IDB WORKING PAPER SERIES Nº IDB-WP-1160

Curbing Grand Corruption in the Contracting Out of Public Services

Lessons from a Pilot Study of the School Meals Program in Colombia

Philip Keefer
Benjamin Roseth

Inter-American Development Bank
Institutions for Development Sector

June 2021

Curbing Grand Corruption in the Contracting Out of Public Services

Lessons from a Pilot Study of the School Meals Program in Colombia

Philip Keefer
Benjamin Roseth

June 2021

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Keefer, Philip.
Curbing grand corruption in the contracting out of public services: lessons from a pilot study of the School Meals Program in Colombia / Philip Keefer, Benjamin Roseth.
p. cm. — (IDB Working Paper Series ; 1160)
Includes bibliographic references.
1. Contracting out-Corrupt practices-Colombia. 2. Political corruption-Colombia.
3. Public contracts-Corrupt practices-Colombia. I. Roseth, Benjamin. II. Inter-American Development Bank. Institutions for Development Sector. III. Title.
IV. Series.
IDB-WP-1160

http://www.iadb.org

Abstract[*]

Do targeted transparency interventions reduce corrupt behavior when corrupt actors are few and politically influential; their behavior imposes small costs on numerous individuals; and corrupt behavior is difficult to observe? Results from a study of informal audits and text messages to parents, meant to curb corruption in the School Meals Program of Colombia, suggests that they can. Theory is pessimistic that transparency interventions can change the behavior of actors who exert significant influence over supervisory authorities. Moreover, inherent methodological obstacles impede the identification of treatment effects. Results substantiate the presence of these obstacles, especially considerable spillovers from treated to control groups. Despite spillovers, we find that parental and operator behavior are significantly different between treatment and control groups. Additional evidence explains why operator behavior changed: out of concern that systematic evidence of corrupt behavior would trigger enforcement actions by high-level enforcement agencies outside of the political jurisdictions where they are most influential.

JEL Codes: D73, H4, H42
Keywords: corruption, audits, contracting out

---

[*] Author contacts: pekeefer@iadb.org; broseth@iadb.org.

This work would not have been possible without the significant contributions of Ruth Persian, Luke Ravenscroft and Monica Wills-Silva from the Behavioral Insights Team (London), the dedicated field work and data collection efforts of Santiago de la Cadena and Tatiana Rodríguez and the collaboration of colleagues from the Government of Colombia, Javier Guillot and Aura Cifuentes, Camilo Cetina, Paula Castañeda, Daniel Gómez, Luis Fernando García, Orlando Chacón, Tathiana Sánchez, Janyther Guerrero, Erika Rincón and María Francisca Sanín. Adelaida Correa, Daniel Jaramillo and Miguel Purroy provided truly invaluable research assistance. Similarly invaluable were the comments and suggestions of Cesi Cruz, Paul Lagunes and Miguel Purroy. We also thank the UK's Foreign and Commonwealth Office for funding the Behavioral Insights Team's contribution to the project and the additional funding and institutional support of the Inter-American Development Bank and Secretaría de Transparencia, the Departamento Nacional de Planeación, the Ministerio de Educación, and the Ministerio de Tecnologías de Información y Comunicación of the Government of Colombia.

Introduction

Do targeted transparency interventions reduce corrupt behavior when there are few, politically influential corrupt actors; the costs of corruption are spread over many individuals, each of whom bears a small cost; and corrupt behavior is difficult to observe? An important body of research has focused on the inverse case, finding that transparency reduces corruption when corrupt actors are numerous and not influential; victims individually bear a relatively large cost; and corrupt behavior is easier to observe. This paper uses evidence from a pilot study, originally designed as a randomized control trial,[1] of a transparency intervention in more than 200 schools that participate in Colombia's School Meals Program (Programa de Alimentación Escolar, or PAE) to explore the significant substantive and methodological implications that these three dimensions have for transparency interventions.

Substantively, theory suggests that targeted transparency[2] interventions should have a larger effect on the behavior of corrupt actors when the information the interventions reveal is more likely to trigger action by victims and when corrupt actors are less able to resist oversight. Prior research typically focuses on this context. The pilot study discussed here yields evidence that suggests transparency interventions—informal audits of the meals served to schoolchildren and text messages to parents—can also reduce corrupt activity in less promising settings. Methodologically, evaluations of transparency interventions to curb corrupt actions by a few are particularly subject to spillover effects that exert a downward bias on estimated impacts. The pilot study yields evidence of the importance of spillover effects in contexts with few corrupt actors but, nevertheless, also reveals significant differences in corrupt behavior between schools exposed to transparency interventions and control schools.

Additional evidence is consistent with two possible mechanisms, the "bottom-up" and the "top-down". An end-line survey reveals that parents in treated schools are much more likely to report participation in school meetings that concern the meals program. In addition, though, the differences between treated and control schools are strongest in the department where national oversight agencies were previously absent, suggesting that operators are specifically concerned about not drawing the attention of those agencies.

[1] As discussed in greater detail below, the pre-analysis plan was designed based on information that we could gather about the modalities of PAE shirking through extensive pre-analysis fieldwork. This information was difficult to gather, ultimately incomplete, and led to the specification of inappropriate tests in the PAP. We present the results of tests that reflect actual operator practices, gleaned from the study itself. However, since these were not pre-specified, we describe the study as a "pilot" and the results as preliminary.

[2] "Targeted transparency" is "the use of publicly required disclosure of specific information in a standardized format to achieve a clear public policy purpose" (Fung, Graham, and Weil, 2007).

Holding constant the total rents extracted across corrupt transactions, a larger number of victims increases the collective action problems of holding actors accountable after revelations of corrupt behavior. Fewer, politically influential corrupt actors, in contrast, can more easily block the implementation of transparency measures, and the more susceptible is the measurement of corrupt behavior to manipulation, the greater the difficulty of implementing transparency measures.

These characteristics are precisely those of the PAE. Departments (states) are responsible for contracting and oversight of PAE operators, but within each department only a few operators provide most school meals. They serve thousands of students who individually bear small losses from operator shirking. Measurement of shirking is susceptible to operator manipulation.[3] Consistent with their political influence, the legislation and oversight of the school meals program exhibit numerous characteristics that make it difficult to prevent operator shirking on their contractual responsibilities. Finally, measurement of corrupt behavior requires on-site monitoring of meals, which operators can observe and to which they can respond strategically. The pilot study revealed ample evidence of this.

These same features also introduce serious barriers to evaluation. The presence of only a few operators requires that treatments be randomized across beneficiary units (schools). Unavoidably, control and treatment schools are served by the same operators. The pilot study yields evidence of significant spillover effects from treatment to control schools. Measurement of operator shirking in the provision of school meals is imprecise, making it advantageous to make multiple assessments. However, because operators respond rapidly to measurement, multiple assessments of school meals in control schools trigger treatment effects in those schools, as well. We find significant evidence of operator responses to the very first informal audit of meals in control schools, at the end of the intervention. Despite the theoretical reasons for which transparency

[3] From the Sistema Electrónico de Contratación Pública (SECOP) of Colombia Compra Eficiente (CCE): https://www.colombiacompra.gov.co/transparencia/conjuntos-de-datos-abiertos (last accessed 3/18/2021), we know that the department of Cesar issued nine million dollars in contracts to *PAE* operators in 2017. The two operators who served the schools in the Cesar sample of the pilot study accounted for 59 percent of the total. The Union Temporal Alimentación del Cesar accounted for 19.4 percent of the *PAE* contract dollars in Cesar and served 34 schools in the sample. The Consorcio PAE Valledupar accounted for 39.2 of the contract dollars and served 52 of the sample schools.
The department of Nariño issued approximately $15 million in contracts, 62 percent of which went to the two operators that served the schools in the Nariño sample. The Union Temporal PAE de Nariño accounted for most of this, 59.4 percent of total contract value in the department, and also for most of the schools: 117 of the schools in the sample. Nariño has many isolated communities that offer less lucrative opportunities and are served by many small operators that collectively account for a fraction of total contracts. They are more similar to the corrupt actors who are the subject of prior research. However, only nine of the schools in the Nariño sample were served by such an operator, La Asociación de Cabildos Indígenas del Pueblo de los Pastos. The value of this group's PAE contracts with the department accounted for 2.5 percent of the total. Three of the schools in the sample receive services from two operators (one, for example, providing the morning meal and another the afternoon).

interventions should matter less in this context, and the methodological obstacles to detecting treatment effects, the pilot study still reveals significant differences in the behavior of school meal providers between schools exposed to transparency interventions and control schools.

