Government Oversight and Inter-Institutional Legibility: Evidence from Colombia

Supporting Information

March 2, 2025

Contents

A1 Corruption and Transparency in Colombia

We first show Colombia's levels of corruption from a comparative perspective. To do so, we rely on the Corruption Perceptions Index (CPI), developed by Transparency International. We focus on the most recent data corresponding to the 2020 measurement of the CPI.[1] Figure A1 shows where Colombia stands vis-á-vis: i) other upper-middle-income countries (top panel); ii) the region (middle); and iii) the world (bottom panel).

Next, we turn to public opinion about corruption in Colombia. We draw from data from the Americas-Barometer by the Latin American Public Opinion Project (LAPOP). We use information from the 2016, 2014, and 2018 waves.[2] In Figures A2a and A2b, we plot responses to the following AmericasBarometer questions:

1. Figure A2a: "In your opinion, what is the most serious problem faced by the country?"
2. Figure A2b, left panel: "Taking into account your own experience or what you have heard, corruption among public officials is...,"
3. Figure A2b, right panel: "Thinking of the politicians of Colombia... how many of them do you believe are involved in corruption?"

Figure A2a shows that corruption consistently ranks among the most frequently cited problems in Colombia. In particular, subsequent to signing of the peace accord (2016), corruption represented the most frequently cited problem in 2018. Figure A2b shows that the modal respondent believes that public officials and politicians are engaged in corruption.

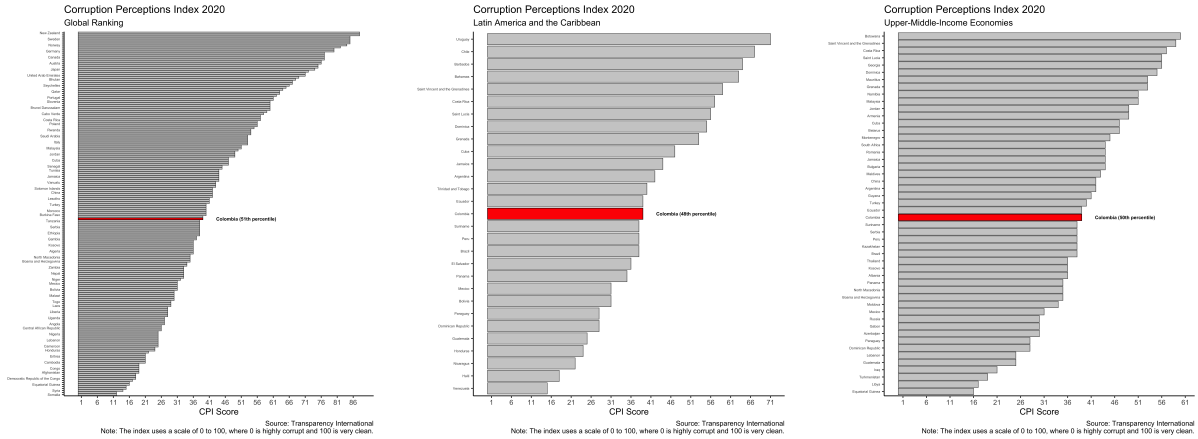
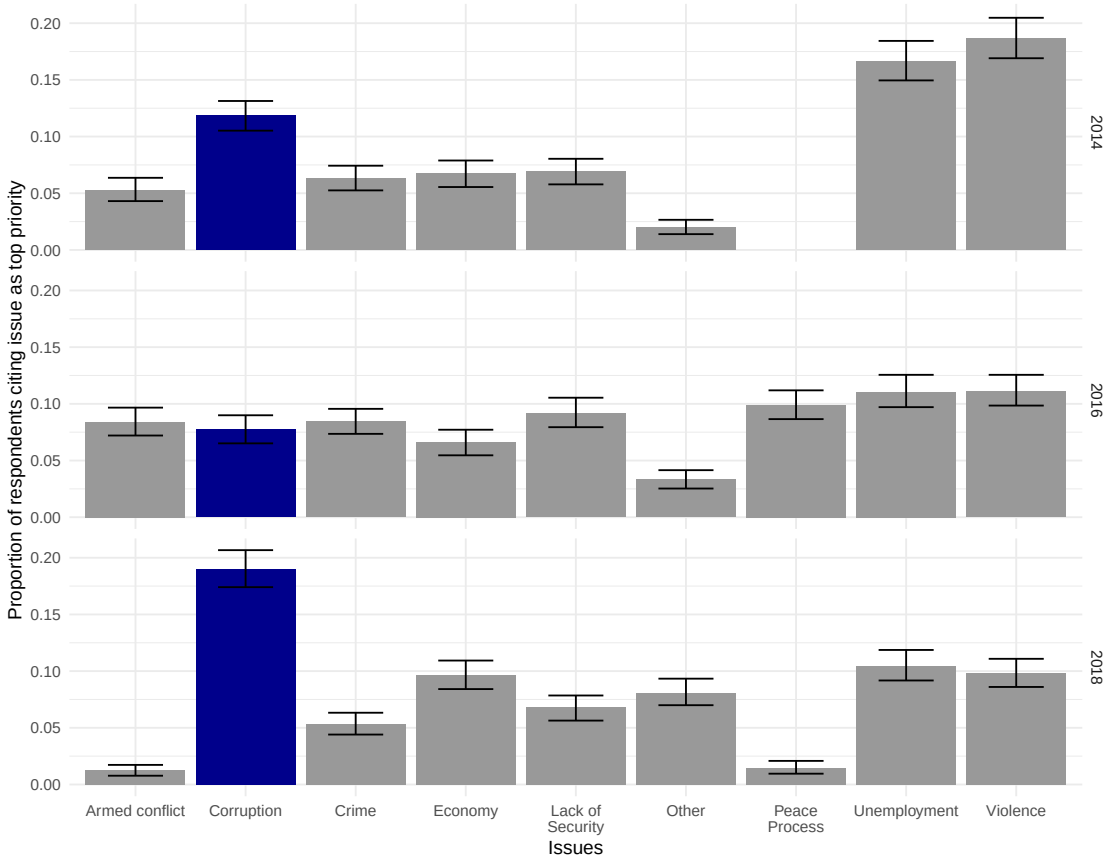


Figure A1: Colombia's ranking in 2020 Corruption Perception Index. The comparison groups are: all countries (left), Latin America (middle), and all upper-middle income countries (right).

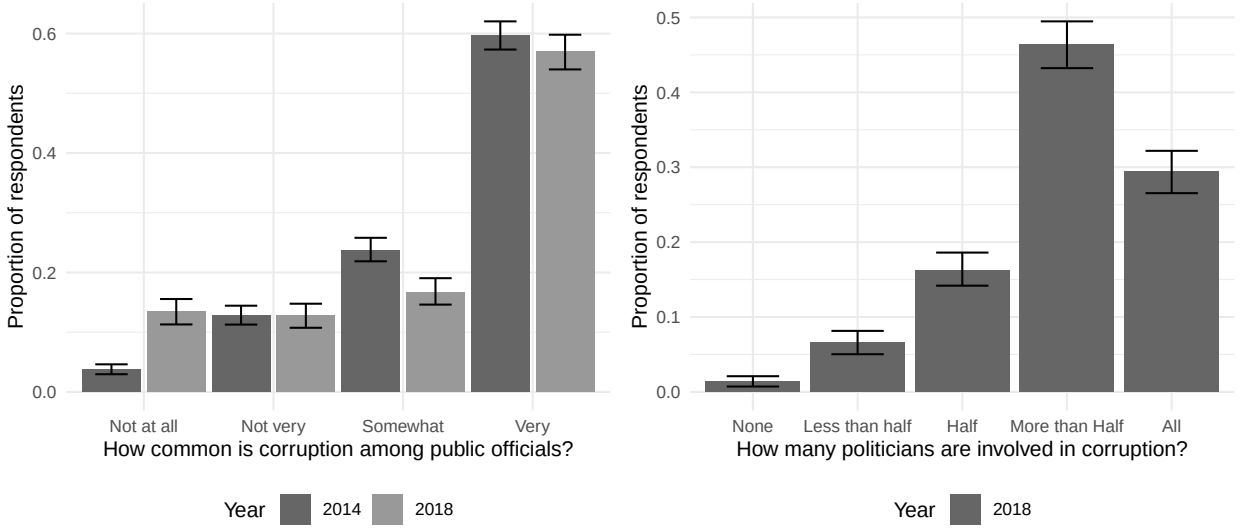
[1] The full report as well as data are publicly available on the NGO's website.

[2] The data for the 2018 was collected between late 2018 and the beginning of 2019. Data for all years are publicly available here.

Figure A2: Perceptions of corruption in Colombia from LAPOP surveys.



