
Social sectors (% of investment)	80.166	11.685**	4.595	0.020	0.015	120
Local administration (% of investment)	14.088	-10.814**	4.192	0.018	0.005	120

Notes: The unit of observation is a municipality-year. The data are from the Audit Reports of 2014–2019. The *Overall Index* is a standardized aggregate (Anderson, 2008) of the four panel indices below, standardized using the control group and signed so that higher values denote better public financial management. **Panel A.** *Any Financial Irregularity* equals one if the auditors flagged any financial-management irregularity; *Fin. Irregularity (% of FADeC)* is the value of the flagged financial irregularities as a percentage of total FADeC transfers received; *Any Ineligible Project* equals one if the municipality spent FADeC funds on a project ineligible for FADeC financing. The *Irregularity Index* aggregates these three measures (higher values denote more irregularity). **Panel B.** *Open* is the Open Tender procedure and *Restricted* combines Restricted Tender and Restricted Consultation; # Open (Restricted) is the number of such contracts and % Open (Restricted) is their share of all contracts awarded that year. The *Procurement Openness Index* aggregates % Open and % Restricted (higher values denote more open procurement). **Panel C.** *% Stoppage*, *% Abandoned*, and *% Ongoing* are the shares of the municipality's works projects that are, respectively, halted, formally abandoned, and under construction. Each measured as a share of *all* development-project contracts audited in the municipality–year. The *Implementation Index* aggregates these three (higher values denote better implementation). **Panel D.** *Social sectors* is the combined share of FADeC investment allocated to all sectors other than the municipality's own administration and the PSDCC/community-managed residual (i.e., education, health, water & energy, sanitation, market infrastructure, rural paths, agriculture, culture, youth & sport, and urban roads & planning); *Local administration* is the share allocated to the municipality's own administrative infrastructure. The *Social Allocation Index* aggregates these two (higher values denote a more pro-social allocation). Control-group means are omitted for the indices, which are normalized to the control group. Randomization-inference p-value based on 1,000 repetitions is reported. All regressions include municipality and year fixed effects, and 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' Beliefs About Corruption and Enforcement of Rules?

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

Notes: The unit of observation is a bureaucrat, and each outcome is standardized to the control group, so coefficients are in standard-deviation units. The *Belief about Rule of Law Index* aggregates the corruption and enforcement sub-components below, with corruption measures reverse-coded so that higher values denote lower perceived corruption and stronger perceived rule enforcement. The corruption sub-components measure whether respondents believe that municipal councilors, other bureaucrats, and city hall vendors—the private firms that win the municipality's procurement contracts—engage in corruption (four-point scale, "not at all" to "frequently"); the enforcement sub-components measure whether respondents believe the rules governing equipment acquisition, public procurement, and staff recruitment are followed (three-point scale, "not at all" to "fully"). We report the p-value in Column (3) and the randomization-inference p-value (1,000 repetitions) in Column (4). Standard errors are clustered at the municipality level. Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Number of meetings held

	Number of City-hall Meetings attended in the last 3 months	
	(1)	(2)
	Overall	Excluding Deliberation
Treatment	-0.460	-1.610*
	(0.896)	(0.847)
	{0.598}	{0.0790}
Control Mean	6.132	6.098**
	(0.625)	(0.622)
	{0.303}	{0.0420}
N	595	595

Notes: The unit of observation is a bureaucrat. The outcome is the number of meetings the bureaucrat reports attending at the city hall in the past three months. Column (1) counts all meetings; Column (2) excludes the three deliberation meetings. Standard errors clustered at the municipality level are in parentheses; randomization-inference *p*-values (1,000 repetitions) are in square brackets. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

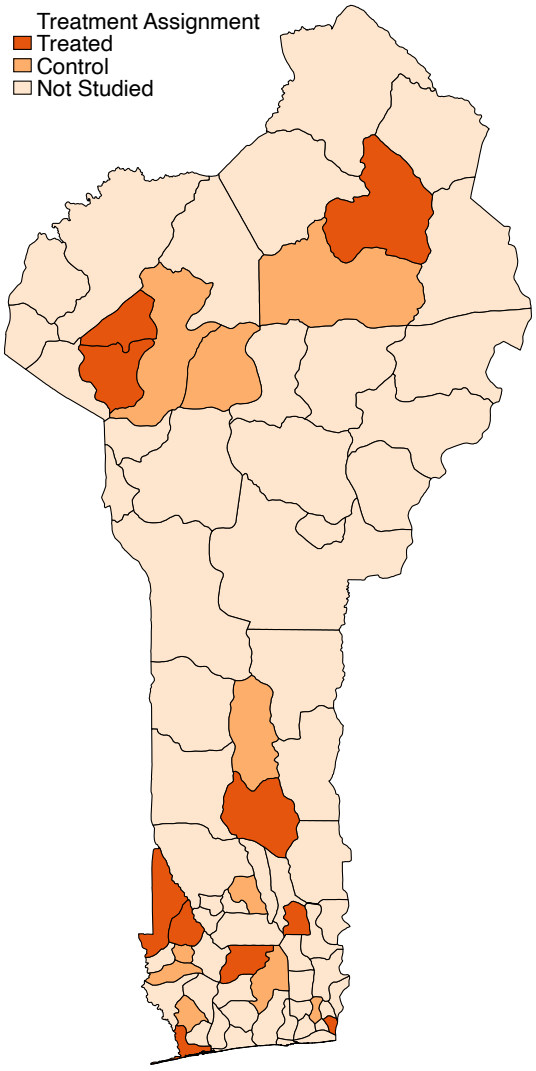
Table 6: Did Deliberation Affect Bureaucrats' Trust in Others?

Dependent Variables	ITT	Standard Error	P-Value	RI P-Value	Obs.
	(1)	(2)	(3)	(4)	(5)
Trust Index	-0.3511**	0.1585	0.039	0.042	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

Notes: The unit of observation is a bureaucrat, and each outcome is standardized to the control group, so coefficients are in standard-deviation units. The *Trust Index* aggregates trust in the municipal council, office colleagues, colleagues in the same department, and other town-hall staff, each measured on a four-point scale ("not at all" to "very much"), with higher values denoting greater trust. We report the p-value in Column (3) and the randomization-inference p-value (1,000 repetitions) in Column (4). Standard errors are clustered at the municipality level. Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

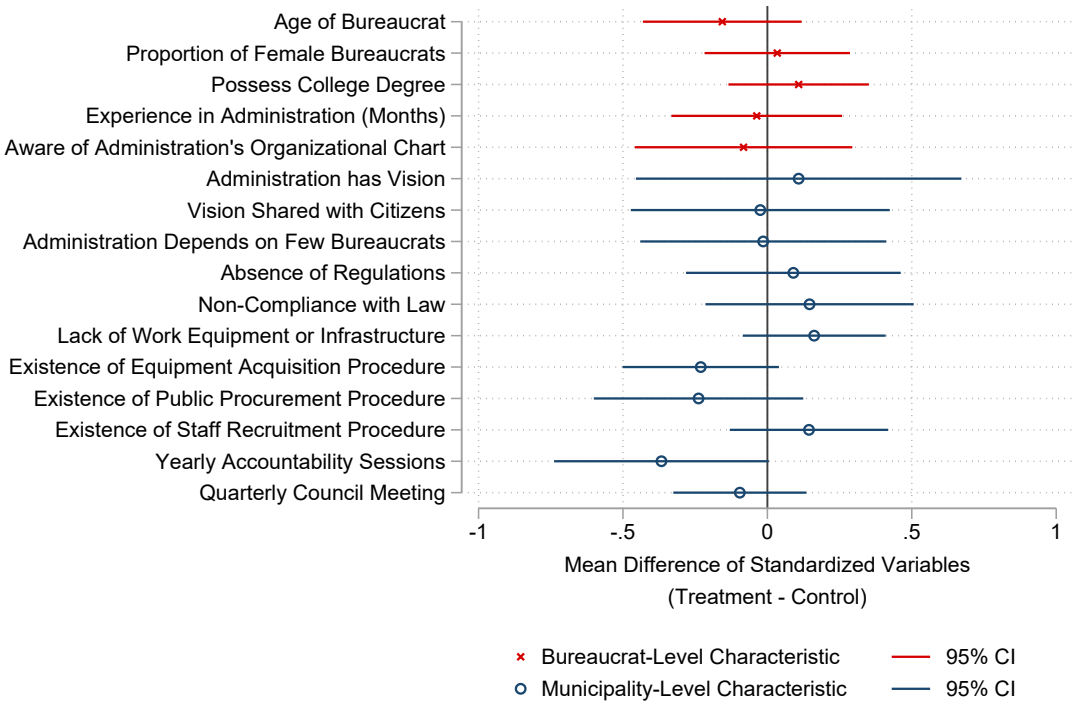
Figures

Figure 1: Spatial Distribution of Treatment Assignment



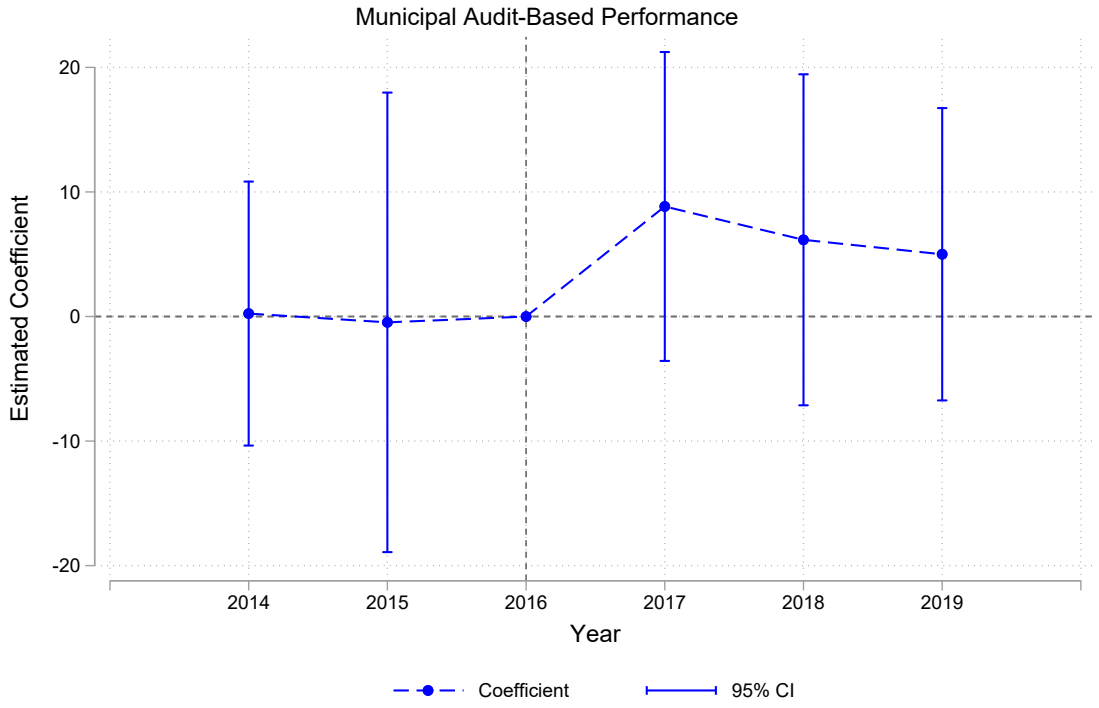
Notes: The figure shows the spatial distribution of the 20 municipalities in the sample. Shading indicates treatment and control assignment.

Figure 2: Balance test on characteristics of bureaucrats



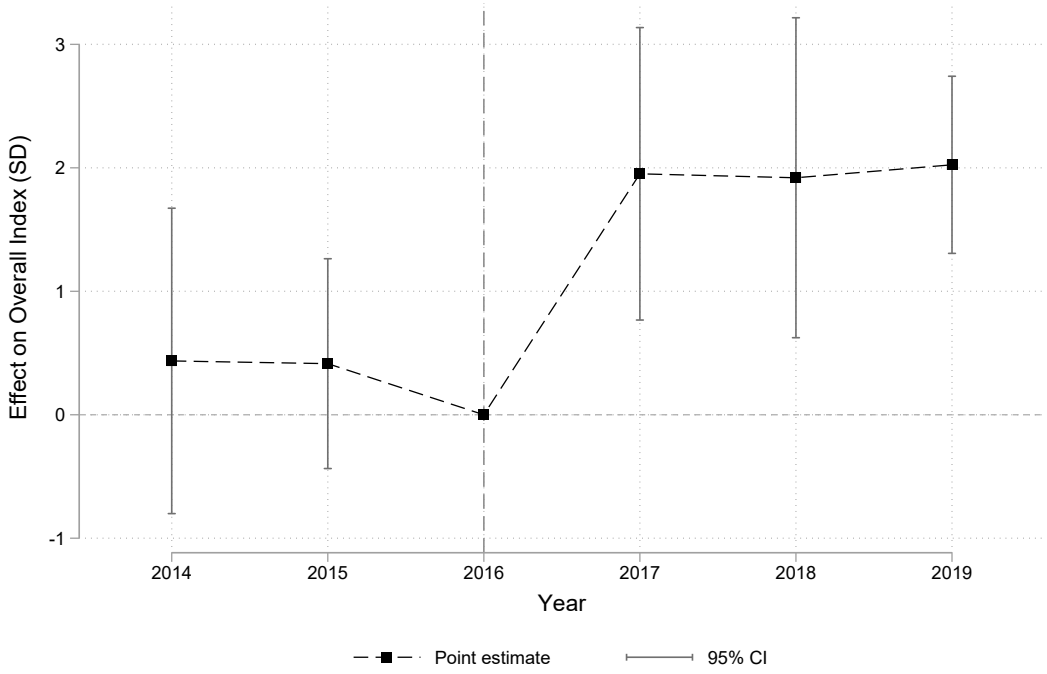
Notes: The variables have been standardized and their mean difference between treatment and control group is plotted along with 95% CIs. The standard errors are clustered at the municipality level. The source of the data is bureaucrat survey.