Corruption by a few actors involving large sums is often labelled “grand corruption.” The pilot study yields numerous lessons for future research and policy reforms related to grand corruption. First, despite the large differences with the corruption examined in the transparency literature, transparency interventions show promise in combatting grand corruption. The pilot study suggests that such an outcome may result from an interaction between bottom-up efforts by parents to hold operators accountable and the prospect of top-down attention of powerful enforcement agencies that is drawn to the systematic information about operator performance collected through the informal audits.

Second, in the absence of formal government audits that are the mainstay of previous research, low-cost, informal strategies to systematically measure operator shirking are available. Those strategies are necessary in the case of shirking in service delivery: operators delivering meals for the PAE can engage in small amounts of shirking across tens or hundreds of thousands of transactions.

Third, even in a context in which governments are acutely concerned about corruption in a sector, low-cost, informal strategies to measure shirking seem not to emerge in equilibrium. Hence, the organization of the informal collection of systematic information regarding behavior should be an object of transparency reform.

Fourth, large service providers in corrupt settings are well-organized and respond quickly to observation by outsiders; in many cases, the response takes the form of efforts to prevent such observation. This raises significant obstacles to the evaluation of policy efficacy. However, and despite such obstacles, the pilot collection of systematic, though unofficial, information about operator performance, combined with text communications with parents, appears to have significantly increased operator compliance with their contractual obligations. However, the presence of an active prosecutor may be key to the success of such initiatives.

The next section of the paper outlines the theoretical importance for transparency interventions of the distinctions between the corruption context we examine in the Colombia school meals program and the types of corruption that are the focus of prior research on transparency. We then identify the methodological challenges that are inherent in the analysis of corrupt behavior in programs such as the PAE. In succession, after describing in more detail the PAE itself, the specific transparency interventions that we implemented, and the design of the pilot study, we present qualitative and quantitative evidence on the theoretical and methodological issues raised by the study of "grand corruption" in service delivery, including information on the mechanisms through which the interventions might improve operator performance.

Theory and Literature

Transparency reform rests on straightforward theoretical foundations: at least since Becker and Stigler (1974), it has been well-understood that agents are less likely to shirk on their obligations to principals, including engaging in corrupt behavior, when principals are better informed about agent behavior (see also, for example, Lui 1986). However, the impact of greater information depends on collective action. If the principal is a country's citizens, control of the agent is a function not only of the principal's information about agent behavior, but also of the principal's—citizens'—ability to overcome the collective action dilemma inherent in monitoring the agent (Persson, Rothstein and Teorell 2013). The agent's ability to evade accountability is also a function of collective action: many agents engaged in small acts of corruption must organize to resist anti-corruption efforts; a small number of agents who each extract large rents can act unilaterally to insulate themselves from oversight.

The impact of transparency interventions also depends on the costs of collecting information about agent performance and agent ability to manipulate the information collection process. If information is collected on only some dimensions of agent effort, agents may increase their shirking on others (see the review in Banerjee, Mullainathan, and Hanna 2012). Less widely recognized is the fact that some types of effort are inherently more observable than others. Some agent behavior – for example, related to the provision of services, such as school meals – leaves no lasting traces and must be measured at the time that it occurs. In contrast, theft or cheaply built infrastructure leave permanent traces of corrupt behavior. Some types of agents are also better able to manipulate the information collection process.[4] These considerations suggest that transparency initiatives should be systematically less effective in settings where, holding constant total rent seeking, there are few corrupt agents and many victims, and the evidence of shirking is more costly to collect.

Ample evidence supports the basic logic of transparency reform – more information triggers greater accountability and lower corruption. However, it does so in contexts where the considerations described above are less salient: government agencies collect information and use the information to decide on possible prosecution,

[4] Many creative efforts to measure corruption, such as those reviewed in Banerjee, Mullainathan, and Hanna (2012), refer to cases in which there are many agents (e.g., service providers in health clinics) with little capacity to avoid observation.

solving the collective action problems of numerous victims; corrupt actors are numerous and cannot easily block the collection of information about their behavior; and evidence of shirking is straightforward to collect.

For example, school directors in Uganda spend central government transfers more carefully when citizens are informed about these transfers (Reinikka and Svensson 2005). The large number of school directors means that they cannot easily lobby to make information about their behavior difficult to observe. At the same time, the relatively large sums involved at each school mean that parents, although numerous, have a stronger incentive to collect evidence and hold directors accountable. Once informed about the amount of transfers schools should receive, given their importance in the school budget, parents and government auditors can easily infer malfeasance by observing (the lack of) spending. Similarly, those public investment projects in Costa Rica that were published on a georeferencing website that facilitated public comments exhibited faster physical and financial execution than non-published projects (Rossi et al, 2021).

In other studies, the collective action challenges confronting many victims are solved by the direct involvement of central government auditors with the power to sanction malfeasance. This mechanism is likely at work in the transparency interventions we explore in the Colombian school meals program. However, in prior studies, in contrast to the PAE, the number of agents is large and information itself is relatively cheap to collect. Mayors of small Indonesian towns curb corrupt behavior in the construction of small, centrally funded infrastructure projects when they know that audits of these projects will be sent to the central government (Olken 2007). So also do hospitals in Buenos Aires (Di Tella and Schargrodsky, 2003). Municipalities in Brazil improve their performance after the publication of central government audits that reveal financial management failures (Avis, Ferraz and Finan 2016)—or, in the case of Puerto Rico, before elections (Bobonis, Cámara and Schwabe 2016). Muralidharan et al. (2019) show that beneficiary farmers are significantly more likely to receive benefits from a large government cash transfer program when officials are informed that the national government will monitor their performance by making phone calls to beneficiaries.

In all these cases, the corrupt actors are numerous and incapable of blocking information collection. The central government, a unitary actor, internalizes the costs of auditing service providers. And in each case, a single piece of evidence (e.g., a single failed audit) can reveal significant malfeasance sufficient to trigger judicial or electoral responses, or both. Results in Lagunes (2018) follow the same logic: he finds that informing Peruvian mayors that civil society groups, supported by the country's anti-corruption agency, are monitoring specific public works projects appeared to reduce project costs. Purroy (2021) argues that a reform of the royalty allocation system in

Colombia reduced municipal corruption by requiring greater transparency in municipal contracting.

Past research does not capture the impact of transparency when a few actors extract large rents from many, typically (though certainly not always) entailing small losses for affected individuals. In contrast, in the school meals program in Colombia only a few, influential operators are responsible for service provision (school meals). They can (and as we show below, do) more easily block information collection. Second, many households are harmed by operator shirking, but individual household losses are small. Hence, the efficacy of transparency initiatives is more likely to depend on the existence of government agencies that solve citizens' collective action problem. Third, in part because of operator behavior and in part because of the intrinsic nature of the services operators provide, information on provider behavior is costly to collect. Evidence of shirking evaporates after the service is provided; shirking can take place on many dimensions; and many service transactions must be observed to distinguish shirking from innocent explanations, such as exogenous shocks such as weather, the number of children absent, and interruptions in wholesale food markets.

The analysis here and the research described above focus on the impact of transparency interventions based on direct evidence of malfeasance. Recent, pathbreaking literature might suggest that such direct evidence is unnecessary. It has exploited machine learning to identify corrupt transactions based on administrative characteristics associated with those transactions, such as price or bidding modality. For example, Dávid-Barrett and Fazekas (2019) use Big Data to identify the influence of political connections on the awarding of government contracts in Hungary and the United Kingdom. Gallego, Prem and Vargas (2020) use a machine learning algorithm to identify Colombian municipalities prone to corruption prior to the Covid-19 pandemic as part of a difference-in-differences identification strategy to associate corruption tendencies with procurement patterns. This work builds on prior research by substituting quantitative assessments of corrupt environments for subjective evaluations.[5] It leaves open the question of whether information about corruption derived from Big Data is sufficient to trigger accountability reactions by voters or judicial actors. Until such research is undertaken, measures of corruption based on direct observations of agent behavior will still be essential, if for no other reason than the fact that for many public services, administrative data do not capture the individual transactions through which corrupt actors earn rents.

[5] Prominent perceptions-based studies analyzing the determinants of corruption include Mauro (1995), Treisman (2000) and Knack and Keefer (1995).

Methodological Challenges of Evaluating Transparency Interventions in Theoretically Adverse Conditions

In those conditions where transparency interventions are most needed, but potentially least effective—when government is least likely to collect information on malfeasance and corrupt actors face the lowest risk of sanction—they are also most difficult to evaluate. These same conditions, the presence of only a few agents, of many principals, and of measurement challenges, also pose significant methodological challenges to evaluation.