(a) Corruption as main problem in the country country



(b) Perceptions on the prevalence of political corruption

A2 Experimental design

A2.1 Inclusion of non-traditional entities

In addition to the near-universe of public sector entities that we describe in the manuscript, the experiment also included a subset of the "non-traditional" entities that are required to fill out ITA. These entities are generally (i) individuals or firms that contract to the state or (ii) political parties and social movements. Below, we include a translation of Article 5 of Law 1712 of 2014 (which we call the "National Transparency Act"), which defines the criteria for inclusion on this list. It reads as follows:[3]

"*The provisions of this law will be applicable to the following persons as obliged subjects:*

a) Any public entity, including those belonging to all Branches of Public Power, at all levels of the state structure, central or decentralized by services or territorially, at the national, departmental, municipal and district levels.

b) Independent or autonomous state and control bodies, agencies and entities.

c) Natural and legal persons, public or private, who provide public functions, who provide public services with respect to the information directly related to the provision of the public service.

d) Any natural or legal person or dependency of a legal person that performs a public function or a public authority, with respect to the information directly related to the performance of his function.

e) Public companies created by law, State companies and companies in which it has participation.

f) Political parties or movements and significant groups of citizens.

g) The entities that administer parafiscal institutions, funds or resources of a public nature or origin.

Natural or legal persons that receive or mediate territorial and national funds or public benefits and do not meet any of the other requirements to be considered obligated subjects, must only comply with this law with respect to that information that is produced in relation to public funds that receive or mediate."

Importantly, in addition to exact blocking on 2019 ITA completion, we also used exact blocking on the traditional/non-traditional distinction. As such, the public-sector entities represented in the main text comprise completely separate blocks from the other entities included in the experiment. Table A1 describes the full experimental sample.

A2.2 Intervention materials

We report and translate the content of the direct communication from the PGN to obligated entities assigned to any direct communication treatment in Table A2. All entities received the core information about ITA. The subsequent experimental treatment conditions were randomized (and crossed).

[3] Authors' translation

	All Obligated Entities*	**Experimental Entities**	**Audited Entities**
Category	Count (n)	Count (n)	Count (n)
PUBLIC SECTOR	6,556	6,556	2,400
National	237	237	200
Territorial	5,928	5,928	2,200
Undesignated	391	391	0
PRIVATE SECTOR	41,938	5,329	0
PGN Priority	5,329	5,329	0
PGN Non-priority	36,609	0	0
PARTIES/MOVEMENTS	168	168	0
Total	48,662	12,053	2,400

Table A1: Sampling of entities in experiment and audit outcome measurement. *This total omits 62 public sector and 38 private sector entities that were randomly sampled and used in a piloting pre-test of intervention implementation.

Table A2: Original and Spanish translation of the letters sent to entities by treatment status

Treatment condition	Original message (in Spanish)	Translation (English)
Information about ITA (core)	Como es de su conocimiento, la Procuraduría General de la Nación, de conformidad con lo dispuesto en el artículo 23 de la Ley 1712 de 2014 "Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional", ha puesto en marcha un sistema de información que permite el registro, seguimiento y monitoreo que automatiza la captura de la información de la "Matriz de cumplimiento normativo de la Ley 1712 de 2014", a través de un formulario de autodiagnóstico. Dicho sistema es la base del Índice de Transparencia y Acceso a la Información – ITA, cuya primera medición se realizó en 2019. Teniendo en cuenta lo anterior, me dirijo a usted con el propósito de recordarle que la entidad que representa tiene la obligación de diligenciar la Matriz ITA para la medición 2020. La plataforma para diligenciar este formulario estará habilitada entre el 15 de septiembre y el 15 de octubre y puede acceder a ella a través de este enlace: https://apps.procuraduria.gov.co/ita/login/. De igual manera, me permito remitir la Directiva No. 026 del 25 de agosto del año en curso, para su conocimiento, cumplimiento y difusión entre los particulares que estén encargados de realizar esta labor. Para cualquier inquietud o comentarios deberá escribir al correo electrónico: soporteita@procuraduria.gov.co. [Other text] De antemano agradezco mucho la colaboración que usted pueda prestarnos, en orden a promover el cumplimiento de esta Directiva.	As you know, the Office of the Inspector Attorney General, following the provisions of Article 23 of Law 1712 of 2014 "Law of Transparency and the Right of Access to National Public Information," has implemented a system of information that allows the registration, follow-up, and monitoring that automates the capture of the information of the "Regulatory compliance matrix of Law 1712 of 2014," through a self-diagnosis form. This system is the basis of the Index of Transparency and Access to Information - ITA, whose first measurement was carried out in 2019. Taking into account the above, I am writing to you to remind you that the entity that you represent must fill out the ITA Matrix for the 2020 measurement. The platform to fill out this form will be enabled between September 15 and 15 October and you can access it through this link: https://apps.procuraduria.gov.co/ita/login/. In the same way, I am sending Directive No. 026 of August 25 of the current year, for your knowledge, compliance, and dissemination among the individuals who are in charge of carrying out this task. For any questions or comments, you should write to the email: soporteita@procuraduria.gov.co. [Other text] Thank you very much in advance for your collaboration in promoting compliance with this Directive.

Table A2 continued from previous page

Training	Adicionalmente, invito a la entidad a visitar la página de la PGN donde se encuentran disponibles dos videos tutoriales donde se explica cómo se debe diligenciar la matriz y la grabación de la capacitación que realizó la PGN en 2019 para instruir y explicar a los sujetos obligados sobre la Matriz ITA. Los videos se encuentran en la parte inferior de la página que se accede a través de este enlace: `https://www.procuraduria.gov.co/portal/ITA.page`.	Additionally, I would like to invite the entity to visit the PGN's website, where two tutorial videos are available explaining how to fill out the matrix and the recording of the training carried out by the PGN in 2019 to instruct and explain to the obligated subjects about the ITA Matrix. The videos are at the bottom of the page, which can be accessed through this link: `https://www.procuraduria.gov.co/portal/ITA.page`.
Oversight (retrospective) for those that complied in 2019	Finalmente, se identificó que la entidad que usted representa diligenció oportunamente la matriz para la medición 2019, dando cumplimiento a la Directiva 006 del 14 de mayo de 2019 proferida por el Procurador General de la Nación. Por este motivo, agradezco su colaboración en la medición anterior y su compromiso para garantizar dentro de la entidad el cumplimiento de lo dispuesto en la Directiva 026 para la medición 2020.	Finally, we identified that the entity you represent filled out the matrix on time for the 2019 measurement, thus complying with Directive 006 of May 14, 2019, issued by the Inspector Attorney General. For this reason, I thank you in advance for your collaboration and commitment to promoting within the entity compliance with the provisions of Directive 026 for 2020 measurement.
Oversight (retrospective) for those that did not comply in 2019	Finalmente, se identificó que la entidad que usted representa no diligenció oportunamente la matriz para la medición 2019, por lo cual se evidencia un incumplimiento a la Directiva 006 del 14 de mayo de 2019 proferida por el Procurador General de la Nación. Por este motivo, agradezco de antemano su colaboración y compromiso para promover dentro de la entidad el cumplimiento con lo dispuesto en la Directiva 026 para la medición 2020.	Finally, we identified that the entity you represent did not fill out the matrix on time for the 2019 measurement, failing to comply with Directive 006 of May 14, 2019, issued by the Inspector Attorney General. For this reason, I thank you in advance for your collaboration and commitment to promoting within the entity compliance with the provisions of Directive 026 for 2020 measurement.
Oversight (prospective)	También quisiera recordarle que el autodiagnóstico y el puntaje que arroja la plataforma no son el resultado definitivo. Dada la importancia del cumplimiento de esta Ley, tal como se realizó con la información del 2019, la información reportada en la medición 2020 pasará por un proceso cuidadoso de revisión de la PGN y puede estar sujeta a un proceso de auditoría de calidad. Lo anterior, de acuerdo con las funciones de vigilancia y prevención de la PGN.	I would also like to remind you that self-diagnosis and scoring that the platform generates are not the final result. Due to the importance of compliance with this Law, as was done with the information reported in 2019, the information reported in the [ITA] 2020 measurement will go through a careful process of oversight by the PGN and may be subject to a quality audit process in accordance with the PGN's functions of surveillance and prevention.

A2.3 Covariate balance

Given our multi-arm experimental design, we report two metrics of balance. Because our common covariates are generally categorical, we construct indicator variables for common categories (i.e., those with ≥ 36 entities, or two blocks), and regress those category indicators on our main regression specification, reported below:

$$\begin{aligned} Y_{ib} =& \beta_0 + \beta_1 \text{Direct Communication}_i + \beta_2 \text{Reminder}_i + \beta_3 \text{Training}_i + \beta_4 \text{Retrospective Oversight}_i + \\ & \beta_5 \text{Prospective Oversight}_i + \epsilon_{ib} \end{aligned} \tag{1}$$

From these analyses we report the p-values associated with several test statistics. First, we report the p-values testing the null hypotheses that $\beta_j = 0$ for each $j \in 1, ..., 5$. This tests for imbalance in individual treatment conditions. These p-values are plotted in Figure A3. In a second test, available upon request, we report p-values from an F-test of the null hypothesis that $\beta_j = 0 \; \forall j \in 1, ..., 5$. This serves as a test of the joint significance of the five treatment indicators. As is suggested by from our results in Figure A3, we fail to reject the null hypothesis (consistent with covariate balance) for any covariate examined at the $\alpha = 0.1$ level.