Figure 3: Event Study Plot of Municipal Audit-Based Performance



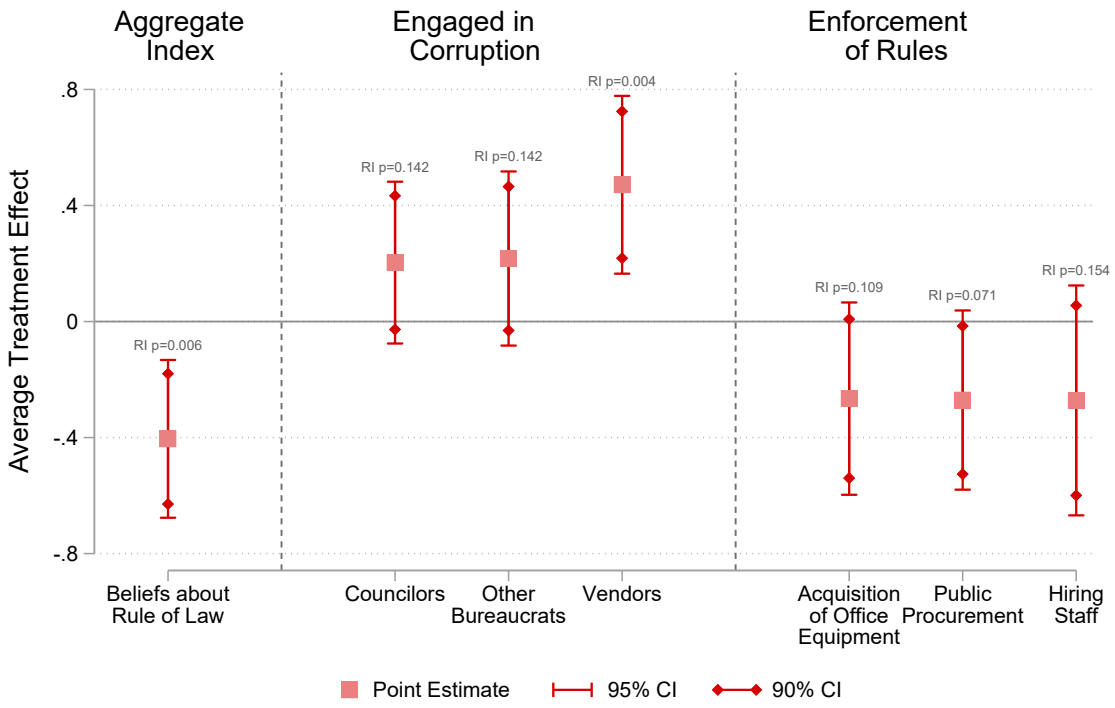
Notes: Municipal Audit-Based Performance is a score built from the audit reports of FADeC. The score is constructed from three sub-components: (i) Record Keeping (ii) Transparency of Public Procurement (iii) Overall Management. Record Keeping is constructed from scores given based on availability of 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 4: Event Study: Effect on the Overall Index of Public Financial Management



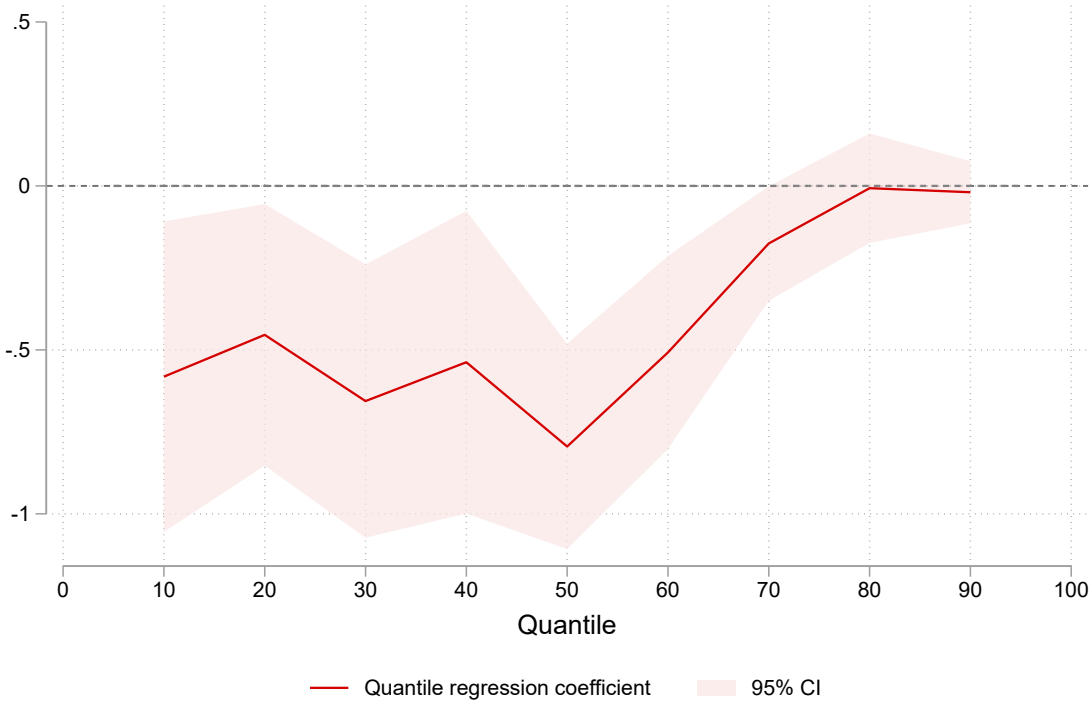
Notes: The figure plots the effect of deliberation on the Overall Index of public financial management, a control-normalised Anderson (2008) index of the four domains reported in Table 3. The unit of observation is a municipality-year and the data are from the CoNaFiL FADeC audit reports, 2014–2019. Each marker is the coefficient on Treatment $\times$ $\mathbf{1}$[Year $= k$], estimated by OLS with municipality and year fixed effects. The omitted reference year is 2016, the last pre-treatment year; the dashed vertical line marks the start of the post-treatment period. Standard errors are clustered at the municipality level and bars denote 95% confidence intervals. A joint test that the pre-treatment coefficients (2014 and 2015) are both zero yields p-value = 0.593

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



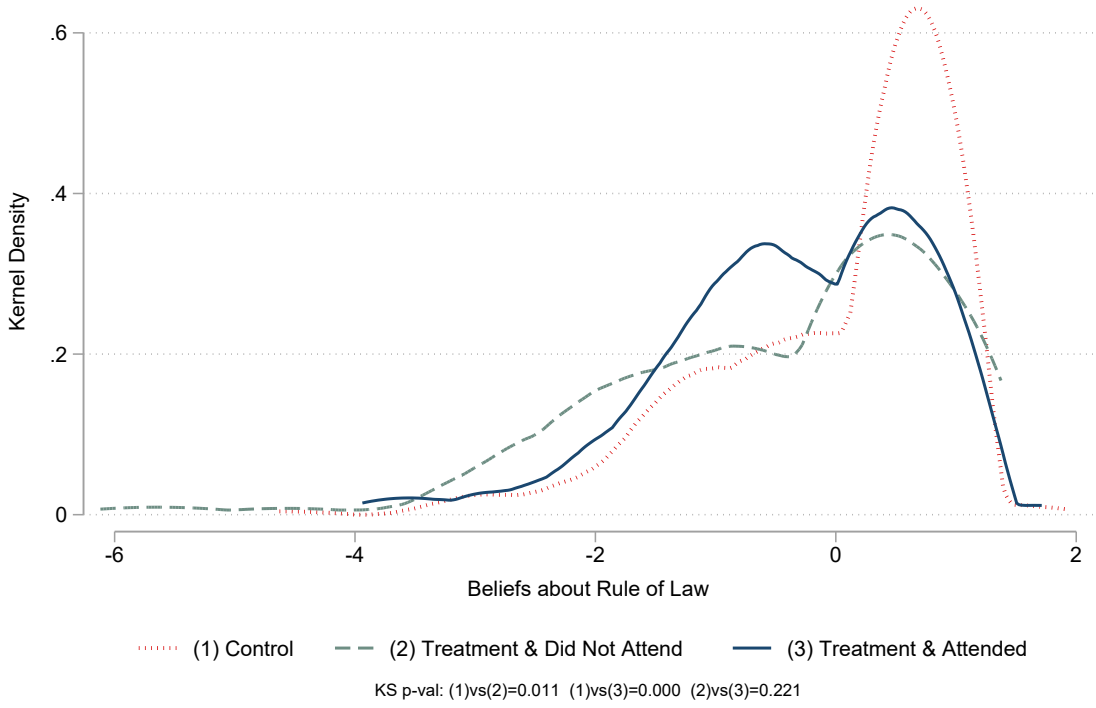
Notes: The figure reports estimates of the effect of deliberation on bureaucrats' beliefs about corruption and enforcement of rules, with 90% and 95% confidence intervals. Randomization inference p-values (1,000 repetitions) are shown above each estimate. The unit of observation is a bureaucrat. Standard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

Figure 6: Quantile Regression Estimates: Did Deliberation Affect Bureaucratic Beliefs about Corruption and Enforcement of Rules



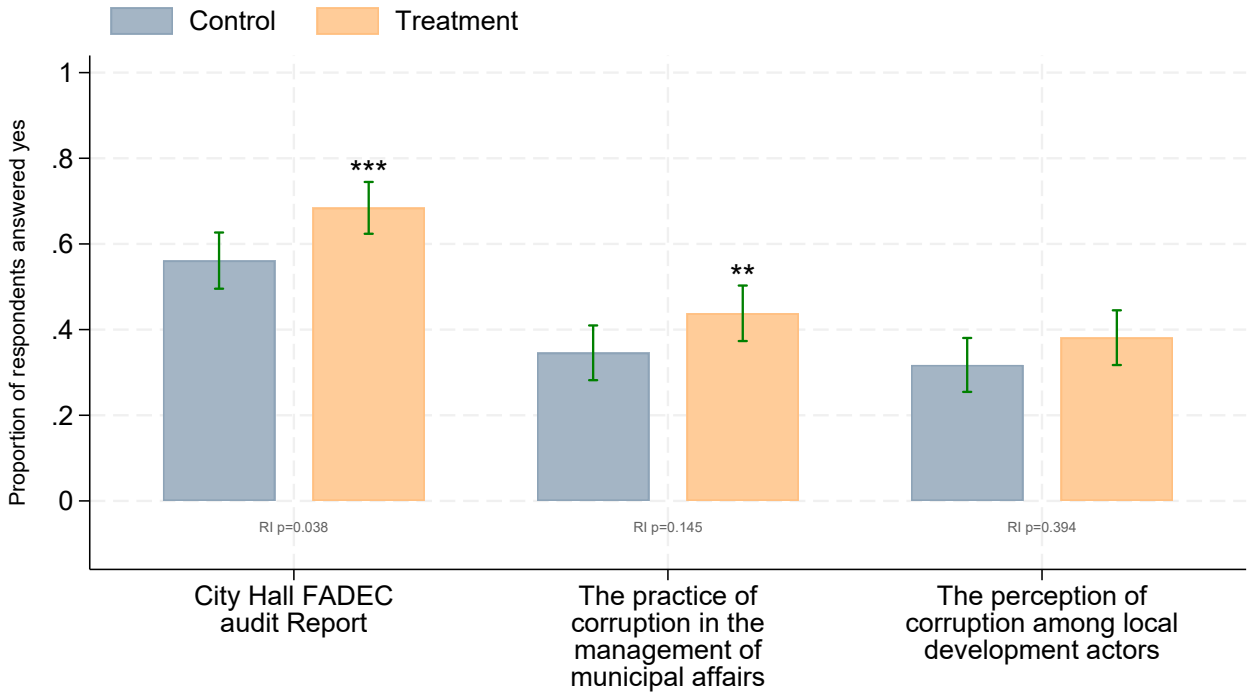
Notes: The figure reports quantile regression estimates of the treatment effect on the index of beliefs about rule of law across quantiles 10 to 90. The shaded area represents the 95% confidence interval. Standard errors are clustered at the municipality level.