All studies evaluating the impact of transparency interventions on corrupt behavior take agents as their unit of observation (e.g., mayors, hospitals, and individual public officials). In theory, the large number of agents raises the chances that the transparency intervention will succeed. It also facilitates evaluation: the sample size is sufficiently large so that when the actors are randomized between treatment and control groups, studies have sufficient power to detect a treatment effect. In contrast, when corrupt activity is controlled by a small number of actors (i.e., when it is grand corruption), this standard research design is infeasible. The only alternative is to randomize the assignment of locations where potentially corrupt transactions occur, such as the schools or health facilities served by potentially corrupt actors. The possibility of spillover effects from treatment to control groups are significantly higher.

Spillover effects are of course possible when there are many agents. Those in the control group could observe the intervention in the treatment group and change their behavior. There is a standard design response to this problem: sampling agents who cannot easily communicate with each other. The theoretical response is more compelling, however. It is unlikely to be rational for agents in the control group to change their behavior, since the control group is characterized precisely by the absence of an intervention to monitor their behavior. Since they confront no increased probability of sanction, they have no reason to reduce the risk of sanction by modifying their behavior.

A small number of agents precludes the standard design response of avoiding communication between control and treatment agents. In addition, however, agents who observe that their behavior is being monitored among some of the beneficiaries they serve have greater reason to believe that the services they provide to all their beneficiaries might be scrutinized. Agents are more likely to change their behavior towards control group beneficiaries for a period if the rents they forego over that period are low relative to the magnitude of the sanctions they might expect if they are detected, and if they believe that additional information that voters or judicial authorities might obtain from control schools will have a significant effect on voter or judicial attitudes towards them.

The grand corruption often involves few actors, and the arguments above immediately apply. However, the theoretical mechanisms through which transparency affects agent behavior always exerts downward pressure on the number of agents that can be included in an evaluation when the agents extract tiny rents from many beneficiaries. In this case, evidence of shirking from a few transactions is insufficient to motivate either voters or judicial authorities to act, in part because any small sample of transactions could exhibit deficiencies for reasons other than agent shirking. Persuasive evidence of corrupt behavior, sufficient to elicit a response from voters or judicial authorities, therefore requires information on many transactions affecting many beneficiaries. A small number of observations is sufficient to fully characterize the behavior of agents in prior research. However, when agent shirking extracts small rents from many beneficiaries, many transactions must be observed. Hence, an evaluation budget that would accommodate a large sample of agents when corrupt behavior can be inferred from low-cost monitoring accommodates many fewer when monitoring is high cost.

Compared to the standard case in the literature, therefore, it is not possible to design evaluations to preclude spillover and the incentives of agents in the control group to react to knowledge of the treatment is greater. This raises an important question that can only be answered empirically. Are spillover effects in this context large enough to prevent the identification of treatment effects? We investigate this issue below with data from the pilot study.

Spillovers are not the only methodological challenge that distinguish the evaluation of transparency interventions in service delivery programs from evaluations in past research. In the studies reviewed earlier, researchers had no contact with the control group and could rely on administrative data to track impact. In the case of service delivery, however, there are no administrative records of the quantity or quality of those transactions, nor do they leave lasting traces, unlike public works projects since services are consumed when they are supplied. Measurement of corrupt behavior requires direct observation of the service delivery transaction. Under these conditions, the measurement of corrupt behavior becomes itself a potential transparency intervention.

Since such measurement must be undertaken in both treatment and control groups, a fundamental obstacle to evaluation arises; measurement of corruption in the control group can be expected to affect agent behavior. As with the spillover effects discussed above, this measurement challenge also injects introduce downward bias downwards the estimate of treatment effects. Data from the pilot study discussed below provides some insight into the extent of that downward bias and how to control it.

Transparency Interventions in Colombia's School Meals Program (PAE)

The PAE is the oldest and largest social service program in Colombia, launched in 1936 to foster better school attendance, support healthy eating, reduce attrition, and improve overall academic performance for the poorest children in the country.[6] Through the PAE, the Ministry of Education provides federal funding to finance one or two meals per day (morning or afternoon snacks and/or lunch) for needy children. The program reaches approximately 60 percent of the students enrolled in public schools, with an annual budget of nearly US$600 million in 2017.

Colombia's 96 administrative organizational bodies (certified territorial entities or ETCs) implement the program. The ETCs consist of all departments plus those municipalities able to meet the certification requirements. Some schools and municipalities provide extra budgetary support to the program to broaden access and many schools provide meals to all children, beyond those meeting eligibility criteria.

Primary responsibility for the oversight of PAE operators in Colombia falls to the department governments that contract them. Within departments, the largest operators that provide the bulk of school meals are few and influential. This contrasts with the schools and municipalities investigated in earlier research: national agencies, over which school and municipal officials exercised little political influence, had primary oversight responsibility for them and subjected them to regular audits. Although PAE operators fall within the jurisdiction of national oversight agencies in Colombia—the Controller General of the Republic (Contraloría) and the Attorney General (Fiscal de la Nación)—these entities target PAE operators infrequently and on an ad hoc basis. In contrast, they permanently monitor municipalities. One trigger for ad hoc targeting could be the revelation of systematic information about operator under-performance. We provide suggestive evidence that operator concern about this effect contributes to their change in behavior after experiencing the transparency intervention.

All meals served must meet detailed requirements established by the Federal Ministry of Education. Operators bid for the concession to provide these meals. They can prepare meals off-site and deliver them ready-made to the school, or they can deliver

[6] Evidence on the educational impact of school feeding programs is mixed, but studies generally do not account for the way in which the programs are administered. In their meta-review, Jomaa, McDonnell and Probart (2011) conclude that school feeding programs improve school enrollment and attendance, but find little evidence that they raise growth, cognition and academic achievement. Alderman and Bundy (2012) similarly find that they are best viewed as transfer programs that can provide a social safety net and help promote human capital investments.

ingredients to schools and arrange for on-site preparation by operator employees, often mothers with children in the schools.[7]

Operators can shirk in many ways on their contractual obligations: they can provide too few meals, serve small portion sizes, or substitute low-cost ingredients not included in sanctioned menus. Scandals erupted in 2016 and led to renewed concerns about corruption in the program. An explosive video from a school in the town of Aguachica, in Cesar, triggered a national uproar and provided the impetus for the government to undertake this project. A teacher secretly filmed the production of a promotional video for the school's meals program. The promotional video showed smiling children with full plates; the clandestine video revealed that the same full plate was passed from child to child for purposes of the video. When filming was completed, the children received their usual meager rations, served directly into their hands. Anonymous threats forced the teacher to leave the country.[8]

The Colombian government requested assistance to develop a response that could be implemented rapidly, without new legislation or budgetary commitments, and agreed to a two-pronged transparency intervention. One intervention consisted of multiple, short-duration, informal audits that yielded systematic—albeit not legally actionable—information on operators' contractual compliance. The other was a program of nine weekly text messages to parents informing them about, among other things, the food items their children should have received on the day of the message.

Theory predicts that the audit intervention could affect operator behavior through two possible channels. First, even if operators are insulated from departmental oversight, they could be concerned that the systematic collection of compliance data, even if undertaken informally by university students and with no legal effect, could lead the Controller General of the Republic (Contraloría) or the Attorney General (Fiscal de la Nación) to divert enforcement resources from other priorities and target them. Both entities are active in enforcing anticorruption laws, but their resources are limited. As discussed above, they cannot easily target investigative resources when malfeasance consists of shirking in many transactions (individual meals to students) that are difficult to observe.

Second, newly informed parents might register more systematic complaints with officials than had previously been possible, prompting either an administrative and

[7] The Ministry of Education describes the program here: http://www.mineducacion.gov.co/1759/w3-article-349942.html [Accessed 9/16/16].

[8] A news report, including the video, can be found here https://www.youtube.com/watch?v=wFVcHyqL3_I [Accessed 1/15/19]. Corruption in the program has been widely covered in Colombia newspapers, for example here: https://www.eltiempo.com/archivo/documento/CMS-16568213 [Accessed 1/15/19].

judicial reaction or a political reaction (e.g., from mayors who had previously tolerated operator behavior). Evidence from Reinikka and Svensson (2005) and Enikolopov, Petrova and Sonin (2018) provides some support for this mechanism. Reinikka and Svensson (2005) show that information about central government transfers to local schools in Uganda significantly increases the fraction of those transfers that reach the schools. In Enikolopov et al., blog posts exposing corruption in Russian state-controlled companies reduced their market returns. Managers of these firms, uncertain about the degree to which their judicial and political connections would insulate them from legal consequences, might therefore have curbed corrupt behavior that had allowed them to boost their profitability.