We assess balance on geographic and organizational characteristics as follows:

- **Department**: Each entity is legally registered with a Colombian department. For entities that work in multiple places, this is where the entity is headquartered or encorporated. In these specifications, Y_{ib} indicates whether an entity is based in a given department. We iterate through all of the departments in Colombia.
- **Legal classification of public sector entities (*naturaleza juridica*)**: For public sector (traditional) entities, this is an indicator for their legal status, i.e., *alcaldía* or local government. We assess covariate balance over common classifications by operationalizing Y_{ib} as an indicator for a given classification. We iterate through all classifications with at least 36 entities. We have preserved the Spanish classification of these entities to reduce confusion in Figure A3.
- **Sector head**: Entities that were in the pure control condition received direction to fill out ITA from sector heads. While this variation is largely captured in our blocking covariates (exact blocking on public vs. other entities and Malahnobis distance minimization on private sector classifications that covary with sector heads), we report balance on sector heads given its relevance to the contrasts that we discuss. We assess covariate balance over common sector head entities by operationalizing Y_{ib} as an indicator for a given sector head. We iterate through all sector head entities with at least 36 entities. We preserve the Spanish names of these entities to reduce confusion in Figure A3.

A3 Ethical considerations

The manuscript describes three sources of data/forms of original data collection.

- The **experiment** conducted in collaboration with PGN: Note that the unit of treatment assignment is the entity (or organization). Thus, although there is some ambiguity as to whether this constitutes human subjects research, given that we focus on the behavior of bureaucrats within these entities, we motivate the experiment as human *participants* research.

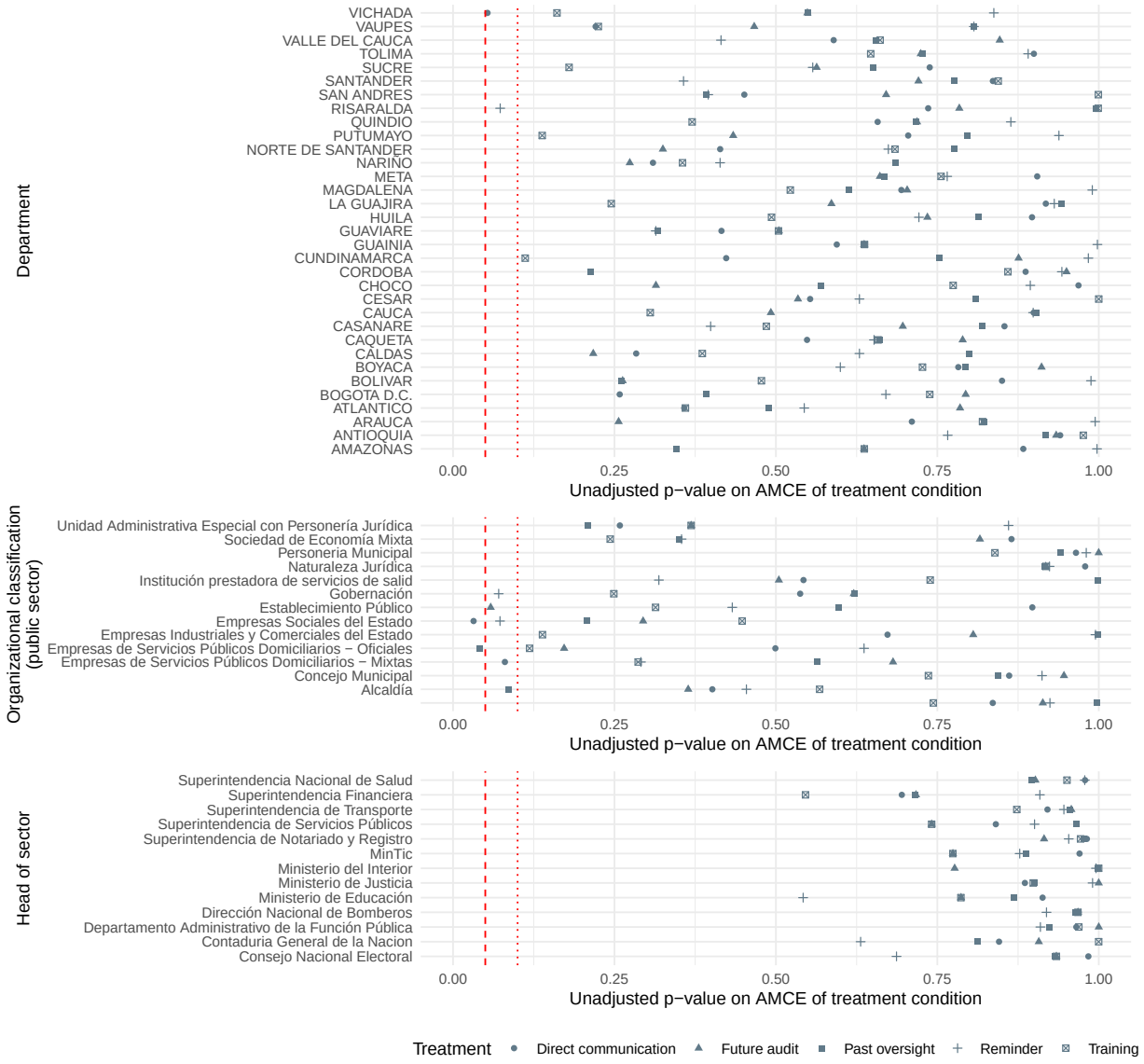


Figure A3: We assess the balance of each treatment condition. This plot graphs the p-values testing the null hypotheses that individual $\beta_j = 0$ in Equation 1 for each outcome indicator along the y-axis.

- The **independent audit** of the ITA data: Note that this is not human participants research. The audit provides an independent measure of the self-reported ITA measure.
- The **semi-structured interviews** with bureaucrats that completed the ITA matrix: These interviews are clearly human subjects research.

Table A3 enumerates the twelve APSA Principles for Human Subjects Research (American Political Science Association, 2020). For each component of the broader research design (each column), we indicate whether an exception is necessary. Note that while "collaboration with government" is not generally considered "human subjects research," it does fall within the broader category of human *participants* defined in the Principles. We therefore discuss the following principles in this light.

	APSA Principles for Human Subjects Research	Experiment	Independent Audit	Interviews
1	Respect autonomy and consider the wellbeing of participants and other people affected by research		N/A	
2	Researcher responsibility to consider the ethics of research.			
3	APSA Principles as standards of conduct.			
4	Power [between researcher and participants]			
5	Informed and voluntary consent.	✓		
6	Avoidance of deceptive or covert research.			
7	Avoidance of harm			
8	Avoidance of trauma			
9	Confidentiality of participant identities			
10	Avoid compromising the integrity of broad political processes	✓		
11	Awareness of relevant laws and regulations governing research and related activities			
12	Shared responsibility			

Table A3: The checkmarks represent exceptions justified below. Note that the independent audit does not constitute human participants research.

We identify two exceptions to the APSA "Principles and Guidance for Human Subjects Research." First, we did not seek informed consent in the experiment. However, under IRB regulations and U.S. federal law, this is only relevant if the experiment is considered human subjects research. Note that all subjects in our study are public officials or, in the case of the non-traditional subjects, representatives of organizations that contract with the state. The American Political Science Association (2020) notes that "the need to protect unconsenting participants from these harms might not apply to some research on public officials and other powerful actors" (p. 3). The "direct communication" treatment and constituent nudges present minimal harm to the bureaucrats working in the entities contacted in our study, who are still human *participants* under the APSA principles. While the broader ITA data collection and use of the data could present risks, these are not risks that are introduced by the randomized treatment. Finally, the PGN administered the treatment as part of their ITA data collection policy. They do not solicit collect consent in their interactions with other entities.