Figure 7: Kernel Density of Bureaucratic Beliefs about Corruption and Enforcement



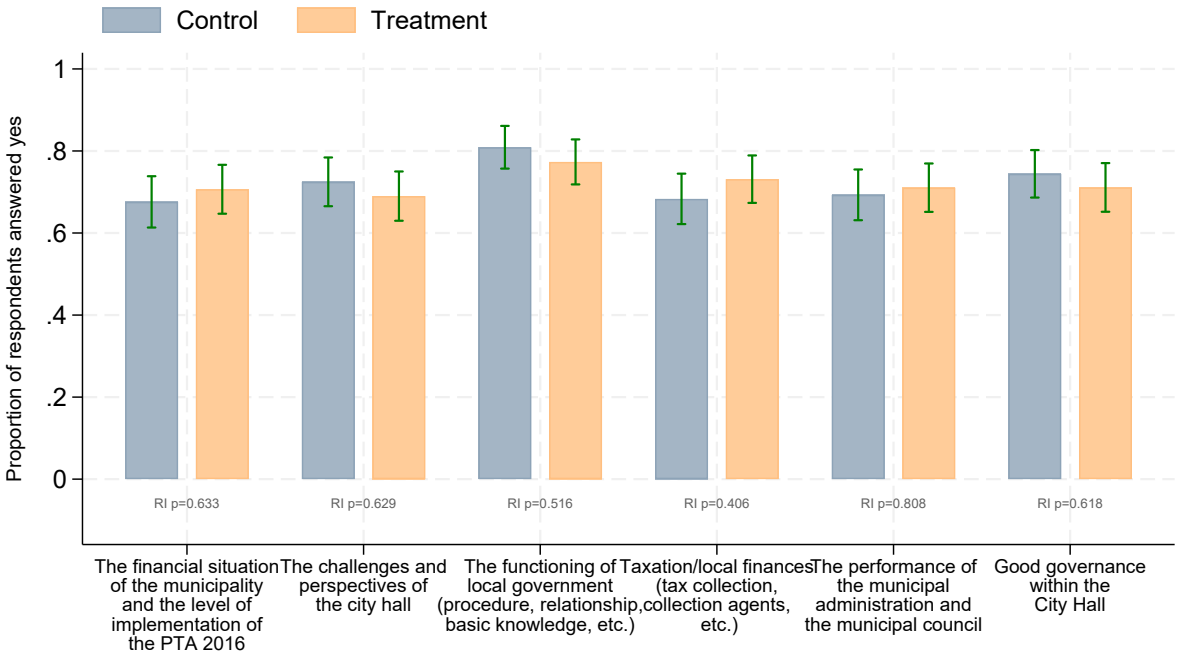
Notes: The figure shows kernel density estimates of the Beliefs about Rule of Law index by treatment and attendance status. Kolmogorov-Smirnov test p-values for pairwise comparisons are reported below the figure.

Figure 8: Topics Discussed During General Meetings in the Municipality: Audit and Corruption Topics



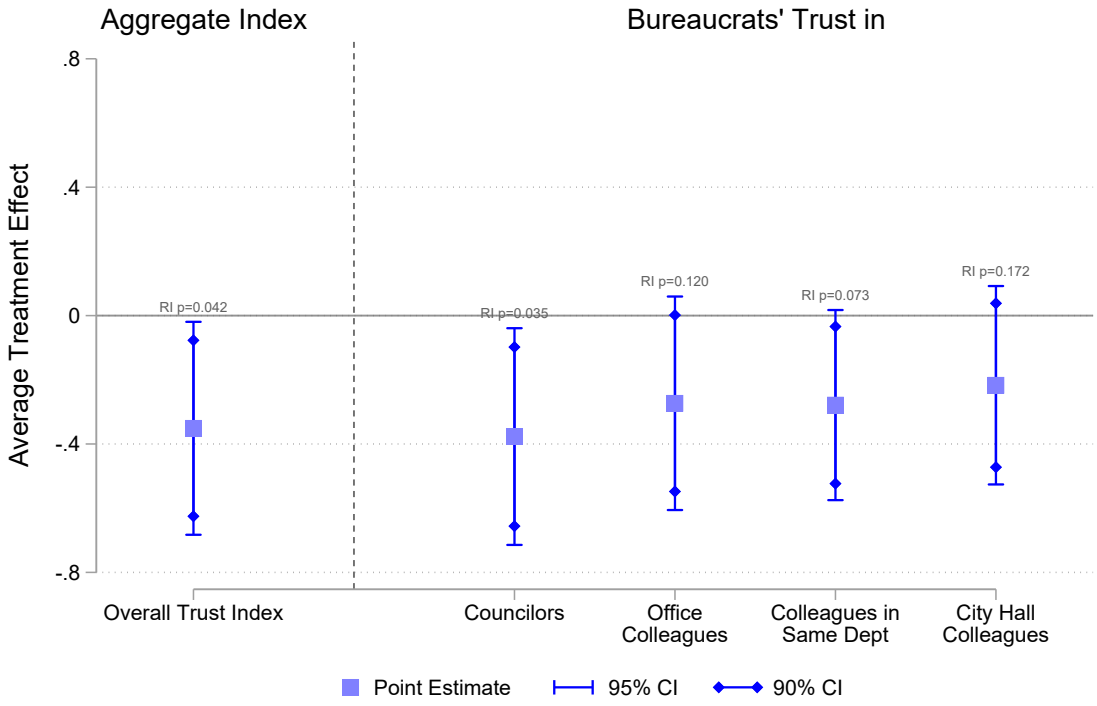
Notes: The figure shows the proportion of respondents who reported that audit and corruption-related topics were discussed during general meetings in the municipality, by treatment status. Error bars represent 95% confidence intervals. Stars denote statistical significance of the difference in means: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

Figure 9: Topics Discussed During General Meetings in the Municipality: Other Administrative Topics



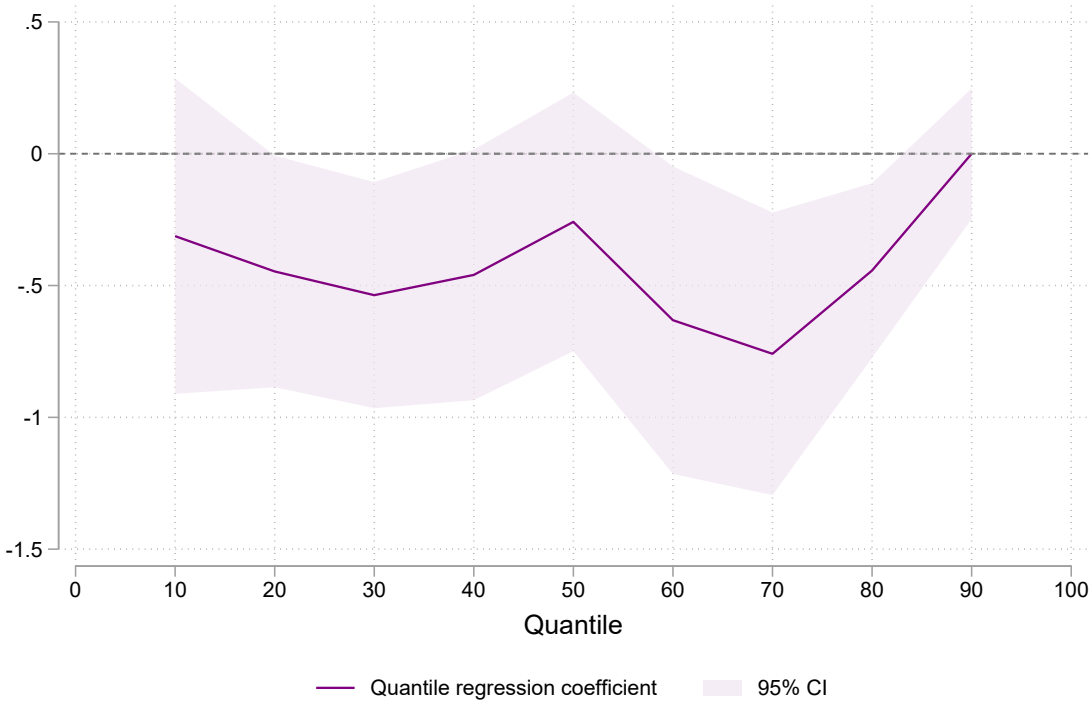
Notes: The figure shows the proportion of respondents who reported that various non-audit topics were discussed during the meetings, by treatment status. Error bars represent 95% confidence intervals. Stars denote statistical significance of the difference in means: * p<0.10, ** p<0.05, *** p<0.01. Randomization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

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



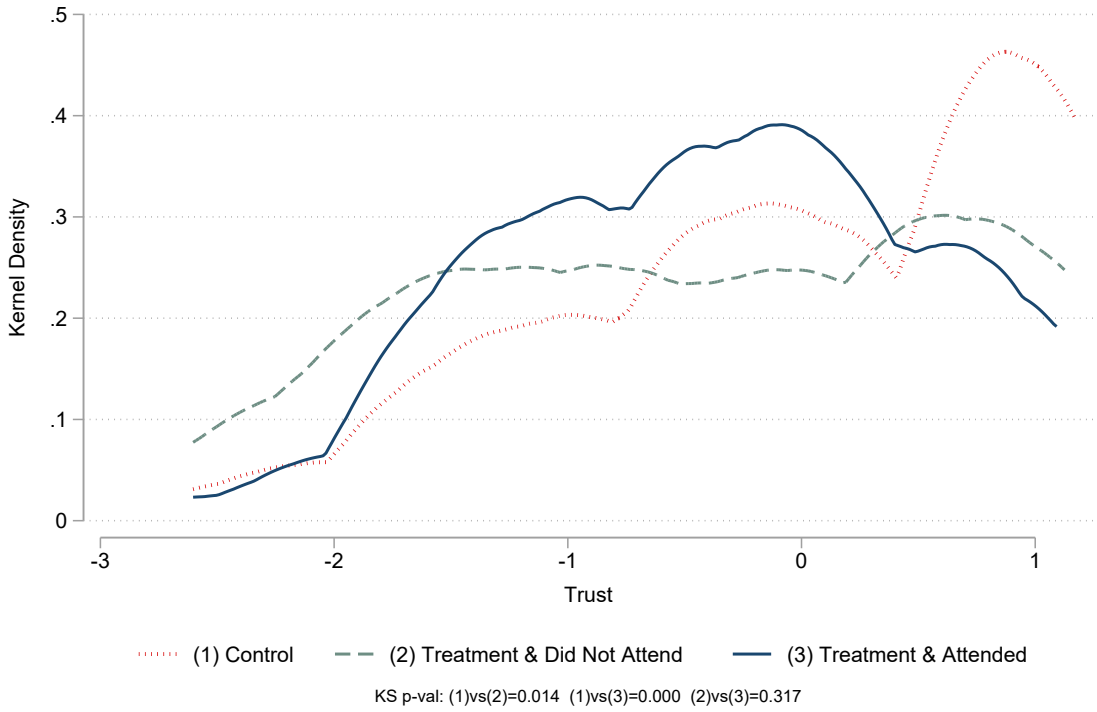
Notes: The figure reports estimates of the effect of deliberation on bureaucrats' trust, with 90% and 95% confidence intervals. Randomization inference p-values (1,000 repetitions) are shown above each estimate. The unit of observation is a bureaucrat. Standard errors are clustered at the municipality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

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



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

Figure 12: Kernel Density of Bureaucratic Trust Index



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

Appendices

A Appendix Tables **66**

B Appendix Figures **78**

C Deliberation Meeting Materials **94**

D Audit-Based Performance Measures **96**

D.1 Data Source and Digitization 96

D.2 Performance Components 96

D.3 Construction of the Performance Measure 96

D.4 Details of the criteria used by auditors 98

E Bureaucrat Survey Measures **101**

E.1 Belief Measures 101

E.2 Trust 102

E.3 Meeting Practices 103

E.4 Meeting Content 103

F Downstream effects on citizen experience **104**

A Appendix Tables

Table A.1: 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

Notes: 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.2: Attrition: Effects on Public Financial management

1(Missing)	Treatment Effect	Standard Error	p-Value	Non-Missing Obs.
% Stoppage	-0.000	0.000	–	100
% Stoppage, carried-over	-0.033	0.092	0.722	83
% Stoppage, new	0.033	0.092	0.722	83
% Abandoned, carried-over	-0.033	0.092	0.722	83
% Abandoned, new	0.033	0.092	0.722	83
% Ongoing, carried-over	-0.067	0.075	0.387	86
% Ongoing, new	-0.000	0.072	1.000	86

Notes: The unit of observation is a municipality-year. Each dependent variable is an indicator equal to one if the corresponding outcome is missing for that municipality-year. The table reports the difference-in-differences estimate of Treatment × Post on this indicator, with municipality and year fixed effects and standard errors clustered at the municipality level. For *% Stoppage* the item is absent for all twenty municipalities in 2014 and present for all of them thereafter, so missingness is fully determined by the audit year and is absorbed by the year fixed effects; the estimate is therefore exactly zero and the test is not defined.