The evidence that grassroots or bottom-up pressure is sufficient to change the behavior of corrupt actors is, however, mixed. Olken (2007) finds that grassroots participation in monitoring of local infrastructure projects increased citizen knowledge of cost overruns but had no effect on corruption. Cruz, et al. (2018) show that merely informing citizens of funds that mayors have at their disposal is sufficient to change behavior, but in unexpected ways: officials found to have shirked on their obligations offset voter disappointment by increasing vote buying in the short run. Weitz-Shapiro and Winters (2016) conclude that voter reactions to information campaigns depend on the credibility of the information; experimenter-generated text messages may or may not be regarded as credible by recipients. Finally, operators in Colombia are more influential in the political jurisdictions within which they operate (departments) than the actors examined in previous research. They may therefore feel more insulated from parental pressure than the public sector officials that are the focus of prior research.[9]

Pilot Design

The interventions were piloted in an exploratory study that included 208 schools in 60 municipalities in two departments. The study was exploratory because key parameters, fundamental for the analysis, could not be defined prior to the interventions. One of these was the dependent variable. Prior to the study, despite substantial field work it was not possible to ascertain the exact dimensions of food provision on which operators would shirk. Another was the speed of operator response to the informal audits, making it

[9] Our intervention does not concern government procurement itself, although this is naturally a central concern in the contracting out of government services (see Broms, et al. 2017 for recent work on the effect of elections on procurement in Sweden). Indeed, some aspects of rent-seeking that we observe are facilitated by contracts that make fiduciary oversight by the government costly and cumbersome. We similarly ignore the electoral effects of corruption, amply documented in Ferraz and Finan (2008), as well as Costas-Pérez, Solé-Ollé and Sorribas-Navarro (2012) and reviewed in De Vries and Solaz (2017).

impossible to specify ex ante whether one, two or three end-line audits in control schools would constitute more accurate measures of counter-factual operator performance.

As the earlier discussion indicated, operators have little incentive to react to a small number of audits affecting a tiny fraction of their schools. Consequently, proof of concept required that we include a large fraction of the schools served by the operators in the study. Further recalling the earlier discussion, theory precluded the usual evaluation design, randomizing the treatment across operators. Instead, the treatment (both interventions) was assigned to 106 schools, 64 in the department of Nariño and 42 in the department of Cesar.[10] In Nariño, treated schools were audited approximately 7.5 times over the course of the semester. A teacher strike shortened the semester in Cesar, where schools were audited approximately five times. The audits were unannounced. The schedule of audits was known only to the research team.[11]

To assess operator compliance with their contractual obligations in control schools, the pilot design included three light audits at the end of the semester. Logistical difficulties in organizing the auditors and the strike-shortened semester in Cesar meant that in about half of control schools, only two audits were conducted. The choice to conduct multiple end-of-semester audits in control schools was made to reflect the tradeoff between greater accuracy—multiple audits allow for more accurate inferences from noisy measures of operator compliance—and the risk of triggering treatment effects in control schools. In fact, we find that the tradeoff was steep: operators appeared to respond rapidly to the first of the audits by improving compliance before the second end-line audit had occurred.

Table 1 details the number of audits in control and treated schools in the two departments. Out of 125 sample schools in 38 municipalities of Nariño, 64 were randomly assigned to the treatment group and 61 to the control. Treated schools received an average of 7.5 audits over the course of 11 weeks. Audits to collect end-line information on control schools took place in weeks 9, 10 and 11; control schools were audited on average 2.8 times in Nariño.[12]

---

[10] The design and implementation of the intervention was the product of collaboration among the Behavioral Insights Team of Great Britain, the Department of National Planning, the Secretariat of Transparency in the Office of the President, the Ministry of Education, the Ministry of Technology, Information and Communications, and the Inter-American Development Bank.

[11] The education departments of Nariño and Cesar were given general information that there would be visits occurring in some schools during the second semester, but not which schools, nor the schedule. The operators were given no information directly. Parents in treated schools were told in workshops that audits were going to be conducted approximately once weekly, but they were not given specific dates. Despite the lack of details, it was clear to the student auditors that operators were aware of their visits. Early in the semester, apparently prompted by information from kitchen workers or other school employees, operator employees occasionally appeared at the schools and asked the students what they were doing.

[12] Different logistical challenges led to a lower pace of audits in the first weeks, which accelerated as the semester went on.

In Cesar, 42 schools received the treatment and 41 schools were in the control arm, across 22 municipalities. The shortened semester left only seven weeks for the intervention.[13] Still, the average treatment school in Cesar was audited five times and the average control school 1.5 times.

Table 1. Audits Summary

	Nariño	**Cesar**
Total schools	126	83
Treatment schools	64	42
Control schools	62	41
Municipalities	38	22
Kickoff workshops	55	39
Audits of treatment schools	477	234
Average audits per treatment school	7.45	5.57
Audits of control schools	179	74
Average audits per control school	2.89	1.80

To increase parental sensitivity to text messages, the pilot plan included kickoff workshops in treatment schools. These also turned out to pose logistical challenges: no workshop could be held in three of 42 treatment schools in Cesar and nine of 55 treatment schools in Nariño.

The least noisy schedule of audits would have involved, for each audit round, examining all schools on the same day. Although we worked with local universities in Cesar and Nariño to recruit more than 40 student auditors, this was too few to simultaneously audit up to 209 schools, many of them hours away from the department capital city. However, we show below that estimates of treatment impacts are robust to controlling for week fixed effects.[14]

Students were recruited from nutrition and related departments in local universities to undertake the audits in the context of ongoing programs (e.g. capstone projects). Since operators can shirk on many margins, the audit methodology directed these students to collect information sufficient to track several dimensions of potential

[13] In Nariño, the first date of the audits was September 6th and the last was November 20th. In Cesar, the first date was October 9th and the last was December 1st.

[14] In the end, of 64 treatment schools in Nariño, about half (30) received their first audit in the fifth and sixth weeks of the semester, and the other half in weeks one through four. After that, the audit rhythm was approximately weekly through the 13th week of the semester. In Nariño, all 62 control schools received two audits and 55 received three. Most control schools received their first audit in week 10, their second audit in weeks 11 or 12, and their final audits in weeks 12 or 13.
In Cesar, 32 of 42 treatment schools received their first audit in the eighth and ninth weeks of the semester, with approximately weekly audits thereafter, through the 15th week of the semester (extended by one week because of the strike). Of 41 control schools in Cesar, 33 received two audits and 8 only one. The first audits were in weeks 13 and 14, the second and final audit in week 15.

operator shirking.[15] There are four modalities of shirking in the case of school meals: serving too few meals, serving too small portions, substituting cheap for expensive food items, and disregarding norms regarding food storage and preparation.[16] We learned from field visits of schools in the two departments, prior to conducting the pilot study, that shirking appeared to take the form of quantity-shaving in both departments and food substitution in Nariño (the department of Cesar placed no formal restrictions on menu substitutions).[17] However, because operators were aware of these field visits, these conclusions could only be tentative.

The light audits contemplated for this study needed to be as unintrusive as possible both to avoid school interruptions and to reduce operator adjustments to the presence of auditors. The measurement of food quantities therefore posed special challenges: the intrusive use of scales to measure portions after they were already served to students was clearly impractical. Instead, we developed a training module for enumerators that allowed them to visually inspect serving sizes of four food types, tubers and derivatives, dairy products, fruits and vegetables, and protein foods, to estimate quantity-shirking.

To measure substitution, auditors collected information regarding substitutions within each of the four food types. Coding and instructions to auditors accounted for the fact that, in Nariño, minimum meal standards differ by age group (children four to six years old; seven to 12 years old; and 13 to 17 years old). In Cesar, authorities did not make these distinctions.

A contrast between these informal audits with audits in prior research highlights the measurement challenges of introducing transparency into corruption in services. Olken (2007) looks at road construction in Indonesia and measured "quantity-shirking", the use of too little asphalt. The measurement of quantity-shirking in construction differs from that in public service provision in two crucial ways, however. First, engineering measurements allow for the precise quantification of shirking in road construction. In

[15] See Winters, Testa and Fredrickson (2012) for a discussion of how the modalities of corruption affect the methodologies available to study interventions to reduce it.

[16] Considerable prior research has documented the importance of bid-rigging. Zamboni and Litschig (2018) conclude that audits in Brazil reduce corruption in local government procurement. Colonnelli and Prem (2017) find that these same audits increase entrepreneurship, improve access to finance and raise levels of economic activity (by increasing operational efficiency in corrupt firms). Mironov and Zhuravskaya (2016) find that government procurement follows political budget cycles and that in more corrupt localities, procurement contracts go to unproductive firms.

[17] Another form of shirking is to provide fewer meals than contracted. We assumed that this form of shirking would pose greater risks to the operator and would not be common. We also discovered, however, that it was not possible to measure shirking on this dimension. Both the numerator and denominator of the key indicator – the fraction of eligible children who received their meals – were difficult to document. With respect to the numerator, the student auditors could not distinguish eligible and ineligible children, nor account for eligible children not served because they were absent from school or chose to ate elsewhere. With respect to the denominator, schools did not maintain lists of eligible children.

contrast, measuring meal portion sizes is difficult and vulnerable to error.[18] Second, shirking in construction can be measured once the road is completed, without affecting the behavior of contractors. In contrast, public services, such as school meals, must be evaluated as they are provided; it is less likely that measurement will escape the attention of operators and not affect their behavior.