Second, as we discuss, the ITA data is used in the PGN's preventative mission. As such, the outputs of our experiment are, in principle, used to target preventative efforts. We view this as a potential social impact on a political process. However, recall that the intervention was implemented by the PGN. The American Political Science Association (2020) notes that "studies of interventions by third parties do not usually

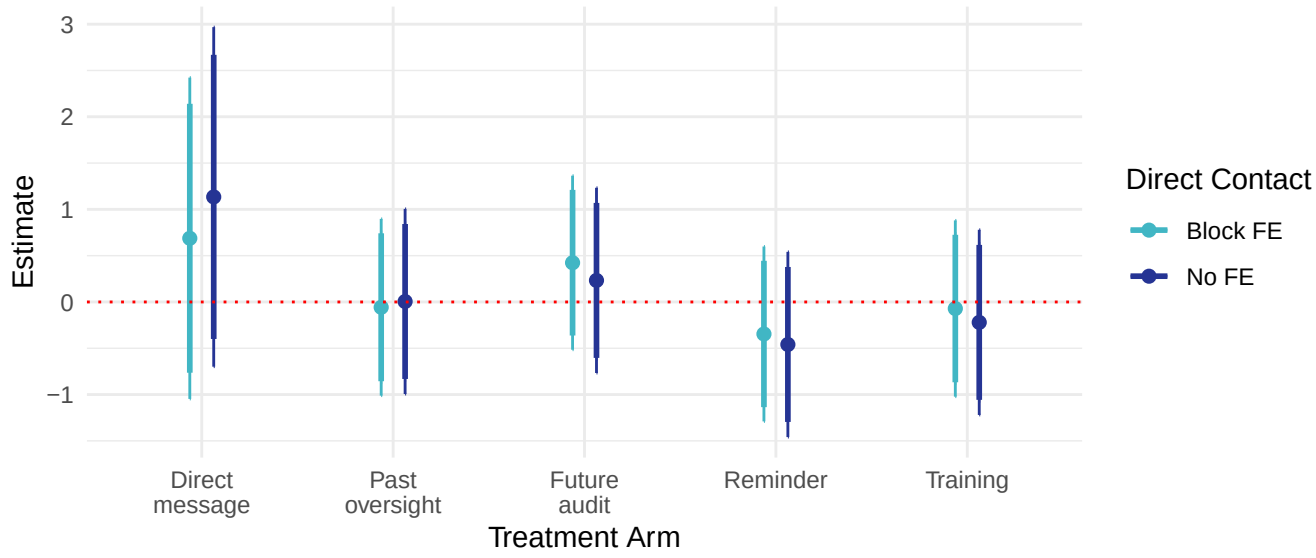


Figure A4: Estimates of the AMCE of experimental treatments on audit-measured quality within the subset of audited entities.

invoke this principle on impact" (p. 14). To promote transparency, we discuss the PGN's goals in our collaboration in the main manuscript.

A4 Audited Items, Microdata

A4.1 Selection into audit sample

Because our independent audit covers only 2,400 of 6,556 public sector entities in the full experimental sample, we consider selection into the audit sample. Per Table 1, it is clear that national entities were sampled at a higher probability than territorial entities. In Tables A4 and A5 we predict selection into the audit sample by estimating the following equation:

$$Y_i = \beta_0 + \boldsymbol{\beta}\mathbf{X}_i + \gamma \text{National}_i + \epsilon_i$$

where $\mathbf{X}_i$ is a matrix of predictors of selection into the audit sample. In Table A4 these predictors include past and current performance on the ITA matrix (in 2019 and 2020, respectively). The inclusion of an indicator for national entities accounts for the differential probability of selection into the audit sample. In Table A5, the predictors include the treatment indicators from the experiment. We do not find evidence of imbalanced selection into the audit. In both tables, all coefficients are very close to zero and we cannot reject the null hypothesis for any predictor.

Figure A4 measures the AMCEs of experimental treatments on the quality measure in the audit (our measure of θ). The outcome ranges from 0-27.75. We find no evidence that the treatments affect audit-measured transparency practices. Our estimates are small in magnitude, and we cannot reject the null hypothesis of no effect for any treatment.

	In audited sample			
	(1)	(2)	(3)	(4)
Submitted data in 2019	-0.001			
	(0.012)			
Entered transparency index system		0.004		
		(0.013)		
Submitted data in 2020			0.008	
			(0.012)	
Transparency index score/100				0.021
				(0.024)
National government entity	0.845***	0.842***	0.840***	0.836***
	(0.025)	(0.024)	(0.024)	(0.026)
Territorial government entity	0.371***	0.370***	0.369***	0.373***
	(0.007)	(0.007)	(0.007)	(0.008)
Intercept	0.000	-0.002	-0.003	-0.014
	(0.002)	(0.006)	(0.005)	(0.016)
Num. Obs.	6556	6556	6556	4446
Adjusted R^2	0.070	0.070	0.070	0.065
Sample	Public sector	Public sector	Public sector	Public sector, completed

* p < 0.1, ** p < 0.05, *** p < 0.01

Table A4: Predicting selection into the audit sample as a function of ITA matrix completion (past and present) and scores. The baseline category for the type of government entity is "unclassified," which comprises $n = 391$ entities. Heteroskedasticity-robust standard errors in parentheses.

	In audited sample			
	(1)	(2)	(3)	(4)
Direct communication	-0.004	-0.004	-0.003	-0.004
	(0.023)	(0.022)	(0.021)	(0.021)
Oversight of past completion	0.019	0.019	0.019	0.019
	(0.013)	(0.012)	(0.012)	(0.012)
Possible future audit	-0.003	-0.004	-0.003	-0.004
	(0.013)	(0.012)	(0.012)	(0.012)
Direct reminder	0.008	0.011	0.008	0.011
	(0.013)	(0.012)	(0.012)	(0.012)
Training	0.010	0.011	0.010	0.011
	(0.013)	(0.012)	(0.012)	(0.012)
National government entity		0.845***		0.893***
		(0.024)		(0.118)
Territorial government entity		0.371***		0.437***
		(0.006)		(0.111)
Num.Obs.	6556	6556	6556	6556
Adjusted R^2	0.000	0.058	0.070	0.074
Sample	Public Sector	Public Sector	Public Sector	Public Sector
Block FE			yes	yes

* p < 0.1, ** p < 0.05, *** p < 0.01

Table A5: Predicting selection into the audit sample as a function of experimental treatments. The baseline category for the type of government entity is "unclassified," which comprises $n = 391$ entities. Heteroskedasticity-robust standard errors in parentheses.

A4.2 Audited items

To keep the audit manageable, we defined a list of nine items to be audited by the independent firm. These were selected based on two criteria:

1. *Relevance*: Items which contain information that citizens would be most likely to consult or use.

2. *Feasibility*: Items that are comparable across entity classes and could be systematized by auditors.

The first criterion led to a list of 6 items. For the latter criterion, we mainly relied on the results of a pilot that we conducted with the goal of assessing the feasibility of the audit itself, and the validity of the instrument. After filtering out those items in which responses were not comparable across entities or were excessively time or data consuming (i.e., items that required downloading and opening multiple files for validation), we defined the following list of items to audit. Note that all items come from the original ITA matrix form.

1. *Transparency button*: There is a designated "Transparency" section on the home page of the obliged subject's website.

2. *Mechanisms for citizen service*: The entity must publish a section where the information on the mechanisms through which the citizen can contact the entity is described.

3. *Form for receiving citizen information*: This item refers to the characteristics of the form for receiving requests for public information. It requires to have the following fields:
   (a) Applicant type (person or company)
   (b) First name
   (c) Surname
   (d) ID Type
   (e) Identification number
   (f) Company's name
   (g) Tax ID number (NIT; for enterprises only)
   (h) Country
   (i) Department
   (j) Municipality
   (k) Address
   (l) Email
   (m) Landline and / or mobile phone
   (n) Content of the request
   (o) Files or document attachments
   (p) Option to choose the response channel
   (q) Information on possible costs associated with the response

4. *Open data*: The entity must publish the data generated by the entity on its website. They must at least have the following: (i) Index of reserved and classified public information; and (ii) Records of Information Assets. Both must be published in open data. NOTE: The publication of these data, regardless of the file format in which it is found (Word, Excel, CSV), must be available in an accessible and reusable way.

5. *Frequently answered questions*: The entity must publish a list of frequently asked questions with the respective answers, related to the entity, its management and the services and procedures it provides.

	Not in public microdata	In public microdata
Did not complete ITA (PGN measure)	615 (25.6%)	89 (3.7%)
Completed ITA (PGN measure)	478 (19.9%)	1,218 (50.8%)

Table A6: Confusion matrix for PGN data versus public microdata.

6. *Mission and vision*: The entity must publish information on its mission and vision in accordance with the creation or restructuring standard or as defined in the entity's quality management system.

7. *Assigned general budget*: The entity must publish the general budget assigned for each fiscal year. (Many times the entities publish the decree defining that budget.)

8. *Management, evaluation, and audit reports*: The entity must publish the management, evaluation and audit reports, including the budget year.

9. *Publication of audit information*: The entities that contract with public resources, or public and private resources, must publish in the SECOP the information of their contractual management with charge of public resources.

In comparing item-level responses to their audit results to measure discrepancies between reported and true scores, we rely on ITA responses present in the public microdata that records responses to each item. When we compare entities for which the PGN has recorded a score to those in the public microdata, we observe some discrepancies. Specifically, there are fewer entities in the public microdata than entities that completed the ITA matrix according to the PGN. This is evident in the lower left cell of Table A6, where nearly 20% of audited entities completed the matrix but are not present in the microdata. A further 3.7% of the sample did not complete the ITA matrix per PGN's match but is in the public microdata. Because bureaucrats self-reported entity names, which often do not match the administrative records, the PGN and the research team conducted separate hand matches between the data inputs and the scores. These 3.7% of entities is suggestive of the lack of overlap in these matches. Ultimately, this suggests that measurement error due to misattribution of scores to entities is quite limited. Our primary concern, which we discuss at greater length when interpreting results, is the absence of some entities from the public microdata.