Table A.3: Heterogeneity based on baseline audit-based performance - Did Deliberation Affect Municipal Performance?

Dependent Variables	Below Mean Municipality Control Mean	Treatment	Treatment × Above Mean	Obs.	P-Value: T + T × Above Mean = 0
	(1)	(2)	(3)	(4)	(5)
Municipal Audit-Based Performance	63.138	10.779***	-7.572***	112	0.312
		(2.605)	(2.225)		
		{0.056}	{0.942}		
Sub-Components of Municipal Audit-Based Performance					
Record Keeping	63.800	18.583*	-18.562**	112	0.997
		(9.860)	(8.692)		
		{0.208}	{0.642}		
Transparency of Public Procurement	68.238	11.061*	-10.110	112	0.848
		(5.623)	(6.057)		
		{0.148}	{0.717}		
Overall Management	65.000	39.935**	-42.201**	72	0.825
		(17.788)	(17.554)		
		{0.129}	{0.249}		

Notes: 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	DID Estimate	Standard Error	P-Value	Randomization Inference P-Value	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Record Keeping	78.493	9.3016	7.7031	0.242	0.216	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

Notes: 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: Effects on Development-Projects, by Contract Type

Dependent Variable	Pre-Period Control Mean	DID Estimate	Standard Error	p-Value	RI p-Value	Obs.
% Stoppage, carried-over	0.835	-5.155**	2.276	0.035	0.044	83
% Stoppage, new	0.668	-2.015	2.198	0.371	0.443	83
% Abandoned, carried-over	0.835	-3.241	2.559	0.221	0.264	83
% Abandoned, new	0.385	-2.848*	1.647	0.100	0.169	83
% Ongoing, carried-over	10.140	1.839	4.526	0.689	0.726	86
% Ongoing, new	31.133	26.941***	6.933	0.001	0.002	86

Notes: The unit of observation is a municipality-year. The data is from the 2014–2019 Audit Reports. The audit reports classify each works contract as *carried-over* or *new*. "Carried-over" contracts (*marchés non soldés*) were signed in an earlier year and remained unsettled at the prior year-end; "new" contracts (*marchés conclus*) were signed in the current year. Each variable is a share computed *within* a contract vintage, not over all contracts. Rows labeled "carried-over" divide the relevant count (stopped, abandoned, or ongoing) by the total number of carried-over contracts; rows labeled "new" divide by the total number of new contracts. The denominators therefore differ across rows, and from the whole-portfolio shares reported in Panel C of Table 3. The carried-over/new classification is reported in the audit reports only from 2016 onward (and for a subset of municipalities in 2015); it is not available for 2014 for any municipality. Randomization-inference p-value based on 1,000 repetitions is reported. All regressions include municipality and year fixed effects, and 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: Effects on Investment Shares by Sector

Dependent Variable	Pre-Period Control Mean	DID Estimate	Standard Error	p-Value	RI p-Value	Obs.
Own municipal administration	14.088	-10.814**	4.192	0.018	0.005	120
Education	27.464	14.421**	6.582	0.041	0.040	120
Health	4.126	1.370	2.561	0.599	0.616	120
Water & Energy	8.834	-4.057	2.448	0.114	0.126	120
Sanitation	2.630	0.272	1.003	0.789	0.776	120
Market infrastructure	12.481	3.252	3.734	0.395	0.406	120
Rural paths	14.438	-1.804	4.598	0.699	0.708	120
Agriculture, livestock & fishing	3.857	1.892	2.093	0.377	0.398	120
Culture & leisure	1.302	0.219	1.632	0.895	0.890	120
Youth & sport	0.331	0.365	0.502	0.475	0.488	120
Urban roads & planning	4.703	-4.247	2.776	0.142	0.146	120
Other (PSDCC/community-managed)	5.746	-0.871	3.494	0.806	0.813	120

Notes: The unit of observation is a municipality-year. The data is from the 2014–2019 Audit Reports. Each variable is the share of total FADeC investment allocated to the sector as a percentage. Randomization-inference p-value based on 1,000 repetitions is reported. All regressions include municipality and year fixed effects, and standard errors are clustered at the municipality level. Significance levels are denoted as: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Effects on Municipal Funding and Budget Execution

Dependent Variable	Pre-Period Control Mean	DID Estimate	Standard Error	p-Value	RI p-Value	Obs.
Payment rate (%)	57.271	6.860	5.608	0.236	0.224	120
Commitment rate (%)	78.546	2.818	5.468	0.612	0.627	120
Log total FADeC received	20.000	-0.039	0.106	0.721	0.722	120

Notes: The unit of observation is a municipality-year. The data is from the 2014–2019 Audit Reports. *Payment rate* is the share of available FADeC resources mandated for payment by the municipality and disbursed by the treasury — the final stage, money that actually reached projects. *Commitment rate* is the share of available FADeC resources the municipality legally engaged during the year — the first stage of budget execution, capturing whether the municipality (its mayor, as authorizing officer) committed the funds available to it. Both measure municipal budget-execution capacity; the central government's role is upstream, in transferring the funds. *Log total FADeC received* is the logarithm of total FADeC investment transfers, which are allocated by the central government. Randomization-inference p-value based on 1,000 repetitions is reported. All regressions include municipality and year fixed effects, and 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: 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

Notes: 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.9: Lee Bounds: Bureaucrats' Beliefs About Corruption and Enforcement of Rules

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

Notes: 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.10: 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**	-0.454**	-0.656***	-0.538**	-0.795***	-0.508***	-0.176**	-0.00687	-0.0193
	(0.241)	(0.203)	(0.212)	(0.235)	(0.159)	(0.150)	(0.0889)	(0.0850)	(0.0483)
	{0.0570}	{0.0280}	{0.00400}	{0.00700}	{0.0150}	{0.0140}	{0.0250}	{0.0490}	{0.110}
Observations	580	580	580	580	580	580	580	580	580

Notes: 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 6. 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.11: 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**	-0.177	-0.574**	-0.074	-0.247	0.006
	(0.154)	(0.202)	(0.241)	(0.355)	(0.308)	(0.251)
	[0.015]	[0.386]	[0.026]	[0.838]	[0.444]	[0.980]
Constant	1.712***	1.712***	2.703***	2.703***	2.237***	2.237***
	(0.121)	(0.121)	(0.175)	(0.175)	(0.144)	(0.144)
	[0.008]	[0.177]	[0.007]	[0.399]	[0.238]	[0.521]
Observations	20	20	20	20	20	20

Notes: 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.12: 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

Notes: 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.13: Lee Bounds - Trust

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

Notes: 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.14: Quantile Regression Estimates: Did Deliberation Affect Bureaucrats' Trust in Others

Quantile	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
Treatment	-0.312	-0.446**	-0.537**	-0.460*	-0.258	-0.631**	-0.759***	-0.443***	0
	(0.305)	(0.224)	(0.218)	(0.242)	(0.250)	(0.297)	(0.273)	(0.169)	(0.126)
	{0.243}	{0.135}	{0.0910}	{0.0860}	{0.127}	{0.0480}	{0.0570}	{0.0250}	{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 11. 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

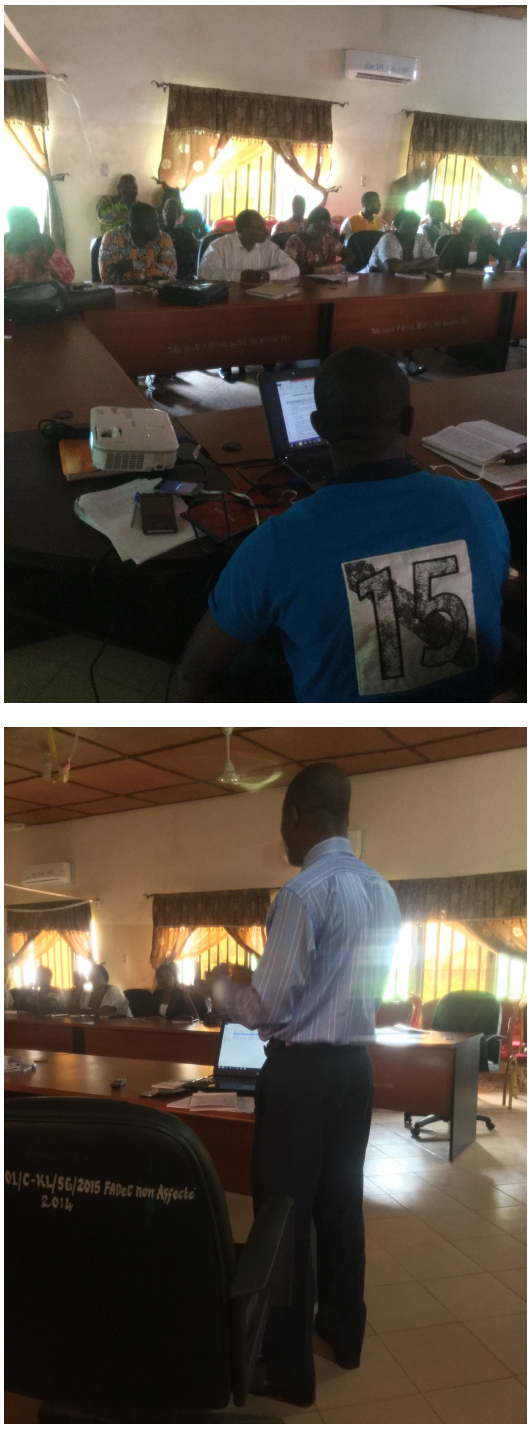
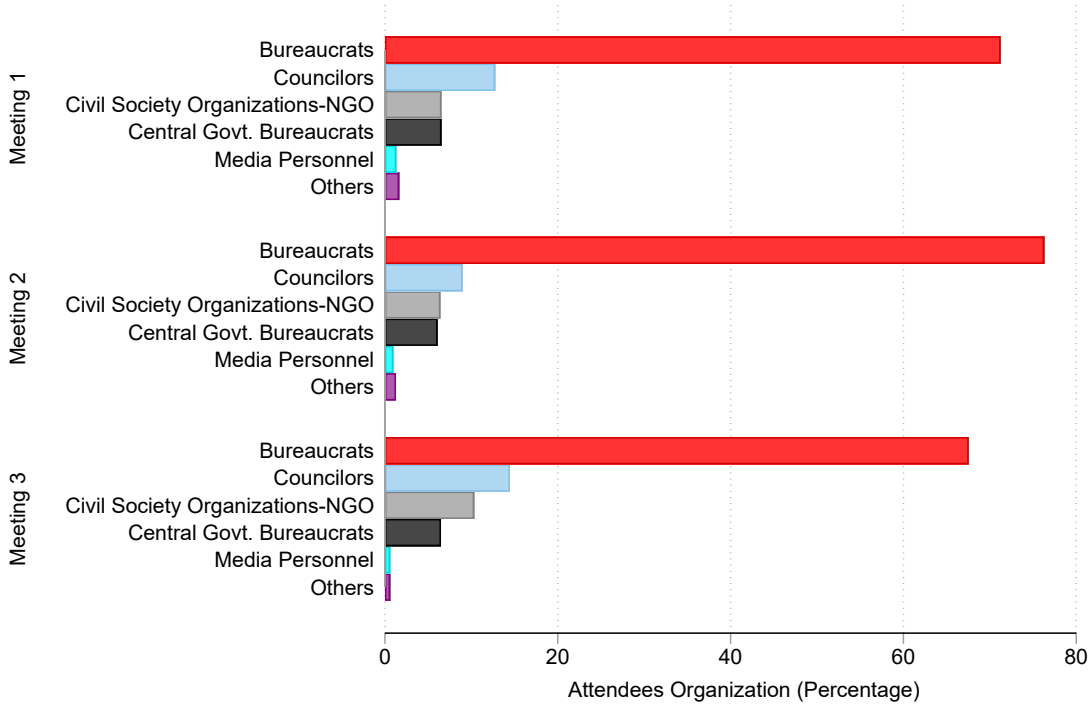
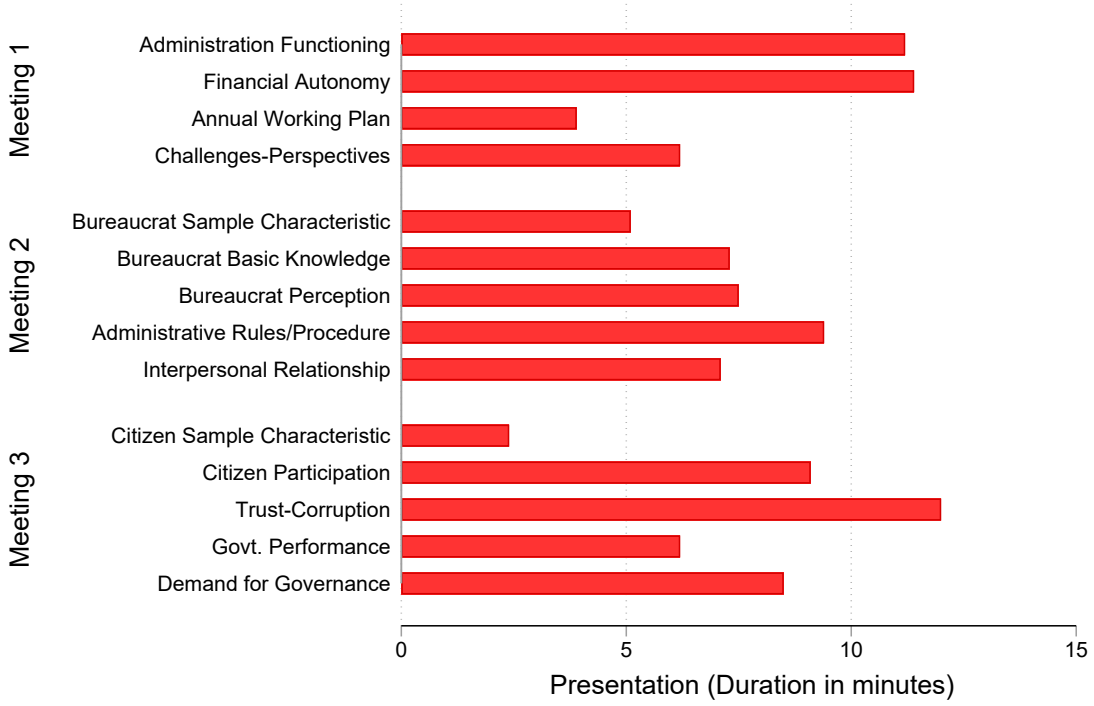