The informal audits were unannounced and lasted approximately 45 minutes. Students collected information about whether the type of food served followed the planned menu and whether the portion size corresponded to menu requirements established by the departmental government. They also checked whether the menu was publicly visible, as required by law; whether the quantity of food in the school storeroom was sufficient to serve the required type and quantity of food to the required number of students prior to the next delivery; that the stored food was expired or would expire prior to the date on which it was planned to be served; and that there was no visible contamination of the food (rot, mold, etc.).[19] All data was collected independently by the students using tools that they brought with them (e.g., they thermometers to check food in the storeroom).

Audit findings were summarized in an index to measure menu compliance by operators. The index aggregates information about insufficient portion sizes, in both departments. While we collected information on substitutions in Cesar, contractual arrangements with operators in that department did not prohibit them from making unilateral substitutions from agreed menus. Hence, substitution is not regarded as "shirking" in Cesar; we exclude substitution in shirking from our evaluation of treatment impacts and include unauthorized menu substitutions only in the index calculations for Nariño.

The index was constructed in the following way. If a required food group was entirely absent (in either Nariño or Cesar) or—in Nariño only—was substituted with another food item without authorization (e.g., beans for chicken, or rice for beans), the index was set to zero for that food group. If the required food group was present or the substitution was authorized, and the portion size met or exceeded the requirements set out by the Ministry of Education, the index was set to one. However, in both departments,

[18] The most precise way to measure portion sizes, using scales to weigh the portions each child received, is obtrusive and impractical. It would have also quickly prompted servers to adjust portions. Instead, we trained the auditors to estimate visually the quantities of key food items served to the children. Knowing that this procedure would be subject to error, we determined that the student auditors should visit schools in the control arm two or three times to reduce the noise in our estimates of operator compliance in control schools.

[19] To address the possibility that shirking took the form of serving spoiled or expired food items or lax hygiene practices, auditors attempted to collect information on contamination (e.g., unclean surfaces); past-expiration food items; and compliance with storage temperature guidelines. However, auditors were able to collect information on all three in only one-third of the audit visits. There was practically no variation and nearly complete compliance on two of these dimensions.

if the required food group was present, but portion sizes were smaller than allowed, the index was set to the fraction of required portion sizes that the operator supplied.

Hence, for any specific food item (in Nariño, a specific food item for a specific age group), the value of the index was set to zero for unauthorized substitutions; it was equal to the fraction (portion sizes served)/(minimum required portion sizes) when there was no substitution but portion sizes were too small; and it was equal to one when the food item was delivered in full compliance with contractual requirements. In Nariño, to construct the index for a specific food item we start with the observed serving size for an age group and divide it by the required serving size; we then average these ratios across the three age groups for each food item. The final index averages the component indices across the food groups (or food group x age group). In Cesar, the index makes no distinctions regarding age groups nor adjustments for substitutions.

The series of weekly text messages sent to parents provided them with information about the PAE and about the menu items their children should expect, lowering the costs to parents of detecting menu non-compliance (see the Appendix for further details). We designed the text messages in a separate, earlier experiment, conducted in eight schools outside of the sample analyzed here. Approximately 2,000 parents were randomized across two treatment groups (and re-randomized) in several successive weeks so that we could make an evidence-based decision regarding key message design elements. Based on family responses to the various treatments, we personalized the messages, asked for closed rather than open-ended responses, included information on three meal items rather than only one, sent them in the morning rather than the afternoon, and focused on benefits for the family's children rather than all children in the department. For details, see de la Cadena et al. (2020). Messages were sent to parents using the Colombian government's texting platform, Urna de Cristal.

Table 2 presents the number of parents receiving text messages in both departments.[20] Across 64 treated schools in Nariño, 3,973 parents received text messages, approximately 60 per school. In Cesar, where schools are much larger, 9,553 parents across 42 treated schools translated into approximately 280 parents per school. Approximately 10 percent of treated parents opted out of receiving text messages in Nariño and approximately five percent in Cesar.

[20] Differences between the beginning and end of semester numbers are due to those parents who chose to opt out of receiving text messages at the beginning of the semester.

Table 2: Parents Receiving Text Messages

	Nariño		**Cesar**	
	Treatment	*Control*	*Treatment*	*Control*
First message	3,973	3,521	9,553	6,870
Last message	3,589	3,372	9,092	6,641
% opted out	9.67	4.23	4.83	3.33

Collective Action, Corruption, and Transparency Reforms: Qualitative Evidence from the Field Visits

We leverage the pilot study to understand how service providers react to transparency interventions in a setting that differs in substantively important ways from those examined in prior research: the actors targeted with the transparency interventions are few, large and powerful; their shirking imposes small costs on many beneficiaries; and shirking is more costly to observe. Two implications of these differences concern the capacity of service providers and service beneficiaries to act collectively. In contrast to prior research, the targeted actors do not confront significant obstacles to collective action and can influence oversight policies. Beneficiaries, in contrast to prior research, confront significantly larger collective action problems, since they are both more numerous and suffer fewer losses in every corrupt transaction. Field visits in preparation for the pilot study yielded evidence on both differences.

To refine the study and tailor the interventions to the situation on the ground, the research team visited multiple schools in Cesar and Nariño, as well as in the department of Huila—also known for difficult PAE management issues—and the capital city, Bogotá. Information gathered from these visits was consistent with the collective action challenges of curbing corrupt behavior in the provision of public services.

In view of the collective action dilemma confronting the many victims of small acts of corruption, reports by PAE beneficiaries of shirking by PAE operators should be rare. The PAE legislation anticipates this dilemma and, recognizing that parents are unlikely to spontaneously demand accountability of meal providers, requires that every school that participates in the program establish a school meals committee. It also requires menus to be posted so parents know what meals to expect. Despite these efforts, in every school visited, the committee was largely moribund—parents were unaware of it, it rarely met, and/or it had never been constituted. The team also interviewed department-level officials who oversee the administration of the PAE. Despite widespread agreement that the PAE is a troubled program, they reported only sporadic parental complaints about food and few or no complaints from the committees themselves.

In contrast to beneficiaries of the PAE, the operators who supply meals to the program should not be deterred by collective action dilemmas. The few operators are politically connected and better positioned to influence the rules governing the collection of information about their behavior. Interviews during pre-pilot field work indicated that rules indeed insulated operators from consequences of non-compliance with their contractual obligations to the school meals program.

Contractor shirking could be curbed if operators were frequently audited for menu compliance. Departments are responsible for audits, but consistent with the influence of operators they are rare. Three factors limit the impact of formal audits on operator behavior, despite the PAE authorizing legislation that mandates them. First, the legislation specifies that audits should cover all aspects of food procurement, storage, and preparation, but not menu compliance.[21] Second, only audits with the characteristics specified in the legislation have legal consequences for operators. Third, central government financing of departmental PAE programs is not conditioned on the performance of these expensive audits. Interviews with officials in all departments revealed that they preferred to divert funding of expensive audits to the direct provision of meals. All confirmed that formal audits of operators are rare.

The contracts between operators and departments also protect operators from oversight. They specify high standards that evidence of contractual non-compliance must meet before departments can hold operators in breach of contract or debar them from future contracts. Parental complaints are insufficient to support charges of breach of contract; so are informal audits such as those in the pilot study. Instead, contractual breaches must be documented directly by under-staffed PAE offices in the departments. Officials interviewed in preparation for the intervention indicated that they can only act against an operator if they observe direct and flagrant evidence of malfeasance on three separate occasions. All indicated that they lacked manpower to review parental complaints in a sufficiently thorough and timely fashion that would allow them to reach the evidentiary threshold.

---

[21] The audit standard that departments are advised to use by the central government involves 15 separate sets of documents/formats, focusing on nutritional and technical details of food preparation and storage (https://www.mineducacion.gov.co/1759/articles-369472_Documento_monitoreo_2018.pdf).
https://www.mineducacion.gov.co/1759/articles-369472_recurso_10.pdf
https://www.mineducacion.gov.co/1759/articles-358483_documentos_04.pdf
Financial and administrative compliance with PAE norms is verified through the submission of documents by operators to the departments. For example, to verify how many meals the operator has prepared, the operator submits forms that the school directors sign indicating those meals – though we verified in the field that directors have no capacity to count the number of meals provided to verify the information on the forms that they sign. The field audits assume, however, that the administrative information is accurate and therefore do not focus on the quantity of food served to students.