A5 Semi-Structured Interviews

We conducted interviews with officials in different public sector entities that filled out the ITA in late 2020. We conducted these interviews in 2021 after the microdata became available. We identified the official responsible for submitting an entity's data available from this microdata. We sent invitations to participate in semi-structured interviews about ITA and data reporting in general. The response rate was 7%.
Our sampling strategy was as follows:

- Identify contacts at public sector ("traditional") entities from public microdata available from the PGN.
  - We eliminate any contacts within the PGN. These were typically individuals that assisted with submission of the ITA upon request by officials in a given entity.
- We stratify along three dimensions:

- Control versus direct message treatment assignment.
- Reported score in three bins: $\{\leq 20, \in (20, 80), \geq 80\}$.
- Elected principal: entities are classified as having an elected principal if the principal became principal by winning an election. The set of entities with elected principals includes: *alcaldías* (local governments); *consejos* (local councils); *gobernaciones* (department governments); *asambleas* (department councils); the presidency; and both houses of Congress (*Cámara de Representantes* and *Senado*).

This stratification helps to ensure that we are considering a variety of entities in the qualitative analysis.

One limitation of the survey evidence is that we do not observe officials who *declined* to submit ITA. Nevertheless interviews with officials that chose to submit data allows us to study how they understand oversight and the role of the PGN.

A6 Ancillary Experimental Analyses

We report several ancillary analyes frrom the experiment, as follows:

1. Figure A5 plots the conditional AMCEs on completion of the 2020 ITA index (left) and the difference in these conditional AMCEs by sector of entities (right). Public sector entities include PGN-designated "traditional" public sector entities. Other entities include private firms, political parties, and social movements. The left plot reports the conditional AMCEs. The estimates in the left panel come from Figure 3. We estimate the difference in conditional AMCEs using the estimators γ_j in the following OLS specification:

$$Y_{ib} = \sum_j \beta_j Z_{ij} + \sum_j \gamma_j \text{Public Sector}_i Z_{ij} + \boldsymbol{\psi}_b + \epsilon_{ib} \tag{2}$$

   Note that j indexes the treatment arms of the factorial design and Z_{ij} is an indicator variable capturing assignment to treatment j. The right panel of Figure 3 plots our estimates of the estimated γ_j's.

2. Figure A6 reports conditional AMCEs on completion of the ITA index among public sector entities, as a function of 2019 ITA matrix completion (i.e., one the lagged dependent variable). We calculate estimated conditional AMCEs and differences in CAMCES from an estimator analogous to (2) where the moderator is 2019 index completion.

3. Figure A7 reports the association between each of the treatment conditions and the reported scores, conditional on completion of ITA. The left-hand side plot reports the estimates from Panel B of Table 3. The right-hand side plot reports differences in these estimates as estimated through an estimator analogous to (2).

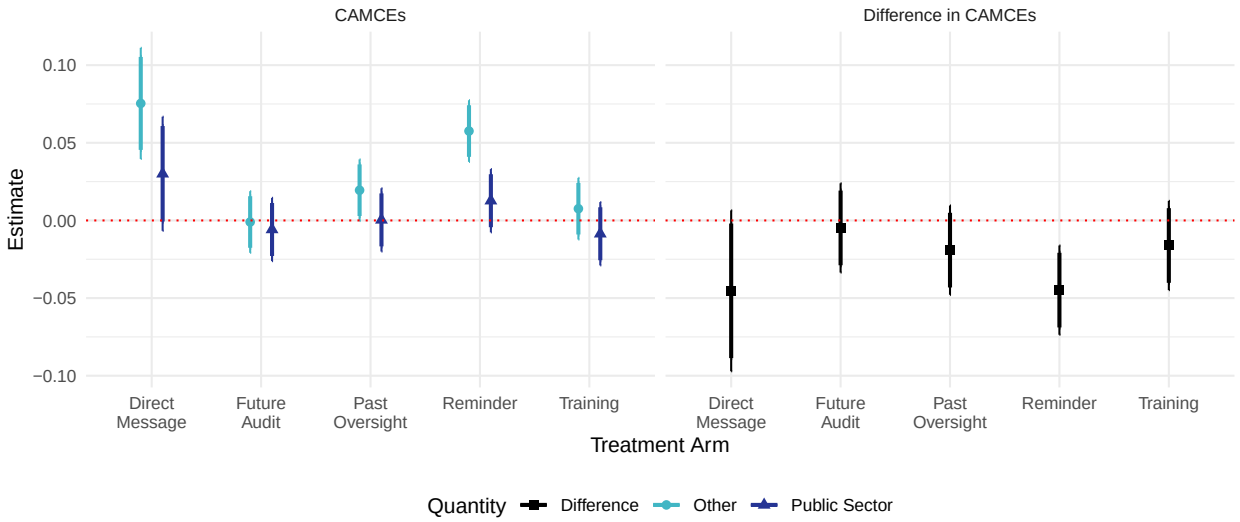


Figure A5: Conditional AMCEs on transparency index completion, among public sector and other (non-public sector) entities (left panel). Differences in conditional AMCEs between public sector and other entities (right). 90% (thin) and 95% (thick) confidence intervals constructed on heteroskedasticity-robust standard errors.

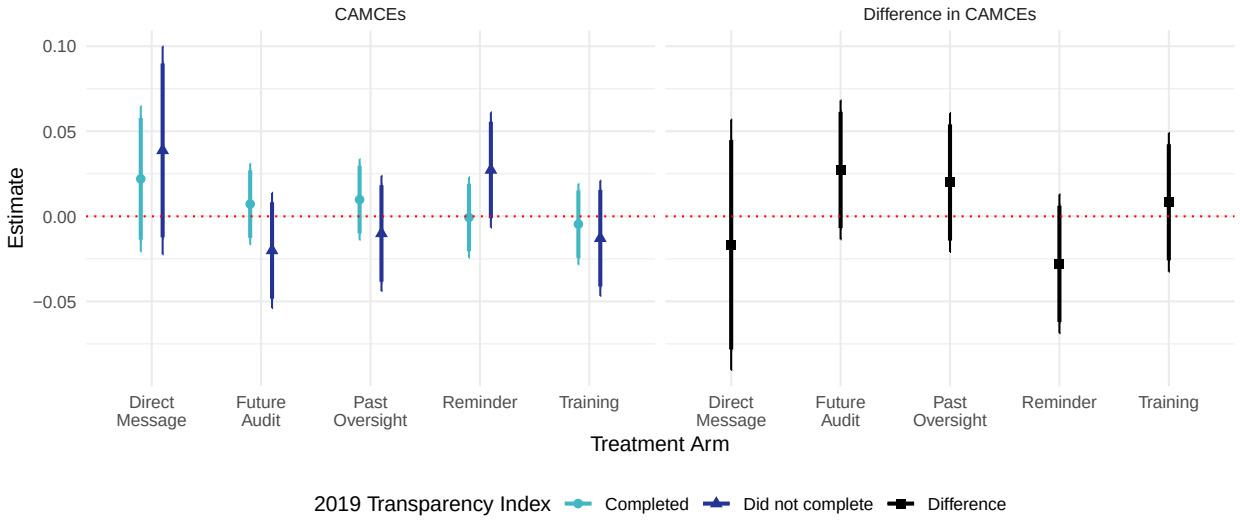


Figure A6: Conditional AMCEs on transparency index completion, among public sector entities a function of index completion in 2019. Differences in conditional AMCEs for 2019 reporters and non-reporters (right). 90% (thin) and 95% (thick) confidence intervals constructed on heteroskedasticity-robust standard errors.