Figure B.2: Attendance by Types of Actors in the Collective Deliberation Meetings



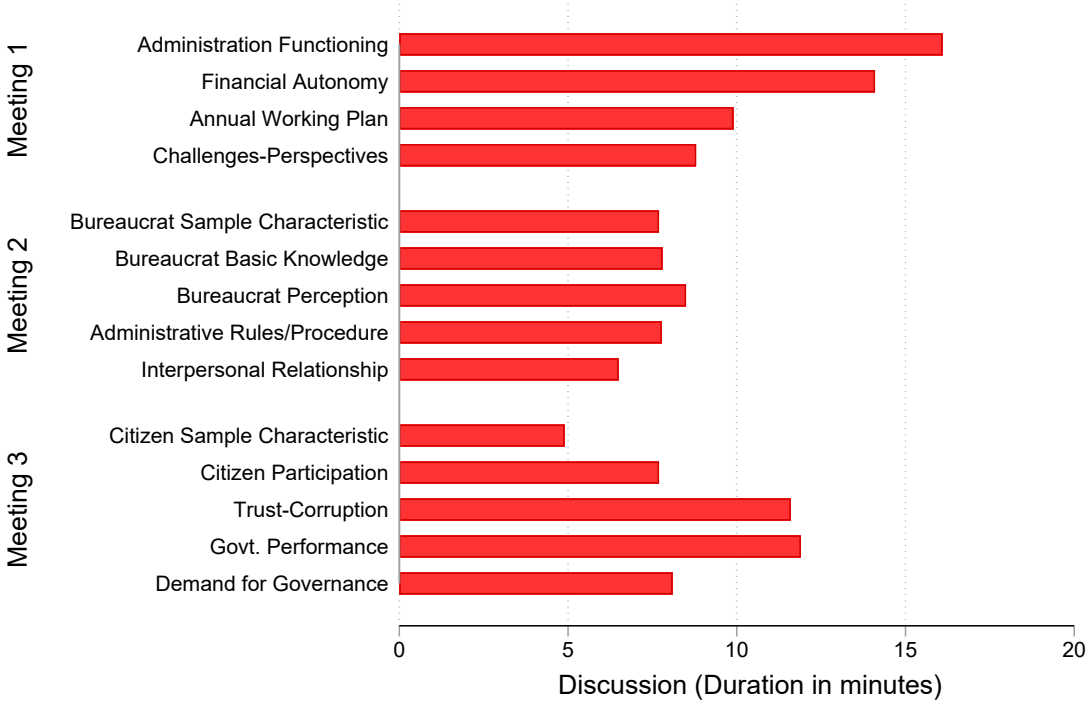
Notes: Data are collected by enumerators before the start of the meeting.

Figure B.3: Presentation topics and their duration



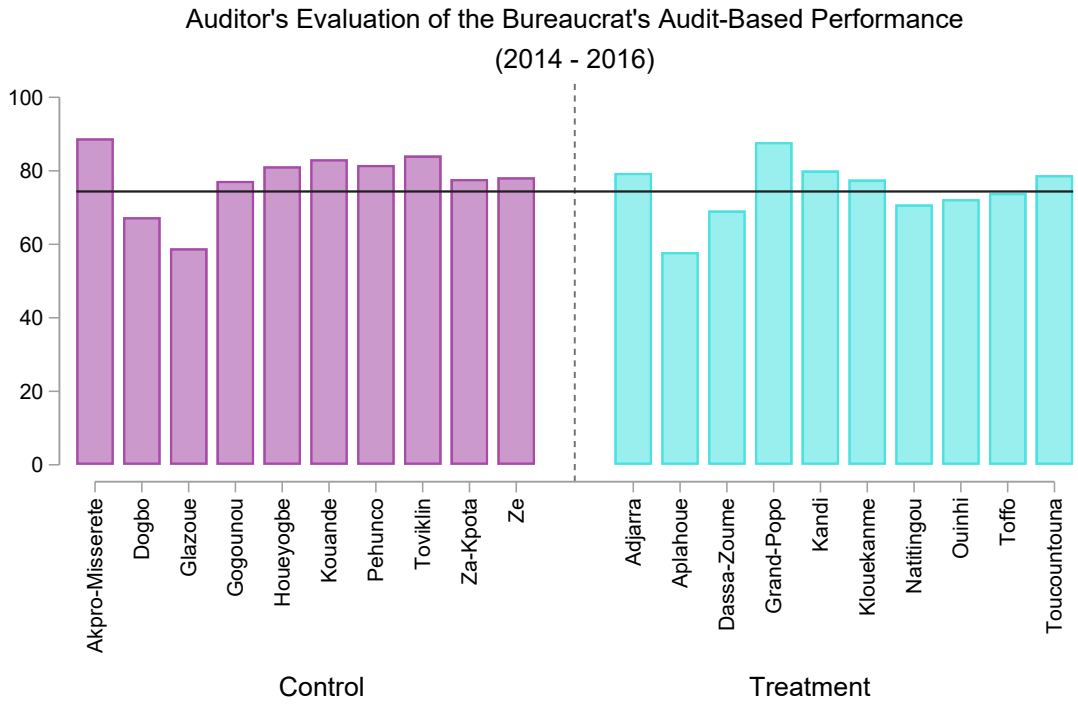
Notes: Data are from meeting notes taken by enumerators.

Figure B.4: Discussion topics and their duration



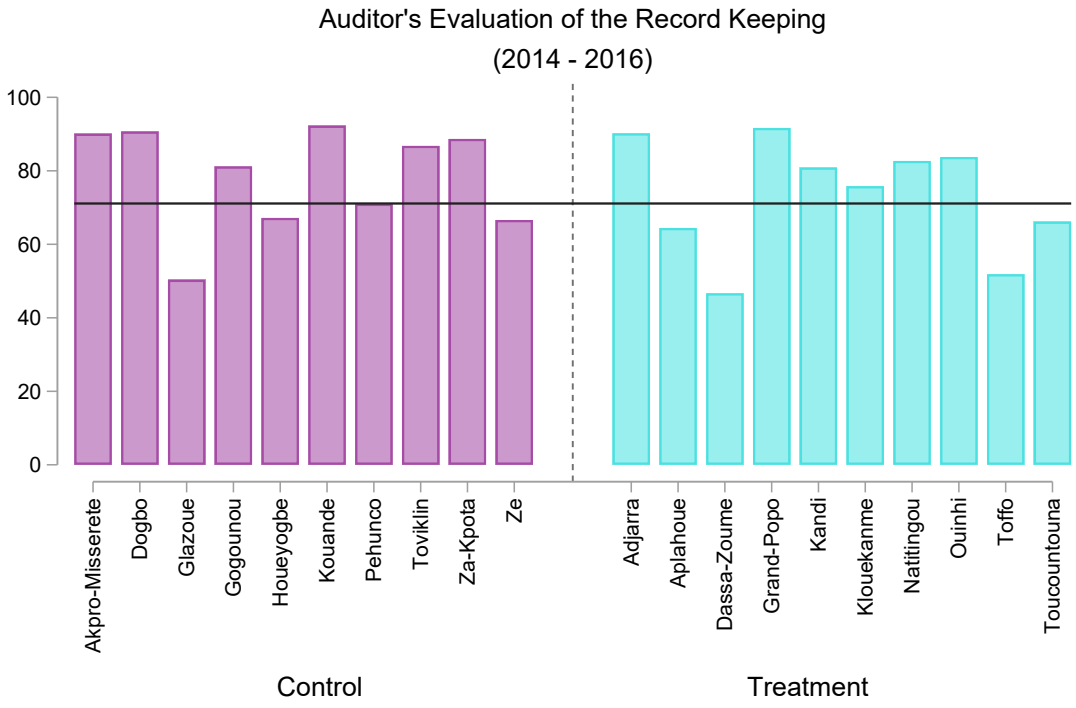
Notes: Data are from meeting notes taken by enumerators.

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



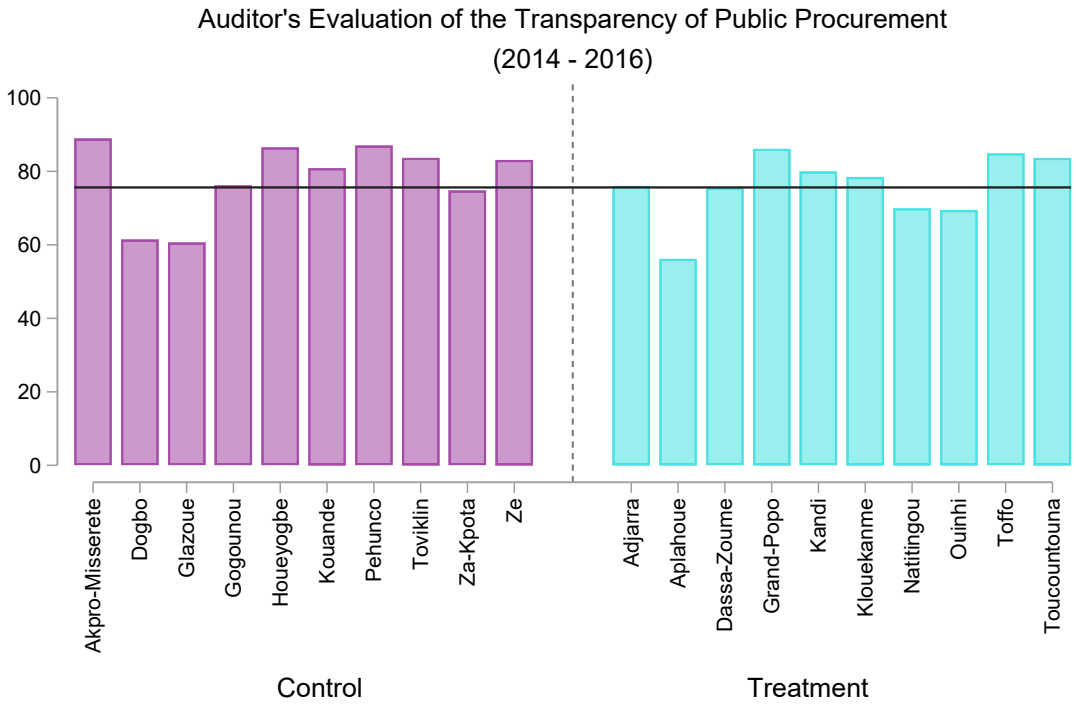
Notes: Municipal Audit-Based Performance is a score constructed from three sub-components: (i) Record Keeping (ii) Transparency of Public Procurement (iii) Overall Management. Record Keeping is constructed from scores on: maintenance of accounting records, existence of documentation room/archive accessible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