The difficulty of collecting information on menu non-compliance can account for the fact that national officials focus disproportionately on corruption at the bidding stage. The infractions that the office of the *Contralaría* pursues are largely based on evidence that a procurement was conducted improperly, not on evidence of under-provision of meals at the school level.

Operator Responses to Audits: Substantive and Methodological Lessons from the Pilot Study

The earlier arguments predict that even operators who are few, large, and powerful should nevertheless be concerned about systematic evidence of shirking reaching national oversight agencies. However, the earlier discussion also observes that their size brings sophisticated organization, raising two methodological questions: how fast do they react to evidence of an audit, and do they react everywhere or only where the audit occurs?

Operators were in fact quick to learn about audits in their schools and could potentially have changed their behavior even during the audits. Moreover, upon learning of the audits anywhere, they could have changed behavior everywhere they were active, both in treated schools that had not yet been treated and in control schools. The pilot study yields evidence of substantial spillover effects and rapid operator responses. However, spillover need not prevent any evidence from emerging of a potential treatment effect since it entails economic losses by operators. To avoid losing rents from menu non-compliance in a school, even if they adjust their behavior in untreated schools, they may do so only partially until they observe audits in those schools.[22] If the operator profit-risk calculus leads them to continue some level of menu non-compliance in untreated (not yet treated and control) schools, it is possible to observe treatment effects.

We assess these spillover effects in several ways, reporting the results below. First, we compare the menu compliance index from the first audits in control schools, conducted at the end of the semester, with those in the first audits in treated schools, conducted at the beginning of the semester. In the absence of spillover effects, the first audits in treatment and control schools should reveal similar levels of menu compliance. If spillover occurred, menu compliance in the first audits of control schools should be

[22] In Reinikka and Svensson (2004), the key finding is that an information campaign increased the receipt of capitation grants in surveyed schools closer to newspaper outlets than in surveyed schools further away, although both sets of schools reported to the same district officials. This implies that district officials did not respond to surveys by increasing capitation grants in all schools, either because they also realized that those further away from newspaper outlets could not cause problems for them, or because only a fraction of their schools was surveyed (about 10 percent of district schools, 250 schools from 18 districts, each with between 59 and 399 total schools).

greater. In fact, menu compliance measured during the first data collection audit of control schools was significantly greater than menu compliance during the first audit of treatments schools.

Second, we assessed the speed of operator reaction by comparing shirking over the first few audits in treated schools (which occurred at the beginning of the semester) and control schools (undertaken at the end of the semester). To the extent that operators reacted rapidly, menu compliance in both treatment and control schools should rapidly improve after the first audit in each, even though first audits in control schools occurred many weeks after first audits in treated schools. In both treatment and control schools, we find significant improvements in menu compliance from the first to the second and third audits.

Third, we ask whether there is any formal evidence of a treatment effect, comparing end-line data in treated and control schools. Substantial spillovers from treated to control schools and the rapid reaction of operators to the first audits in control schools bias against finding any observed treatment effect: menu compliance in the first control audit is biased upwards by spillovers from treated to control schools, and compliance in the second and third control audits is upwardly biased because of operator reaction to the first audit. Consistent with this, there is no difference in end-line *averages* across treatment and control schools. However, menu compliance reported in the *first* control school data collection audits is significantly lower than menu compliance reported in end-line audits of treated schools.

We randomized schools into control and treatment groups, stratifying by department and by whether their contract called for meals to be prepared on-site, in school facilities, or off-site, in the contractor's kitchens. Control and treatment schools were statistically indistinguishable on observables, as Table 3 illustrates. Key characteristics that might affect the costs of operator compliance or non-compliance are the number of breakfasts, lunches, and afternoon snacks in the operator contract; whether the meals are prepared at the school (on-site), or off-site for subsequent delivery to the school; whether the school had a contract for morning, lunch, or afternoon meals; and the number of students. The two arms exhibit statistical balance across all of these.

Table 3: Balance Table, Control vs. Treated Schools

	Mean treatment	Mean control	Difference	*p*-value, significance of difference
Which meal observed at first audit?				
Morning Snack	0.835	0.783	0.052	0.342
Lunch	0.155	0.179	-0.024	0.645
Afternoon Snack	0.010	0.038	-0.028	0.187
Number of AM meals in the contract	88.835	87.981	0.854	0.972
Number of lunch meals in the contract	28.728	46.179	-17.451	0.399
Number of PM meals in the contract	15.447	24.491	-9.044	0.545
Prepared off-site (vs. on-site)	0.330	0.368	-0.038	0.568
Total students (from SIMAT, national government statistics)	115.777	140.774	-24.997	0.445

Some variables exhibit larger differences, albeit still insignificant: the contracted number of lunches and afternoon snacks is notably larger in control schools, as is the total number of students. However, the results we report are insensitive to the inclusion of controls for these observables in the estimations reported below.

Table 4 summarizes values of the menu compliance index (quantity-shirking and improper substitution) at key junctures: the first and last audits in the treatment and control schools. Compliance increased significantly (one-half of a standard deviation) between the first and last audits in treatment schools; between the first and last audits in the control schools (about one-fifth of a standard deviation); and, compared to first audits of treatment schools, the first audits of control schools reveal higher compliance (by approximately one-third of a standard deviation), pointing to spillovers from treated schools to control schools.

Table 4: Menu Compliance, Treated vs. Control Schools, First and Last Audits

	Mean	Standard deviation
Control		
First audit	.53	.38
Last audit	.61	.34
Treated		
First audit	.42	.35
Last audit	.58	.33

Table 5 explores spillovers from treated to control schools in more detail, estimating compliance differences between the first audit in treated schools, early in the semester, and the first audit in control schools, late in the semester. The first specification has no controls, the second controls for operator and municipal fixed effects and for school characteristics. Consistent with the presence of spillovers, compliance in control schools is significantly higher. The coefficient -0.11 indicates that compliance measured in treatment schools during their first audit was approximately one-third of a standard deviation worse than compliance measured in the control schools in their first audit. Estimates are nearly the same in the two specifications.

Table 5: Spillovers from Treatment to Control Schools

	Menu compliance	
	(1)	(2)
=1 for first audit in treated school, =0 for first audit in control	-0.118**	-0.108***
	(0.051)	(0.037)
Controls	No	Yes
Observations	209	196
R2	0.026	0.716
Municipality FE	No	Yes
Operator FE	No	Yes

Notes: Balance controls are type of meal, mode of preparation, number of meals in the contract and the total students of the school. Changes in observations responds to the subsamples conditions and singleton observations are dropped following Correia (2015) when municipality fixed effects are included. Cluster standard errors at the school level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

It is of course impossible to conduct a formal test for spillovers, which would require comparing end of semester compliance in control schools run by operators who were treated with schools served by operators who were not treated. Indeed, the very reason that spillovers arise is because the small number of operators prevents operators from being randomly assigned to treatment and control groups. Hence, we cannot exclude the possibility that operator menu compliance increased over time for reasons unrelated to our intervention, leading to greater menu compliance in the first control audit than in the first treatment audit. However, estimates that examine the speed of operator response argue against this interpretation. As Table 6 demonstrates, improvements in control schools were observed at the end of the semester and in treatment schools at the beginning of the semester.

Qualitative evidence that accumulated throughout the implementation of the pilot pointed to rapid operator responses to the presence of auditors, reflecting the size and sophistication of operators' organizations. The student auditors reported that their arrival at schools frequently drew the immediate attention of operators; in some cases, operators' area supervisors arrived at the school in less than an hour to question auditors about their presence. In addition, operators called the departmental PAE administrators to complain about the audits. This anecdotal evidence suggests that we should expect to observe rapid changes, not gradual, in operator behavior. Table 6 provides evidence of the speed of operator reactions in both control and treatment schools.

Our concern here is within-school changes in operator behavior, not spillovers across schools, examined in Table 5. Therefore, all specifications in Table 6 control for school fixed effects. The first column examines control schools. These received two or three audits at the end of the semester. The first coefficient demonstrates a large and significant increase in compliance from the first audit to the second or second and third audits. In that short period, the menu compliance index rose .21 of a standard deviation compared to the first audits of control schools.