A7 Evaluating Monotonicity of Selection into Reporting

In Table 3, we show that direct communication from the PGN increased rates of reporting by entities assigned to treatment. In the decomposition of the post-treatment estimand on scores in Appendix A8, we

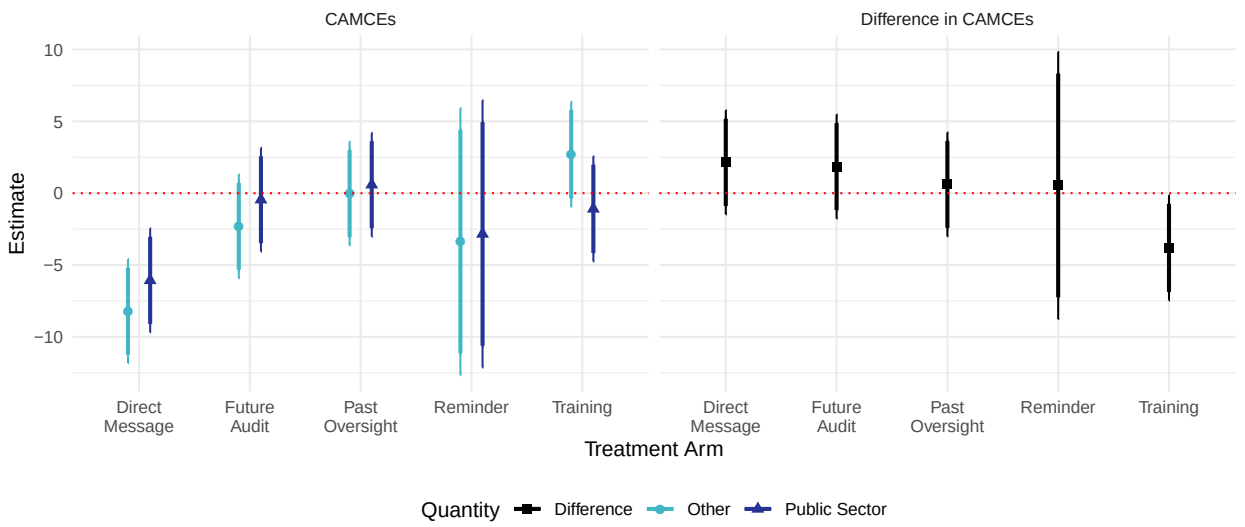


Figure A7: Conditional association between treatments and transparency index scores, among public sector and other (non-public sector) entities (left panel). Differences in conditional associations between public sector and other entities (right). 90% (thin) and 95% (thick) confidence intervals constructed on heteroskedasticity-robust standard errors.

invoke Lee (2009) bounds to estimate interval estimates of the treatment effects on always-reporters. Lee bounds assume monotonicity (or no defiers) on selection into reporting. In this analysis, we provide support for that assumption. To evaluate the assumption of monotonicity, we use machine learning to estimate CATEs across a large set of pre-treatment covariates. Specifically, we use a generalized random forest model proposed by Athey, Tibshirani, and Wager (2019). We employ each of the following covariates (as binary indicators for each level): department, level of entity (national or territorial), administrative classfication (per Colombian government classification scheme), central/decentralized administration (per Colombian government classification scheme), ITA completion in 2019, type of entity (for large categories), and an indicator for an organization with a tax identification number (in DIAN). This yields a matrix of 61 predictors.

Figure Figure A8 depicts the predicted CATEs for each entity. It shows that while 27% of the estimates are negative, none are statistically distinguishable from zero at the $\alpha = 0.05$ level. In contrast, we estimate precisely-estimated positive treatment effects for 13.5% of entities. We interpret our inability to detect (statistically) a negative treatment effect for any institution as evidence consistent with our assumption of monotonicity.

A8 Decomposition of the Post-Treatment Estimand

Denote the outcome, reporting $R_i \in \{0, 1\}$ and the reported score $S_i \in [0, 100]$. Table 3 reports estimates of the post treatment estimand:

$$\mathcal{P} \equiv E[S(Z = 1)|R(Z = 1) = 1] - E[S(Z = 0)|R(Z = 0) = 1]$$

Consider a binary treatment $Z \in \{0, 1\}$ where Z corresponds to direct contact from the PGN. We invoke an

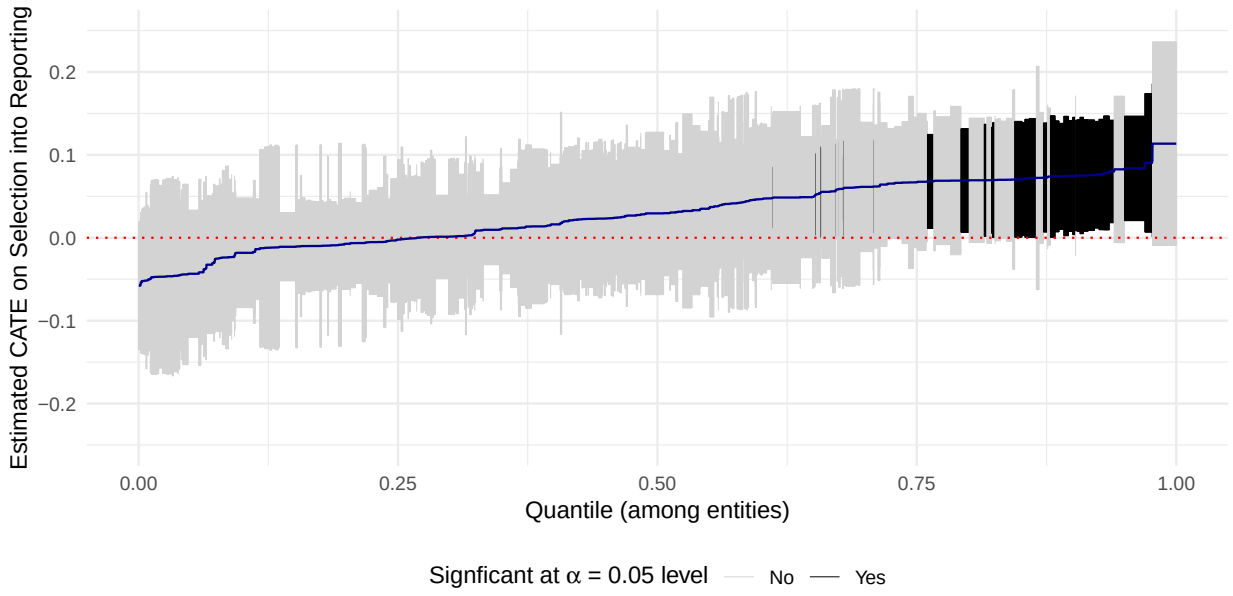


Figure A8: Predicted CATE for each entity ($n = 6,556$) estimated by generalized random forest estimator. Each interval represents a 95% confidence interval. The navy line is the ECDF of the CATEs.

assumption of monotonicity: this implies that no entity that would have reported without direct communication failed to report because of the direct communication. Under the assumption of monotonicity, there are three potential (causal) types in the data: always reporters (indicated by $j = A$), never reporters (indexed by $j = N$), and if-treated reporters (indexed by $j = T$). We denote the shares of each type as π_j where $\sum_j \pi_j = 1$.

Now, consider each term in $E[S(Z = 1)|R(Z = 1) = 1] - E[S(Z = 0)|R(Z = 0) = 1]$:

$$E[S(Z = 1)|R(Z = 1) = 1] = \frac{\pi_A}{\pi_A + \pi_T} E[S(Z = 1)|j = A]| + \frac{\pi_T}{\pi_A + \pi_T} E[S(Z = 1)|j = T] \tag{3}$$

$$E[S(Z = 0)|R(Z = 0) = 1] = E[S(Z = 0)|j = A] \tag{4}$$

Simplifying terms we can express $\mathcal{P}$ as:

$$\begin{aligned} \mathcal{P} = & \underbrace{\frac{\pi_A}{\pi_A + \pi_T} \left(E[S(Z = 1)|j = A] - E[S(Z = 0)|j = A]\right)}_{\text{Change in scores reported}} + \\ & \underbrace{\frac{\pi_T}{\pi_A + \pi_T} \left(E[S(Z = 1)|j = T] - E[S(Z = 0)|j = A]\right)}_{\text{Change in composition of reporters}} \\ = & \frac{\pi_A}{\pi_A + \pi_T} CATE + \frac{\pi_T}{\pi_A + \pi_T} \left(E[S(Z = 1)|j = T] - E[S(Z = 0)|j = A]\right) \end{aligned} \tag{5}$$

where $CATE$ is the conditional average treatment effect among always-reporters ($j = A$). We use Lee (2009) trimming bounds to bound $CATE \in [CATE_L, CATE_U]$. With these bounds, Equation (5) similarly provides bounds on the average score of if-treated reporters, $E[S(Z = 1)|T]$. Rearranging Equation (5) we derive the bounds:

$$E[S(Z=1)|T] = \Big[\frac{\mathcal{P} - \frac{\pi_A}{\pi_A+\pi_T}CATE_U + \frac{\pi_T}{\pi_A+\pi_T}E[S(Z=0)|j=A]}{\frac{\pi_T}{\pi_A+\pi_T}}, \\ \frac{\mathcal{P} - \frac{\pi_A}{\pi_A+\pi_T}CATE_L + \frac{\pi_T}{\pi_A+\pi_T}E[S(Z=0)|j=A]}{\frac{\pi_T}{\pi_A+\pi_T}}\Big] \tag{6}$$

Note that it is straightforward to calculate point estimates of $\mathcal{P}$, π_A, π_T, and $E[S(Z=0|j=A)]$. With these point estimates and the Lee (2009) bounds, $CATE_L$ and $CATE_U$, it is straightforward to bound $E[S(Z=1)|T]$.