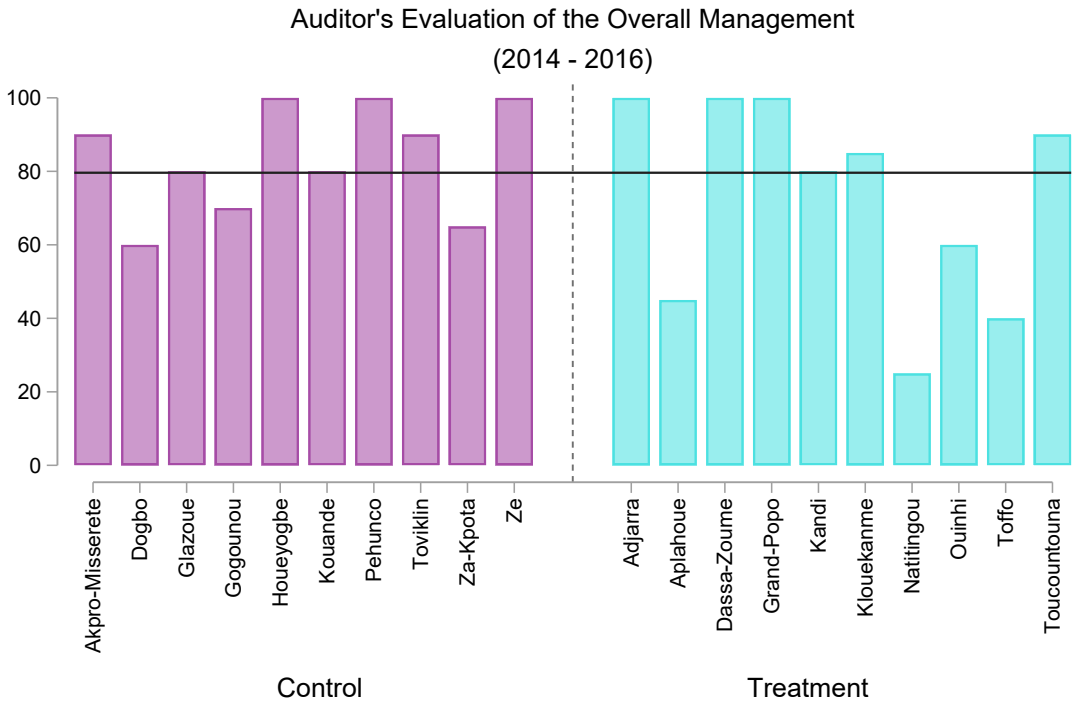
Notes: Record Keeping is constructed from scores on: maintenance of accounting records, existence of documentation room/archive accessible to the public and equipped with storage furniture, statements of transferred resources, updated physical registers and stock-taking by the authorized officers.

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



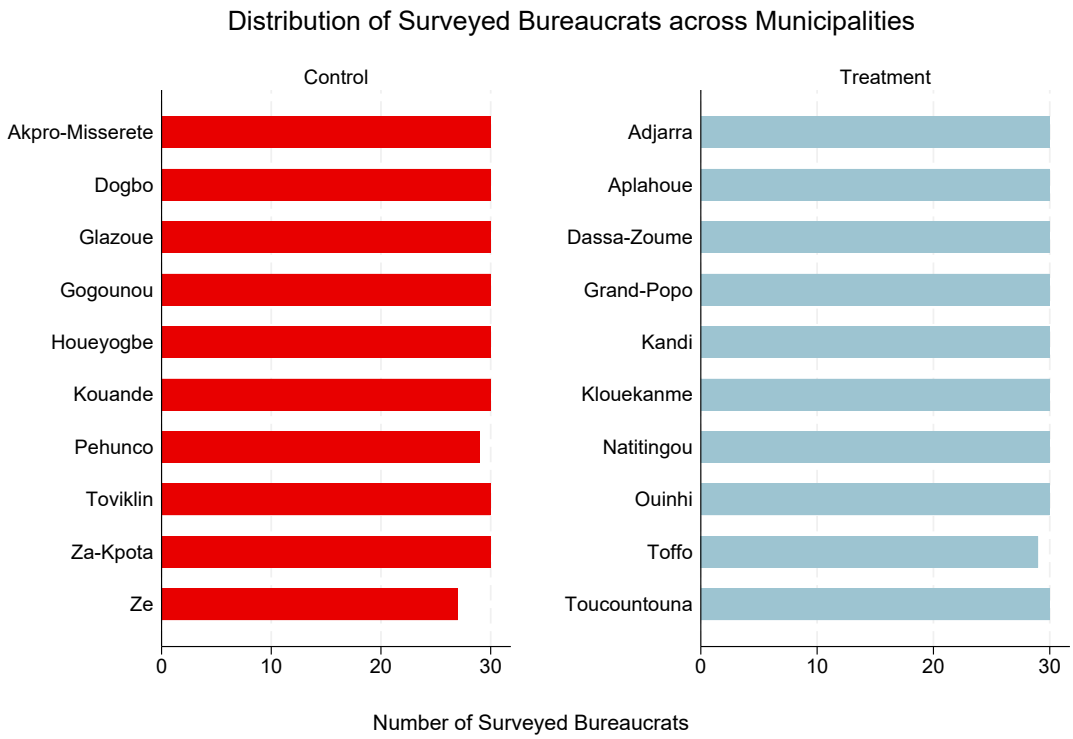
Notes: Transparency of Public Procurement is constructed from scores given on: Secretariat of the organizations responsible for public procurement is functional, the number and exhaustiveness of the reports published by them, details on FaDeC funds and spending, compliance with public procurement procedures,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

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



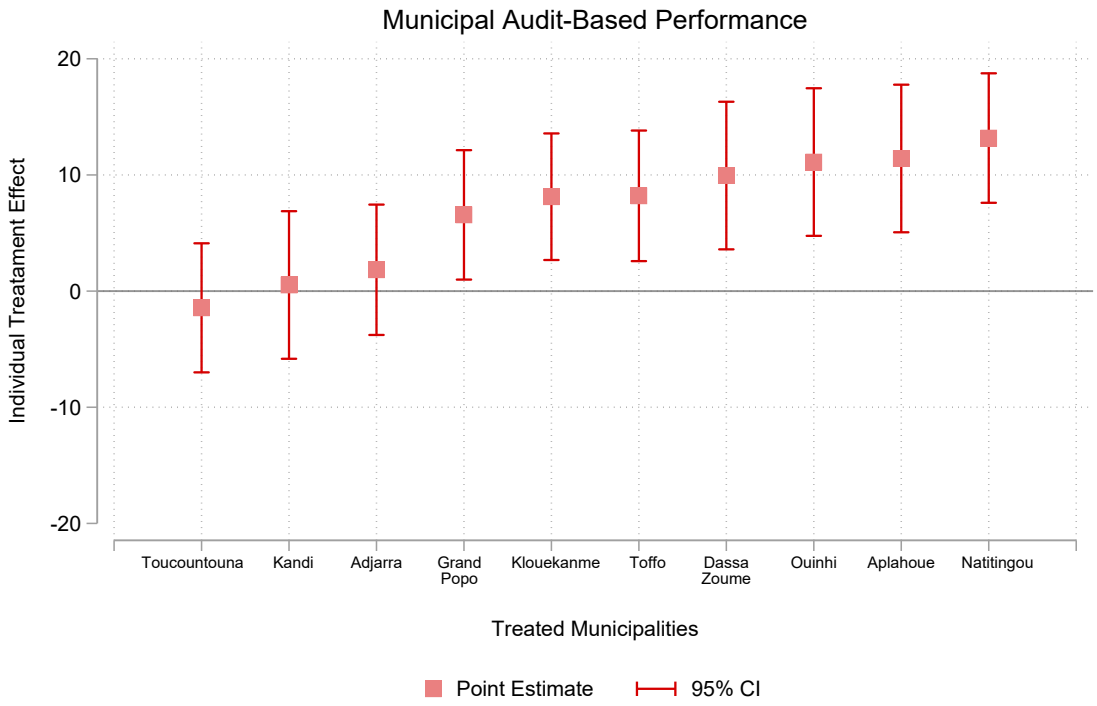
Notes: Overall Management is constructed from scores given on effectiveness of the role of coordinator of local services by the Secretary General, and management of certificates of civil status (certificates of birth, death, marriage and divorce)

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



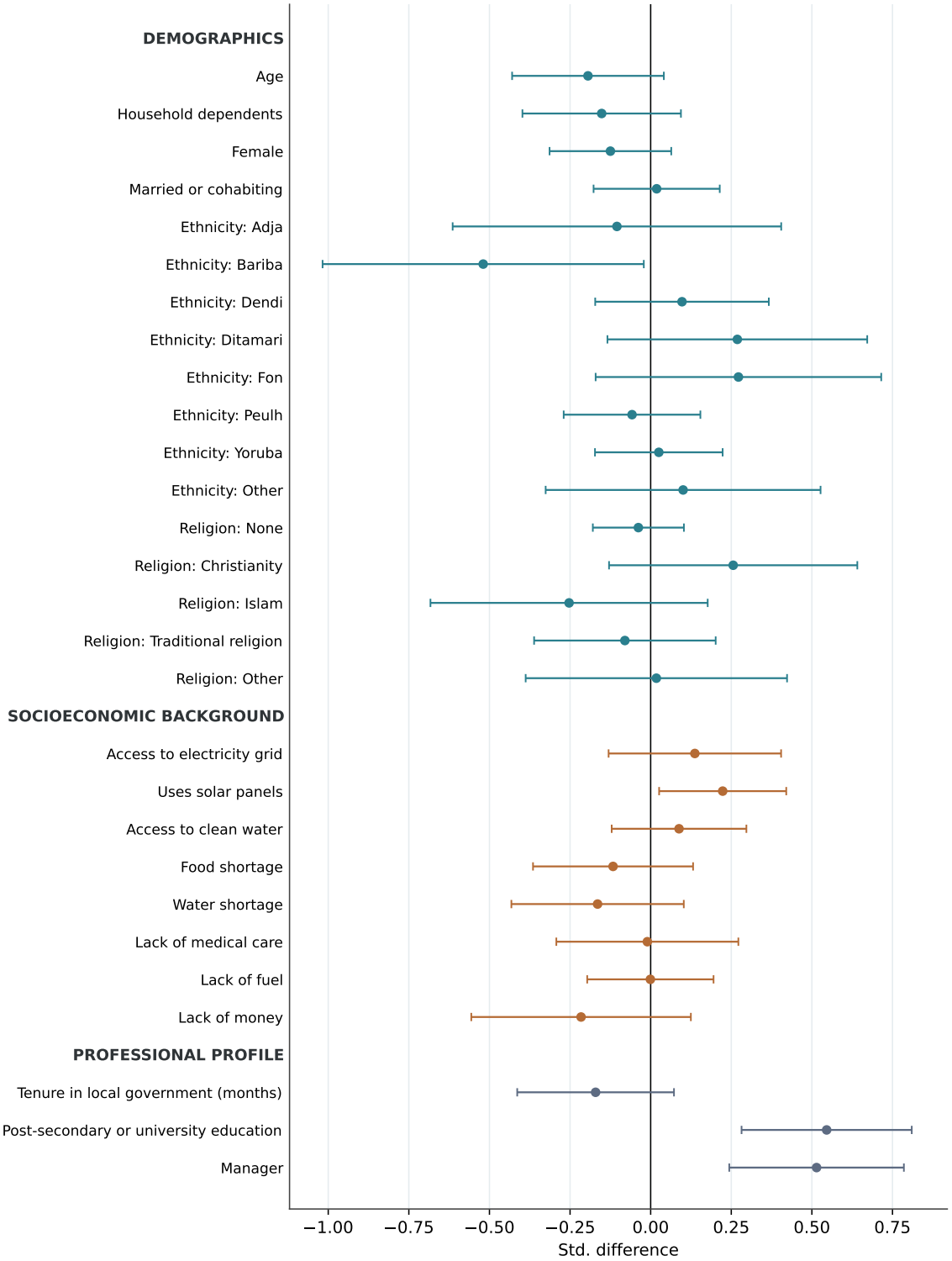
Notes: For the endline survey conducted in December 2016, 30 bureaucrats were randomly selected within each of the 20 municipalities. However, there were some exceptions: 27 in Ze and 29 each in Pehunco & Toffo, bringing the total number of observations to 595.