Table 6: Speed of Operator Reaction, Control Schools

Dependent variable: menu compliance

	Control schools, **end of semester**	Treatment vs control schools, **end of semester**	Treatment schools, **beginning of semester**
	(1)	(2)	(3)
First audit (=0) versus next one or two audits (=1) (*see note*)	0.079** (0.035)	0.079** (0.035)	
Last two audits vs third from last X treated (*see note*)		-0.089* (0.043)	
Beginning of semester: Audits 2 and 3 (= 1) vs Audit 1 (= 0)			0.100*** (0.034)
Observations	245	561	316
R^2	0.682	0.660	0.674
School FE	Yes	Yes	Yes

Notes: In the first column, the first row coefficient compares the first visit with the last two visits in control schools that had three data collection audits, and the first with the second visit in controls schools with only two data collection audits. The second column includes treated schools, the last two visits versus the third-from-last. The first row estimates the difference in control schools; the second row, with the interaction term, the effect in treated schools. The third column looks only at treated schools at the beginning of the semester and compares compliance at the first audit with the second and third audits. Cluster standard errors at the school level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The results in the first column could be the consequence of unobserved events at the end of the semester that motivated operators to comply more assiduously with their menu obligations. However, if this had been the case, we should observe a similar increase in compliance in the final two audits in treatment schools (*T* and *T-1*) compared to audit *T-2*, where *T* denotes the number of the last audit. The second regression examines this possibility by comparing changes in audit performance in treated and control schools at the end of the semester. Two variables are of interest. The first is an indicator variable that is 1 if the audit was the last and second to last (for control schools with only two audits, last only); and 0 if it was the third to last (for control schools with only two audits, second to last). The second is the interaction of this indicator variable and the treatment variable. The coefficient in row 1, column 2 tells us, as in column 1, that compliance in control schools increased .079. The coefficient in row 2, column 2, rejects the conjecture that compliance increased in all schools at the end of the semester; if anything, compliance in treated schools declined.

This evidence is consistent with the conclusion that compliance improvements in control schools resulted from the rapid reaction of operators to audits. An alternative explanation is that compliance improvements in these schools were the product of an unobserved end-of-semester shock in control schools. If this explanation were correct, however, we should not observe similarly rapid changes in operator behavior in treatment schools at the *beginning* of the semester. The results presented in the final column of Table 6 demonstrate that this is not the case: treated schools exhibited rapid improvements in menu compliance from the first to the second and third audits.

This specification replicates that of column 1 in Table 6, but now compares compliance during the second and third audits in treatment schools with compliance during the first audit, all at the beginning of the semester. The coefficient is large, significant, and comparable both to the treatment effect observed in the main results earlier and to the change in compliance observed in control schools at the end of the semester. By the third audit, operators had increased compliance by .28 of a standard deviation relative to compliance measured in the first audit.

The results summarized in Tables 5 and 6 show substantial spillovers and rapid operator responses. The remaining question is whether, despite these effects, it is still possible to detect treatment effects by comparing treated and control schools at end-line. Are end-line measures of the menu compliance index significantly higher in treatment than in control schools?

One way to make such a comparison is to average the end-line audits in the control and treated schools. The average of multiple audits has the advantage of greater accuracy, given the noise inherent in compliance measurement. However, the average of the menu compliance index across all (two or three) control audits is subject to *both* sources of upward bias of the compliance measure: spillovers from treatment to control schools and within-school behavioral changes after the first control audit. Given this, it is unsurprising that there is no evidence of a treatment effect if we compare the average of all control audits with the average of the last three audits in the treatment schools.

The *first* audit of control schools reflects only one of these biases, the possible spillover from treatment to control schools, but is immune to the other, the within-school behavioral response of operators to audits. In fact, tentative evidence of a treatment effect emerges from comparing end of semester values of the menu compliance index in treated schools with the index calculated based on the first audit of control schools.

Most of the improvement in treatment school compliance occurred earlier in the semester since operators reacted rapidly. In principle, therefore, any end-of-semester audits in treatment schools fully reflect the treatment and could be used in the comparison with the first control audit. However, comparisons could be sensitive to the timing of the first control audits and the end-of-semester treatment audits. These were not possible to match up cleanly—first control audits took place in various weeks and coincided with different treatment audits—sometimes the penultimate, sometimes the third from last, etcetera. Hence, we simply compare the first control audit to various combinations of the last three treatment audits:

$$Treated_i = 1\ for\ i \in \{Treated\}\ and\ audit\ numbers \in \{T, T-1, T-2\}$$
$$Treated_i = 0\ for\ i \in \{Control\}\ and\ audit\ numbers \in \{1\}$$

where the unit of observation is the school, indexed by *i*, and *T* denotes the final audit that a school receives. We examine multiple combinations of the end-of-semester observations of treated schools, including the average of all three end-of-semester observations. The specification we test is:

$$Menu\ Compliance\ Index_i = \beta_0 + \beta_1 Treated_i + \boldsymbol{X_i \beta_2} + \upsilon_h + \omega_j + \epsilon_i.$$

It was not possible—again, because of the low number of operators—to balance treatment and control schools at the level of operator and municipality.[23] However, unobserved operator and municipal characteristics are likely to influence average treatment effects. Unobserved characteristics of operator organizations dictate how rapidly and widely they react to evidence of audits. Those reactions can vary by municipality depending on unobserved relationships between operators and mayors. Mayors differ in their sensitivity to the electoral consequences of shirking in the PAE, and in their ability and willingness to make appeals to higher level government officials to

[23] It is possible, for example, that treatment schools might disproportionately be located in municipalities where mayors are less (or more) attentive to school meals and therefore respond less (or more) to the treatment.

intervene. Hence our main specifications control for fixed operator and municipality effects, υ_h and ω_j, respectively.

Controlling for both types of fixed effects, identification is based only on comparisons between control and treated schools run by the same operator in the same municipality. This excludes treatment or control schools in municipalities where their operator has no corresponding comparison school (control or treatment school, respectively). We show, however, that the differences between control and treatment schools with both municipal and operator fixed effects are similar in magnitude and significance to differences estimated with only operator or only municipal fixed effects.

Finally, we control for school-specific variables $\boldsymbol{X_i}$, all of which are statistically balanced across treatment and control schools, but which nevertheless are not identical. These variables are dummies for the type of meal served (lunch or snack), whether it was prepared on site or delivered ready-to-eat; and the number of students in the school, in total and by age group.

Table 7.1 reports the evidence of end-line differences between treatment and control schools. The estimates compare the last three audits in treatment schools with the first audit in control schools. Each treatment audit is a separate observation, but the results are the same if we instead use the average values for the treated schools (see Table A.1 in the Annex).

Table 7.1: Treatment Effects on Menu Compliance *(unauthorized substitutions and quantity-shaving)*

	Dependent variable: menu compliance index						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment = 1 for any of last three treatment audits, = 0 for first control audit	0.067 (0.046)	0.091*** (0.034)	0.073** (0.034)	0.079** (0.033)	0.060* (0.035)	0.081** (0.033)	0.076** (0.037)
Balance Controls	No	No	No	Yes	Yes	Yes	Yes
Observations	419	411	419	410	419	410	408
R^2	0.007	0.503	0.338	0.523	0.355	0.524	0.539
Municipality FE	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**Yes**
Operator FE	**No**	**No**	**Yes**	**No**	**Yes**	**Yes**	**Yes**
Week FE	**No**	**No**	**No**	**No**	**No**	**No**	**Yes**

Notes: Balance controls are type of meal, mode of preparation, number of meals in the contract and the total students enrolled in the school. Changes in observations responds to the subsamples conditions and singleton observations are dropped following Correia (2015) when municipality fixed effects are included. Cluster standard errors at the school level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The first three columns highlight the relevance of unobserved operator and municipal effects. The first column specification controls for neither, an inappropriate specification since unobserved municipal and operator characteristics are likely to influence the estimates. Estimates are nevertheless large (.19 of a standard deviation), but not significant. The regression in column 2 adds controls for municipal fixed effects, column 3 adds operator fixed effects to the specification in column 1. In both cases results are highly significant. Columns 4 and 5 add balance controls to the specifications in columns 2 and 3. These have no impact on the estimates, as expected given the high level of balance on observables between control and treatment schools. Column 6 controls for both operator and municipal fixed effects and balance variables. Given the theoretical importance that both operator and municipal characteristics could play in reactions to the audits, this is our preferred specification. Results remain economically and statistically significant and largely the same as the previous estimates.

Finally, estimates in column 7 add a control for week fixed effects, to address the possibility that unobserved circumstances that arose in the specific weeks that treated and control schools received their end-line audits could account for observed differences in menu compliance. It was not possible to ensure that each audit in the sequence was given to all target schools in the same week (i.e., that in week *n* of the study all schools received their *n*th audit). Appendix Table A.2 demonstrates the dispersion of audits across weeks, indicative of the logistical challenges of conducting the study: five times as many enumerators would have been needed to ensure simultaneous audits. The estimates in column 7 demonstrate that this dispersion has little effect, results are essentially unchanged from column 6. We exclude week fixed effects from the main specification, however, since their inclusion requires that we rely for identification only on those treated and control schools located in the same municipality, served by the same operator, and occurring in the same week.