Table A7 reports our interval estimates of the CATE of treatment on the scores reported by always-reporters. We generate uncertainty estimates using bootstrapping. Table A7 reports the 95$^{\text{th}}$ percentile of the upper bound estimates, $(Q_{95}(\widehat{CATE}_U)$. Because this decomposition employs ATEs instead of AMCEs in this analysis, we consider multiple definitions of treatment. In Panel A we consider the ATE of direct communication treatment (pooled over content) versus pure control, as in the top interval estimates in Figure 2. In Panel B, we consider treatment as direct communication and any individual manipulation of content versus pure control, as in the subsequent four estimates in Figure 2. In Panel C, we compare factorial treatment cells (individually) to control. Note that the number of observations in the treatment group decreases substantially in each set of panels. Because we want to test whether the upper bound of the interval estimate is negative, we assess the sign of the upper bound of the 95% confidence interval on the upper bound of our interval estimate. Note that in Panel C, the cell sizes are substantially smaller, rendering estimates much noisier.

A9 Audit Data: Ancillary Analyses

We conduct several ancillary analyses of the audit data, as described below:

1. Table A9 reports all coefficient estimates from the specifications reported in Table 4.

2. We examine the relationship between two measures of administrative capacity and transparency practices and reporting behavior, to assess how administrative capacity influences reporting dynamics, as follows:

    - **Municipal-level capacity**: We use the National Planning Department's (DNP) index of municipal performance to measure geographical variation in administrative capacity. This measure evaluates local government outputs, which cover a subset of the entities we study. We have this measure for 952 of the 953 municipalities represented in the audit, minimizing missingness.
    - **Entity-level capacity:**We use data from the Administrative Department of Public Administration's (DAFP) index of institutional performance to assess administrative capacity at the entity level. This index is constructed from self-reports (like the ITA) and is not required for all entities mandated to report ITA data, resulting in higher missingness—we have scores for only 1,210 of 2,400 audited entities.

    Table A8 reports the associations between these measures of municipal and institutional capacity and our measures of reporting behavior from the audit. In Columns 1-2, we show that the probability

Direct communication content				Lee Bounds		Upper bound
Past	Future	Reminder	Training	$\widehat{CATE}_L$	$\widehat{CATE}_U$	$Q_{95}(\widehat{CATE}_U)$
PANEL A: DIRECT COMMUNICATION (ANY)						
Any	Any	Any	Any	-7.97	-5.23	-1.37*
PANEL B: DIRECT COMMUNICATION ALONG EACH MARGIN						
✓	Any	Any	Any	-7.67	-4.59	-0.58*
Any	✓	Any	Any	-8.19	-5.65	-1.39*
Any	Any	✓	Any	-9.38	-5.60	-1.69*
Any	Any	Any	✓	-8.52	-6.14	-1.81*
PANEL C: DIRECT COMMUNICATION BY FACTORIAL CELLS						
				-2.98	-0.40	5.09
✓				-6.64	-6.64	0.56
	✓			-8.22	-4.64	1.11
✓	✓			-6.77	-2.43	3.00
		✓		-10.26	-5.66	-0.23*
✓		✓		-7.70	-1.37	3.43
	✓	✓		-6.94	-0.67	4.82
✓	✓	✓		-9.95	-9.95	-3.33*
			✓	-7.30	-7.30	0.46
✓			✓	-7.52	-4.22	1.36
	✓		✓	-6.95	-6.95	-0.27*
✓	✓		✓	-5.95	-2.58	3.12
		✓	✓	-11.07	-7.58	-1.79*
✓		✓	✓	-8.45	-3.40	1.97
	✓	✓	✓	-12.34	-9.17	-3.02*
✓	✓	✓	✓	-8.47	-8.47	-0.04*

Table A7: Lee trimming bounds on the conditional average treatment effect (CATE) among the principal stratum of always-reporter entities. $\widehat{CATE}_L$ and $\widehat{CATE}_U$ correspond to the lower and upper bounds, respectively. To rule out compositional effects in isolation, we test the one-tailed null hypothesis, H_0 : $\widehat{CATE}_U \geq 0$. The 95th percentile of the bootstrapped distribution of $\widehat{CATE}_U$, ($Q_{95}(\widehat{CATE}_U)$) allows us to test this hypothesis. * corresponds a rejection of H_0 at the $\alpha = 0.05$ level.

	Completed ITA $\mathbb{I}(r \neq \emptyset)$		Audit score (θ)		Distortion $(r-\theta)$		| Distortion | $(\mid r-\theta \mid)$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Muncipal capacity (standardized)	0.018^{+}		0.851***		0.673***		-0.655***	
	(0.009)		(0.232)		(0.146)		(0.128)	
Institutional capacity (standardized)		0.116***		2.572***		0.568**		-1.259***
		(0.011)		(0.279)		(0.217)		(0.186)
Num. Obs.	2400	2400	2400	2400	1696	1696	1696	1696
Sample	All	All	All	All	Completed	Completed	Completed	Completed
Indicator for missing capacity measure	yes	yes	yes	yes	yes	yes	yes	yes
DV mean | Capacity measure present	0.707	0.807	16.773	21.137	-0.119	0.341	3.678	3.298
DV std. dev | Capacity measure present	0.455	0.395	11.929	9.683	7.141	6.301	6.122	5.379
DV range	$\{0,1\}$	$\{0,1\}$	$[0, 27.75]$	$[0, 27.75]$	$[-27.75, 27.75]$	$[-27.75, 27.75]$	$[0, 27.75]$	$[0, 27.75]$

$^{+}p < 0.1$,* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$

Table A8: Associations between each of the capacity measures and measured parameters capturing reporting behavior. All regressions are estimated by OLS with heteroskedasticity-robust standard errors.

of completion of the ITA matrix increases in both measures of administrative capacity. Comparing the two estimates, a one standard deviation increase in municipal capacity increases the probability of reporting by 1.8 percentage points whereas a one standard deviation increase in institutional capacity increases the probability of reporting by 11.6 percentage points. Columns 3-4 show that audit-measured transparency scores are increasing in municipal capacity. Again, a one standard deviation increase in institutional capacity is associated with a audit score that is 2.57 points higher (on the 27.75 point scale), whereas a one standard deviation increase in municipal capacity is associated with only a 0.85 point increase in audit scores. Columns 5-8 report distortions in reporting among the entities that reported ITA transparency scores. First, note that Columns 6-7 document an *increase* in the degree of over-reporting as a function of administrative capacity. Thus, even though these entities already have higher compliance with the audited transparency practices (as shown in Columns 3-4), they over-report by a greater amount than low-capacity institutions. Ccolumns 7-8 show that despite this systematic over-reporting by higher-capacity entities, higher-capacity entities also report with *less noise* (greater accuracy) than low-capacity entities. These distortions include both intentional (d) and unintentional (ε) distortions. Combining the estimates in columns 5-6 and 7-8 suggests that lower capacity introduces large unintentional distortions that outweigh the score manipulations that we document in columns 5-6.

3. Table A10 considers the possibility of interactions between audit-measured scores on the transparency index and administrative data in determining reporting behavior. It estimates specifications of the form:

$$\begin{aligned} Y_i = &\beta_0 + \beta_1 \text{Audit Score}_i + \beta_2 \text{Capacity}_i + \beta_3 \text{Capacity Missing}_i + \\ &\beta_4 \text{Audit Score}_i \times \text{Capacity}_i + \beta_5 \text{Audit Score}_i \times \text{Capacity Missing}_i + \epsilon_i \end{aligned}$$

We report estimates of β_1, β_2, and β_4 in the table for each outcome (completion, distortions, and absolute value of distortions) and both capacity measures. The principal finding of this analysis is that at low levels of transparency (when the Audit Score is low, or near zero), municipal capacity increases noise (Columns 5-6). However, as the underlying level of transparency increases, noise in reporting decreases in municipal capacity.

4. Figure A9 decomposes the conditional standard deviation of reported scores by audit measured scores

(x-axis) and the two measures administrative capacity. Consistent with the idea that low-capacity entities face higher costs of effort, which would lead them to exert less effort, we observe higher-variance in the reported scores from these entities for all subgroups. However, even within the high- and low-capacity medians, we observe similar patterns in the conditional variance. Lower-transparency entities report scores with higher variance.

5. Table A11 reports the association between audit-measured quality and the probability of completion of the ITA matrix. It reveals substantial positive selection into reporting. A one standard deviation increase the audit-measured quality score (11.9 points) corresponds to a 15.19 percentage point increase in the probability of completing the ITA matrix. This association is substantially and significantly attenuated for entities led by elected principals. However, there is still modest positive selection for these entities.

6. We estimate the conditional average treament effects of the direct communication treatment at different levels of audit-measured transparency. Here, we estimate CATEs in four bins of audit scores, j. In (7), β_j serve as the estimators of the CATEs in each bin. We examine whether an entity reports and the distortion in reports (conditional on reporting) as the outcomes, Y_i.

$$Y_i = \sum_{j=1}^{4} \beta_j \text{Direct Communication}_i \times \text{Audit score in } j_i + \sum_{j=1}^{4} \gamma_j \text{Audit score in } j_i + \kappa \text{National entity}_i + \epsilon_i \tag{7}$$

A10 Simulating Government Use of Data

In Figure 7, we consider what the PGN might see under different audit-targeting policies, $\rho(p)$. In this section, we report the functional forms that we use for each form of targeting. Note that what varies in the analysis is how the $\rho(r)$ varies in r, not the absolute level or frequency of targeting. For this reason, we represent each non-zero rate of targeting scaled by some arbitrary constant, $k \in (0, 1]$.