Figure B.10: Treatment Effects in the different Treated Municipalities



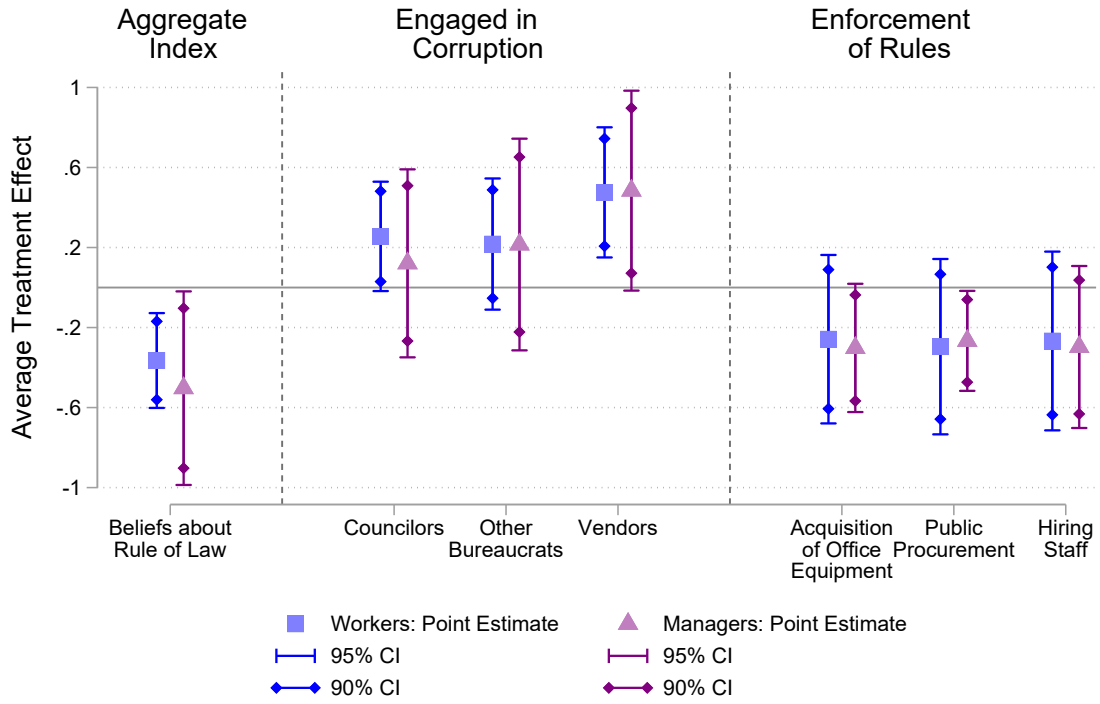
Notes: The figure shows estimated treatment effects on municipal audit-based performance for each treated municipality individually, using a difference-in-differences specification with municipality and year fixed effects.

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



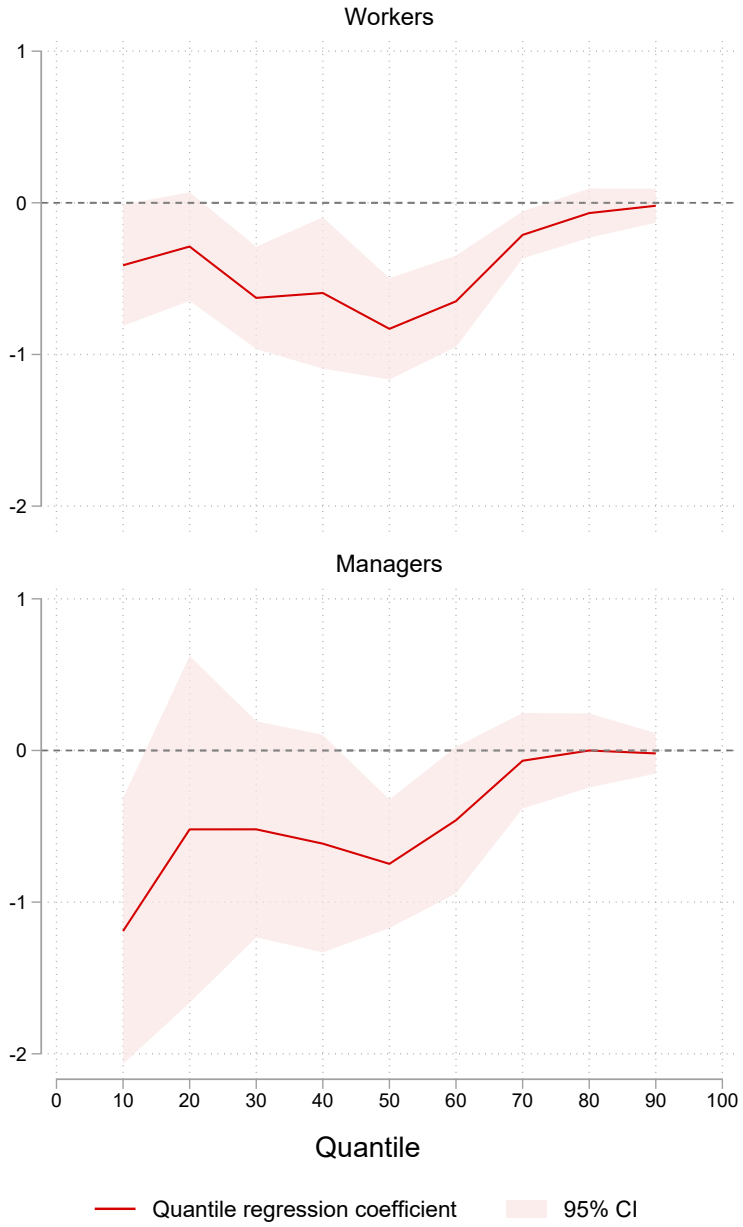
Notes: Each point is the coefficient from a separate regression of a standardized respondent characteristic on an indicator for attending at least one deliberation meeting. Bars show 95% confidence intervals using standard errors clustered by municipality.

Figure B.12: 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.

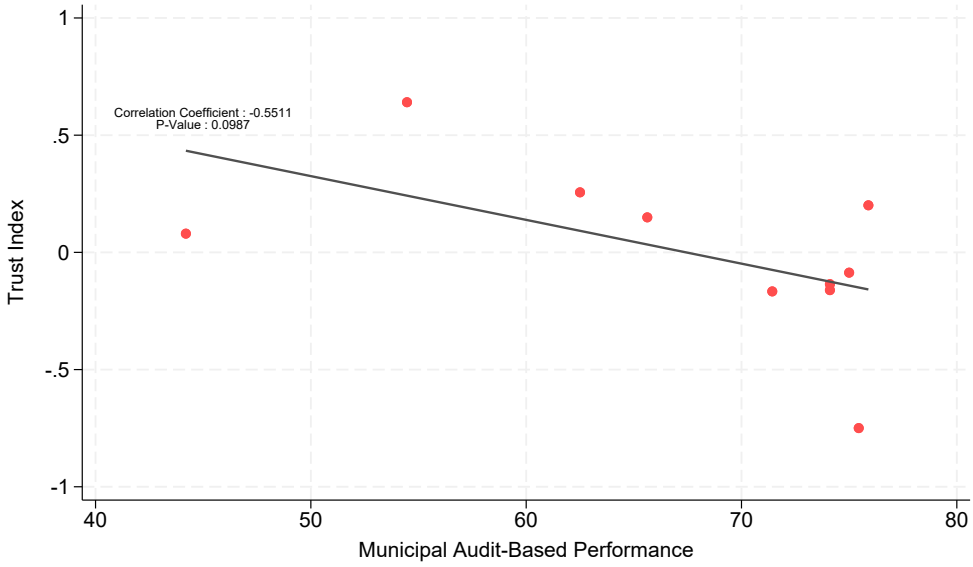
Figure B.13: Workers and managers: Quantile estimates of the effects of deliberation on bureaucratic beliefs about rule of law



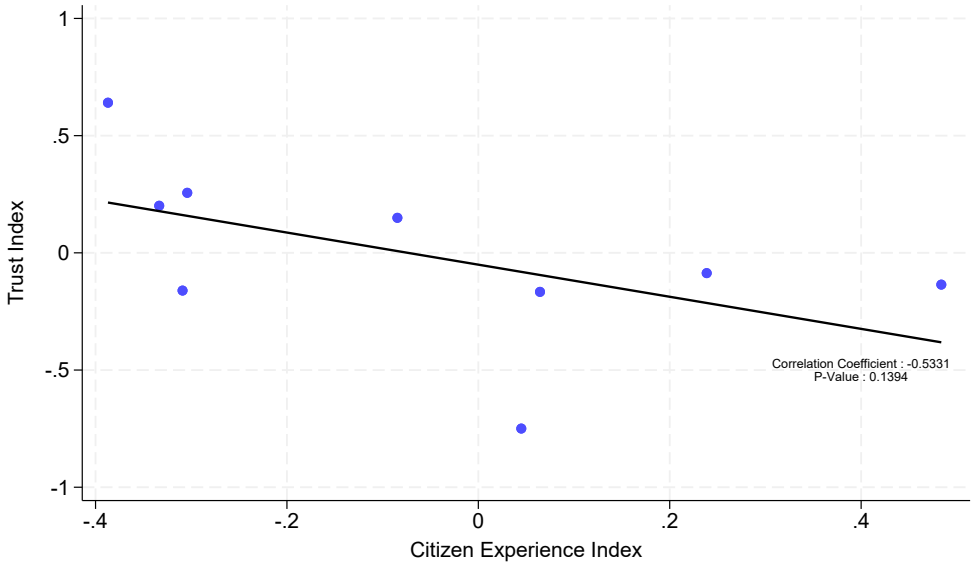
Notes: The figure reports quantile regression estimates of the treatment effect on the Beliefs about Rule of Law index, separately for workers and managers, across quantiles 10 to 90. The shaded area represents the 95% confidence interval. Standard errors are clustered at the municipality level.

Figure B.14: Relationship between Trust and Performance Outcomes in Control Municipalities

(a) Trust and Audit-Based Performance

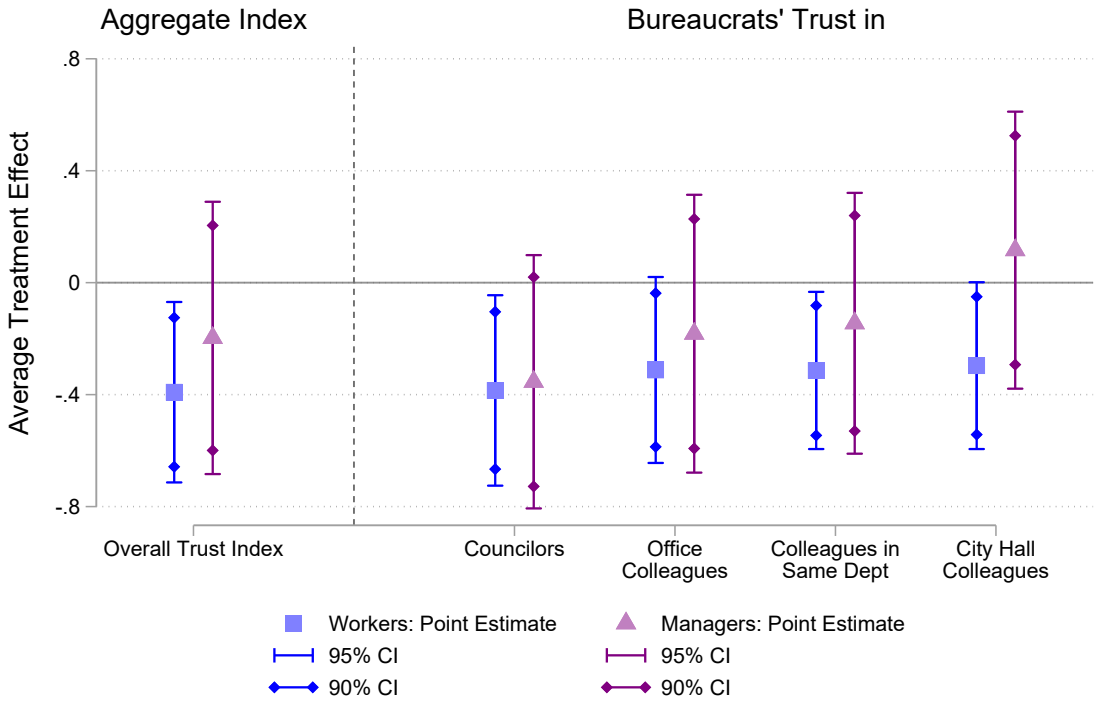


(b) Trust and Citizen Experience



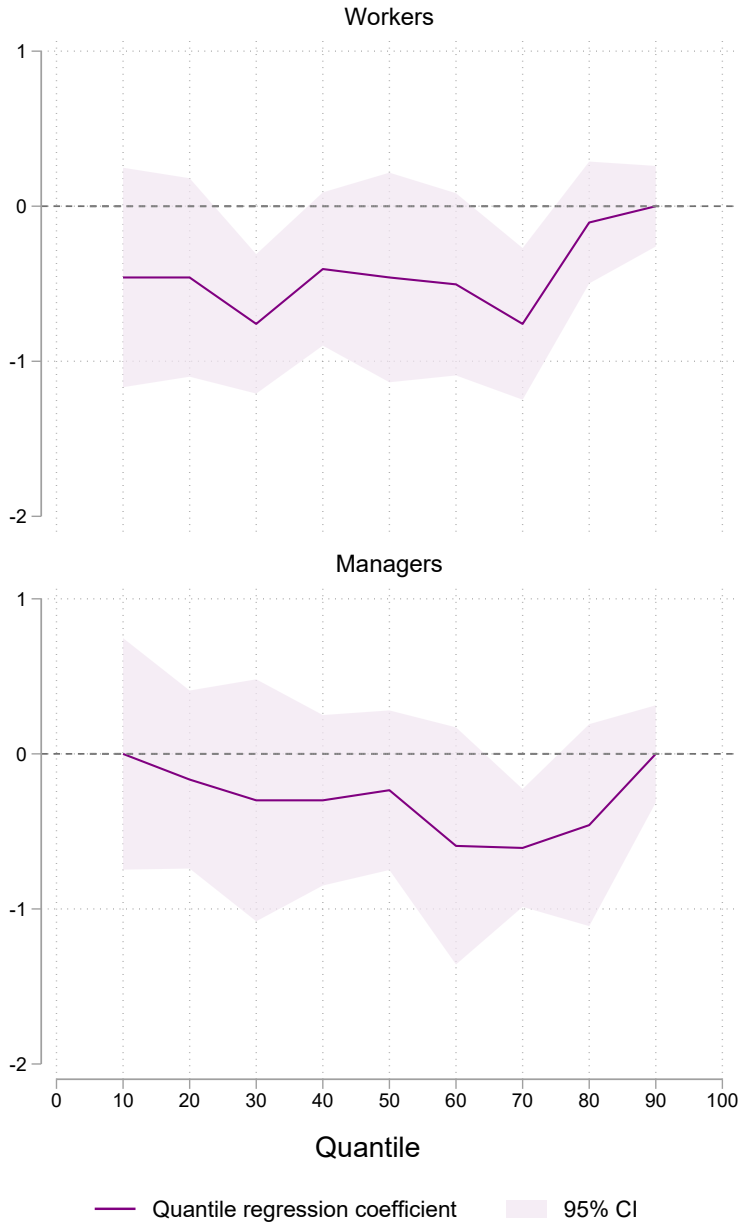
Notes: Each panel plots, across the control municipalities, the relationship between the *trust index*—a standardized aggregate of bureaucrats' trust in the municipal council, their office and departmental colleagues, and other town-hall staff—and a measure of local government performance. The top panel uses *municipal audit-based performance*, the CoNaFiL audit score expressed as a percentage of the maximum attainable score; the bottom panel uses the *citizen experience index*, a standardized aggregate of Afrobarometer measures of citizens' perceptions of corruption, living conditions, and access to public services. Each panel reports the correlation coefficient and its p-value.

Figure B.15: 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.16: 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.17: 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**

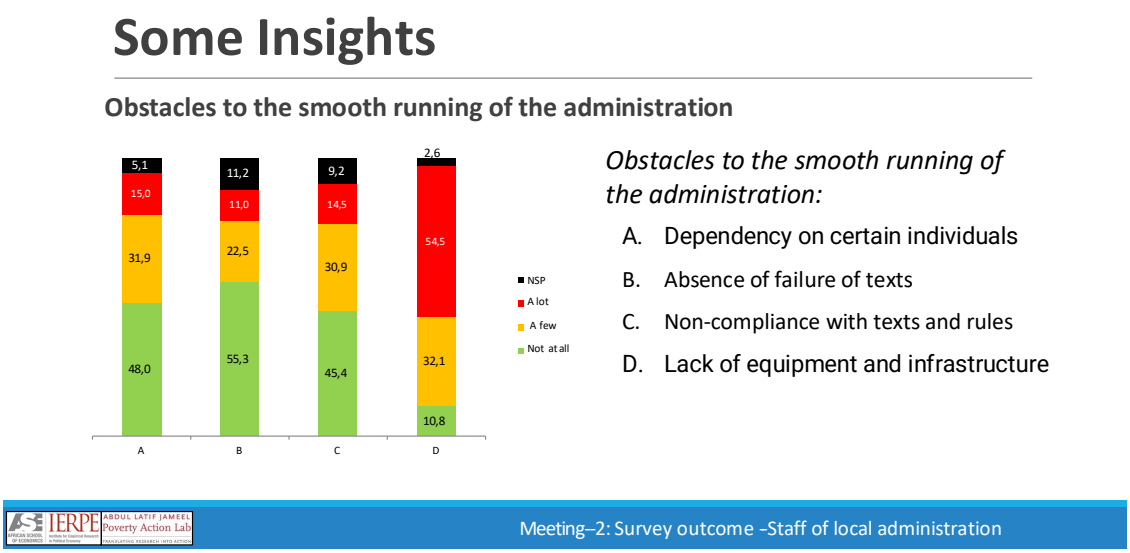
- The local treasury has informed the municipality of resource availability;
- Very long effective time taken to make BTRs available compared to the date of issue;
- Auxiliary account records are updated to reflect the availability of resources to commit to spending needs.

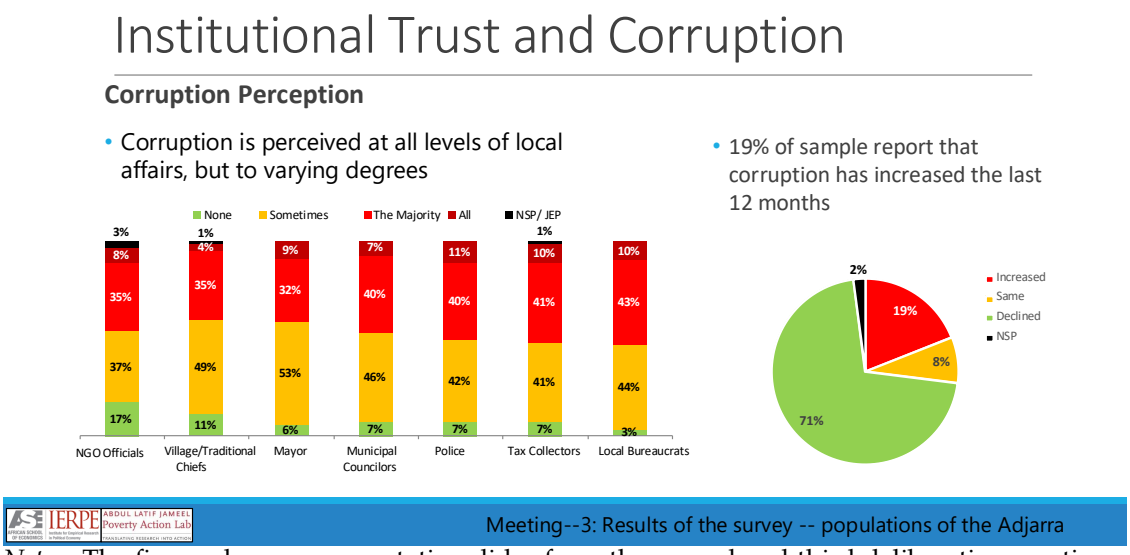
✓ **Risks**

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





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

D Audit-Based Performance Measures

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

D.1 Data Source and Digitization

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

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

D.2 Performance Components

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

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

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

Tables D.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.

$$\text{Municipal Audit-Based Performance}_{mt} = \frac{RawScore_{mt}}{MaxScore_t} \times 100$$

where $RawScore_{mt}$ is the total score obtained by municipality m in year t, and $MaxScore_t$ is the maximum attainable score in that year. Table D.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 Enforcement of Rules. 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 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.4 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

F Downstream effects on citizen experience

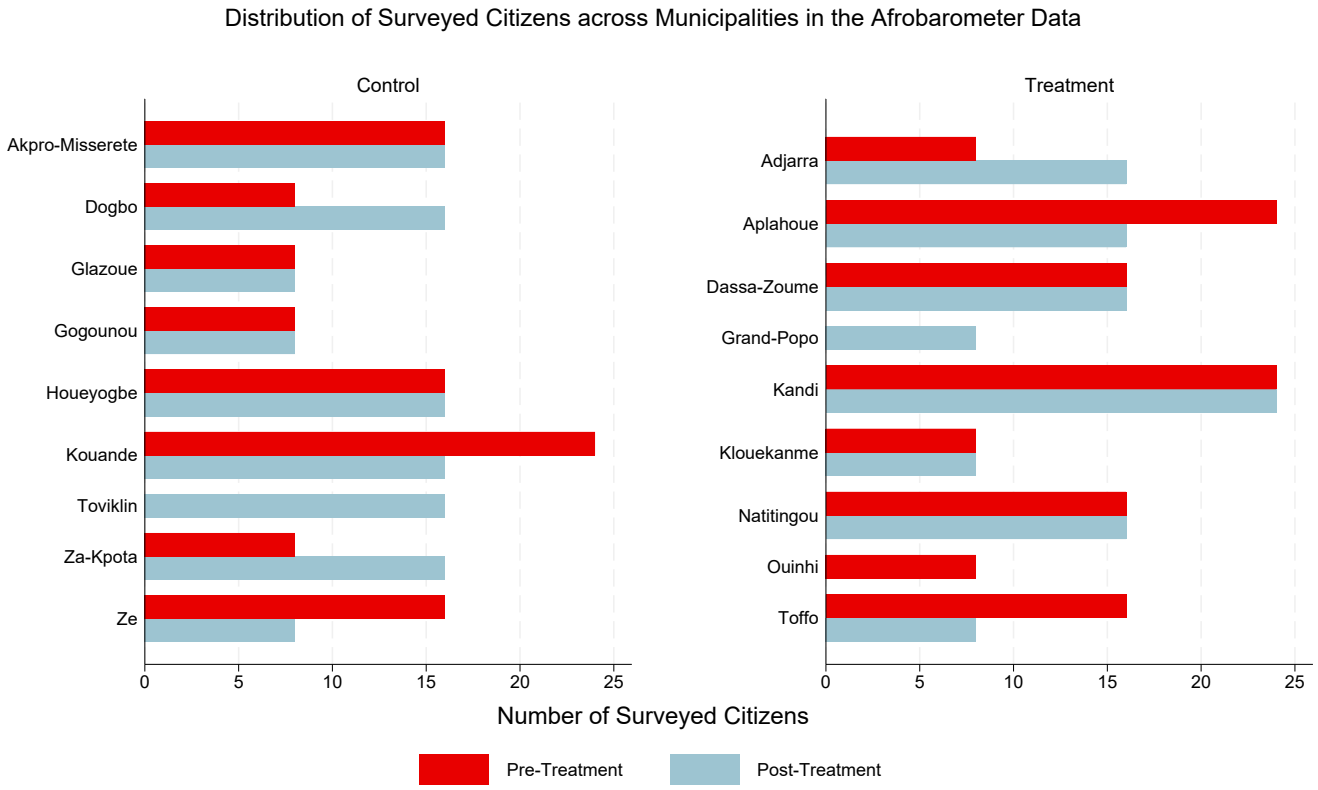
While our primary outcome is administrative performance as measured through municipal audits, an important question is whether these improvements translate into better citizen experiences with local government. To examine downstream effects on public service delivery, we use data from the Afrobarometer, 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. Figure F.1 presents details of the number of citizens surveyed across treatment and control municipalities.

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 (Anderson, 2008). We estimate effects using the same specification as in Equation 1, replacing the outcome with citizen experience.

Table F.1 presents the main results and Tables F.2 and F.3 present heterogeneity analysis based on baseline citizen experiences with the local government. Results are discussed in Section 5.4 of the paper.

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	DID Estimate	Standard Error	P-Value	Randomization Inference P-Value	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Aggregate Index						
Citizen Experience Index	.	0.1668	0.1498	0.281	0.375	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

Notes: The unit of observation is a citizen-year, and the analysis uses Afrobarometer data for 2014 and 2017. Panel A reports indicators for whether citizens perceive corruption among the council and government officials and whether such officials go unpunished. Panel B reports indicators for whether citizens feel that their present living conditions are good and that they are better off than others. Panels C and D report, respectively, indicators for whether citizens had difficulty accessing—and whether they had to pay bribes to obtain—(C) identity documents (e.g., a birth certificate, driver's license, passport, voter's card, or government permit) and (D) household services (water, electricity, and sanitation). We report the *p*-value in Column (4) and the randomization-inference *p*-value (1,000 repetitions) in Column (5). Standard errors are clustered at the municipality level. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table F.2: Heterogeneity based on baseline citizen experience

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

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 F.3). Above- vs. below-mean municipalities are classified by the baseline (2014) citizen experience index. The estimating equation is $y_{ict} = \beta_1(T_c \times Post_t) + \beta_2(T_c \times Post_t \times Above_c) + \gamma_t + \delta_c + \varepsilon_{ict}$ 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 Control Mean	Treatment	Treatment × Above Mean	Obs.	P-Value: T + T × Above Mean = 0
	(1)	(2)	(3)	(4)	(5)
Panel A: Corruption					
Corrupt among Govt Officials	1	-0.0293 (0.0385) {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

Notes: 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 F.2. 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.