- Targeting is not based on reported scores:
  - All entities:

$$\rho(r) = k$$

  - Non-respondents only:

$$\rho(r) = \begin{cases} k & \text{if } r = \emptyset \\ 0 & \text{else} \end{cases}$$

  - Respondents only:

$$\rho(r) = \begin{cases} k & \text{if } r \in \mathbb{R} \\ 0 & \text{else} \end{cases}$$

	Reported score on audited items			Total reported score		
	(1)	(2)	(3)	(4)	(5)	(6)
Audit score	0.509***	0.509***	0.504***	0.731***	0.733***	0.731***
	(0.025)	(0.025)	(0.025)	(0.073)	(0.073)	(0.075)
National Indicator	3.124***	3.039***	3.260***	17.323***	17.202***	17.298***
	(0.481)	(0.492)	(0.541)	(1.276)	(1.290)	(1.465)
Direct Communication		-0.742	-0.743		-4.177+	-4.174+
		(0.738)	(0.737)		(2.429)	(2.429)
Oversight of past completion		-0.290	-0.284		0.262	0.262
		(0.427)	(0.427)		(1.396)	(1.396)
Possible future audit		0.414	0.422		-0.226	-0.221
		(0.432)	(0.432)		(1.396)	(1.398)
Reminder		-0.548	-0.555		-3.145*	-3.141*
		(0.427)	(0.427)		(1.396)	(1.397)
Training		0.266	0.273		0.187	0.186
		(0.427)	(0.427)		(1.392)	(1.392)
Elected entity head indicator			0.451			0.202
			(0.435)			(1.460)
Intercept	9.886***	10.637***	10.535***	59.240***	64.274***	64.207***
	(0.607)	(0.804)	(0.809)	(1.767)	(2.433)	(2.455)
Num. Obs.	1307	1307	1307	1696	1696	1696
Adjusted R^2	0.339	0.339	0.339	0.121	0.125	0.124

$^{+}p < 0.1,^{*}\, p < 0.05,^{**}\, p < 0.01,^{***}\, p < 0.001$

Table A9: This table reports all coefficients estimated by the specification reported in Table 4. Heteroskedasticity robust standard errors in parentheses.

- Targeting is based on reported scores
  - Target zero scores

$$\rho(r) = \begin{cases} k & \text{if } r = 0 \\ 0 & \text{else} \end{cases}$$

  - Targeting low (non-zero) scores

$$\rho(r) = \frac{rk}{100}$$

  - Targeting middling scores

$$\rho(r) = k\left(\frac{50 - |r - 50|}{100}\right)$$

  - Targeting high (non-perfect) scores

$$\rho(r) = \frac{k(100 - r)}{100}$$

	Completed ITA $\mathbb{I}(r \neq \emptyset)$		Distortion $(r - \theta)$		| Distortion | $(\mid r - \theta \mid)$	
	(1)	(2)	(3)	(4)	(5)	(6)
Audit score (0-1 scale)	0.373***	0.175***	-7.685***	-10.196***	-1.938***	-5.156***
	(0.021)	(0.039)	(0.471)	(0.877)	(0.457)	(0.844)
Muncipal capacity (standardized)	-0.076***		1.030*		1.106**	
	(0.018)		(0.412)		(0.390)	
Institutional capacity (standardized)		0.138***		0.542		0.786
		(0.029)		(0.825)		(0.790)
Audit score × Municipal capacity	0.132***		0.040		-2.240***	
	(0.022)		(0.464)		(0.447)	
Audit score × Institutional capacity		-0.056		0.877		-2.210**
		(0.034)		(0.890)		(0.855)
Num. Obs.	2400	2400	1696	1696	1696	1696
Indicator for missing capacity measure	yes	yes	yes	yes	yes	yes
DV mean | Capacity measure present	0.707	0.807	-0.119	0.341	3.678	3.298
DV std. dev. | Capacity measure present	0.455	0.395	7.141	6.301	6.122	5.379
DV range	{0, 1}	{0, 1}	[-27.75, 27.75]	[-27.75, 27.75]	[0, 27.75]	[0, 27.75]

$^{+}p < 0.1,^{*}\, p < 0.05,^{**}\, p < 0.01,^{***}\, p < 0.001$

Table A10: This table investigates interactions between audit-measured transparency scores (rescaled to a 0-1 interval) and both measures of municipal capacity. All regressions are estimated by OLS with heteroskedasticity-robust standard errors.

– Targeting perfect scores

$$\rho(r) = \begin{cases} k & \text{if } r = 100 \\ 0 & \text{else} \end{cases}$$

Supplementary Appendix: References

American Political Science Association. 2020. "Principles and Guidance for Human Subjects Research." Available at `https://tinyurl.com/y5vm6cem`.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *The Annals of Statistics* 47 (2): 1148–1178.

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

Figure A9: Disaggregation of Figure 5 by medians of the municipal and institutional capacity measures. This figure decomposes the conditional standard deviation in reports by entities above and below the median of state capacity (right).

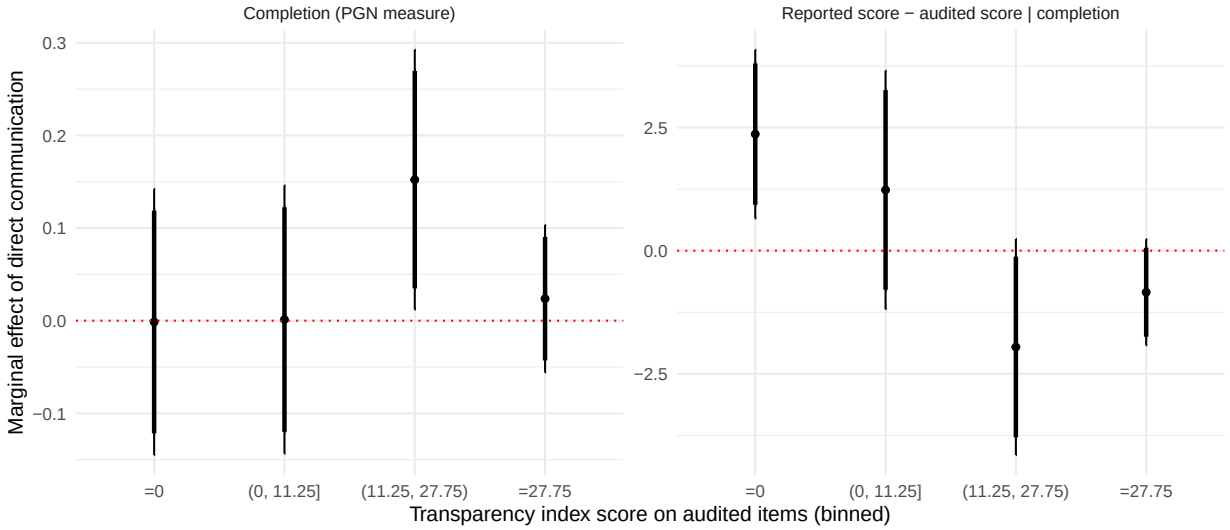


Figure A10: Estimates CATE of direction communication treatment by latent levels of transparency across audited entities in experimental sample. 95% confidence intervals are reported. Note that the right panel conditions the sample on submission, so care should be exercised in interpretation of these estimates as CATEs.

	Completed ITA Matrix (×100)					
	(1)	(2)	(3)	(4)	(5)	(6)
Audit score	1.292***	1.274***	1.562***	1.524***	1.258***	1.258***
	(0.078)	(0.079)	(0.094)	(0.098)	(0.080)	(0.080)
Elected			35.873***	36.210***		
			(3.754)	(3.759)		
Audit Score : Elected			-1.282***	-1.245***		
			(0.163)	(0.165)		
National government entity					11.683	11.683
					(9.699)	(9.699)
Audit Score : National					-0.038	-0.038
					(0.378)	(0.378)
Intercept	48.999***	48.809***	41.332***	40.994***	48.655***	48.655***
	(1.741)	(1.750)	(1.945)	(1.956)	(1.769)	(1.769)
Num.Obs.	2400	2400	2400	2400	2400	2400
Adjusted R^2	0.114	0.111	0.150	0.149	0.119	0.113
Reweighted		✓		✓		✓

+ p $<$ 0.1, * p $<$ 0.05, ** p $<$ 0.01, *** p $<$ 0.001

Table A11: Selection into reporting as a function of entity characteristics. The outcome is rescaled by a factor of 100 to facilitate legibility of estimates. This means that a coefficient of "1" refers to a marginal effect of one percentage point. Reweighting indicates the use of weights equivalent to the inverse of the probability of sampling. Heteroskedasticity-robust standard errors in parentheses.