Table 7.2 demonstrates that the differences reported in Table 7.1 are not an artefact of the specific treatment audits chosen for the comparison. The specifications in Table 7.2 follow the specification of column 6, Table 7.1, controlling for municipal and operator fixed effects and including balance controls. Whether one compares the control audit to the last two audits in treated schools, to the second-to-last audit or to the third-to-last audit, significant differences emerge. Note that sample sizes drop according to the number of treatment audits that are included in the comparison.

Table 7.2: Treatment Effects on Menu Compliance, Alternative End-line audits in Treatment Schools

	Last 2 audits in treated schools	2nd-to-last audit in treated schools	3rd-to-last audit in treated schools	Last audit in treated schools	Last 3 audits in treated schools, the same menu compliance index for both departments	
					Index accounts for menu substitution	Index ignores menu substitution
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment = 1 for audits indicated in column headings, = 0 for first control audit	0.072**	0.084*	0.093**	0.053	0.071*	0.079*
	(0.036)	(0.045)	(0.041)	(0.046)	(0.038)	(0.043)
Observations	306	195	193	195	410	398
R^2	0.518	0.612	0.668	0.559	0.373	0.339
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Operator FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Balance controls are included in all specifications: the type of meal, mode of preparation, number of meals in the contract and the total students of the school. Changes in observations responds to the subsamples conditions and singleton observations are dropped following Correia (2015) when municipality fixed effects are included. Cluster standard errors at the school level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The last audit in treatment schools, conducted right before the Christmas holidays, is the exception and reveals no significant difference with the first end-line audit in control schools. However, observers alerted us prior to the pilot study that the last week of the semester, before Christmas and after exams, is anomalous: school attendance tends to be low and unpredictable, challenging circumstances both in which to supply school meals and to monitor menu compliance. It is, however, also possible that operators no longer responded to the audits in that week since their performance in that off-week would have little effect on the conclusions that authorities might reach based on the information accumulated in the previous seven or more audits. Of course, it is also a possible question of low power due to the relatively small sample size.

An additional robustness issue concerns the definition of corrupt behavior and the measure of menu compliance. In Table 7.1 we use different measures in Nariño and Cesar because of the different legal definitions of shirking in the two departments: in both, quantity-shaving is prohibited, but only in Nariño must menu substitutions be authorized by the departmental PAE authorities. However, parents and enforcement authorities might care more about quantity-shaving than menu substitutions, or that they object to menu substitutions even if these are formally allowed. In the first case, it would be more appropriate to exclude menu substitutions from the Nariño index when examining responses to the intervention; in the second case it would be more appropriate to include menu substitutions in the Cesar index.

The last two columns of Table 7.2 examine the robustness of the results in Table 7.1 to the use of these alternative indices. The first uses the "Nariño" definition for both departmental indices, adjusting the Cesar index to account for menu substitutions. The second uses the "Cesar" definition for both indices, adjusting the Nariño index to exclude consideration of menu substitutions. The preferred specification, from column 6 of Table 7.1 is used, controlling for municipal and operator fixed effects and including balance controls. The resulting estimates are similar in magnitude to those in Table 7.1 and modestly significant with *p*-values of approximately .06.

Mechanisms

Both treatments analyzed in the pilot study, informal audits and text messages to parents, could affect "bottom-up" mobilization by parents to improve the quality of school meals. They could have encouraged parents to confront operators directly, to lobby municipal and state officials, or by to bring operator shirking to the attention of national authorities. The informal audits could also have directly changed operator behavior, independent of parental mobilization, out of fear that systematic information on hundreds of schools would have drawn the attention of national authorities (though not departmental authorities, given their influence in departmental politics). An end-line telephone survey of 4,205 households with children enrolled in the sample schools yields experimental evidence on the first mechanism; other quantitative and qualitative information is suggestive that the second also played an important role.

Collective action by parents to improve meal quality is difficult. Individual households may be unaware both of operator obligations and the institutional channels through which they can mobilize, such as the school meals committees. Parents might further believe that even their concerted action would be insufficient to shift operator behavior directly or that lobbying of municipal and state officials would be ineffective. The text messages and informal audits could have reduced these obstacles. The text messages informed them of operator obligations—the meals that their children should receive; of the existence and purpose of the school meals committees; and of opportunities to voice their concerns. The informal audits would have signaled the interest of a well-organized outside group in improving operator performance in their schools, raising expectations about the potential efficacy of parental mobilization.

Evidence from the end-line telephone survey of parents in treated and control schools indicates that the two interventions significantly increased parental participation in the oversight of school meals. There are two established venues for such oversight, the school meals committee and *ad hoc veedurías*, meetings with parents, organized by the school, to review any aspect of school performance, including school meals. Column 1 of Table 8.1 indicates that 18.5 percent of respondents in control schools reported attending a meeting of the school meals committee. The number was more than one-third (6.5 percentage points) larger in treated schools. Column 2 examines parental attendance at oversight meetings that addressed school meals. In control schools, 20.6 percent of parents reported attending such a meeting; nearly half again as many (9.5 percentage points more) parents in treated schools reported attending such an oversight meeting.

Table 8.1: Parental Engagement with the School Meals Program, Treated versus Control Schools

	(1)	(2)	(3)	(4)	(5)
	Did you attend a meeting of the school meals committee this semester?	Did you attend an oversight meeting (veeduría) this semester where school meals were discussed?	How often do you speak with other parents about the meals that your children receive in school?	This semester, did you observe (supervise) the delivery of food supplies to the school or the serving of meals?	Has the school meals committee in the school your child attends met this semester?
Treated	0.065***	0.095***	-0.031	-0.020	0.028
	(0.024)	(0.024)	(0.023)	(0.027)	(0.029)
Constant	0.185***	0.206***	0.662***	0.268***	0.269***
	(0.015)	(0.013)	(0.017)	(0.019)	(0.020)
Obs	4,205	4,205	4,205	4,205	4,205
R-squared	0.006	0.012	0.001	0.001	0.001

Notes: Ordinary least squares, xluster standard errors at the school level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; “Constant” is also the control mean, as regressions did not include controls.

A large fraction of parents in control schools, 66.2 percent, report speaking with other parents about school meals, essentially the same as in treated schools (Column 3). This high fraction demonstrates the salience of school meals to parents. The two treatments therefore seem to close the gap between the number of parents who are concerned about school meals, those who talk about meals with other parents, and the number who take some sort of action to improve them, such as attending oversight meetings.

Parents were informed in the text messages that they could also observe the delivery of food supplies or the serving of meals, but this was not enough to prompt them to engage in either activity. In the control schools, 26.8 percent of parents either observed the delivery of supplies or meal service (Column 4). This fraction is essentially the same in treated schools and is a likely consequence of the different costs and benefits to parents of the different oversight activities. With respect to costs, to observe food deliveries or meal service, parents must be at the school early in the day, well before students arrive, when they are normally engaged in household tasks, or already at work, and in any case not routinely at the school. The meetings take place after school, when parents are more likely to be at the school anyway (to pick up their children) or are not working. With respect to benefits, the institutional process for bringing parental observations regarding food delivery and meal service to the attention of operators is less clear to parents than the process for bringing the conclusions of the oversight meetings.

Column 5 verifies that the opportunity for parents to attend school meetings was the same in control and treatment groups; there is no difference between the two groups in the fraction of respondents who say the committee met. The fraction, however, is low: fewer than 30 percent of parents agree that the committee met. The treatments increased participation in school meetings but did not increase the number of meetings held. The results in Table 8.2 demonstrate that the effects on meeting attendance documented in columns 1 and 2 of Table 8.2 are highly robust to a large number of controls and municipal fixed effects.

Table 8.2: Parental Attendance at Oversight Meetings, Treated versus Control Schools, Robustness

	(1)	(2)	(3)	(4)
	Did you attend a meeting of the school meals committee this semester?		Did you attend an oversight meeting (veeduría) this semester where school meals were discussed?	
Treated	0.081***	0.077***	0.112***	0.107***
	(0.024)	(0.022)	(0.020)	(0.019)
Controls	No	Yes	No	Yes
Municipality FE	Yes	Yes	Yes	Yes
Observations	4,204	4,204	4,204	4,204
R-squared	0.046	0.064	0.047	0.067

Notes: Ordinary least squares. Controls included are: economic stratum; respondent relationship to the household children; gender of respondent; if respondent is the mother or knows how to read and write; marital status, education level, hours worked if employed, earnings at work; participation in community/volunteer work in the previous three months; if respondent has a phone or smartphone; if respondent knows how to use a computer; the frequency of text messages their school received; if the respondent's residence has electricity or water and sewerage connections; if the family belongs to "Familias en Acción"; the mode of transportation and travel time to school; and if the respondent feels part of the community, neighborhood, and school's parent community. Clustered standard errors at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

It is possible that the significant treatment effects reported in columns 1 and 2 of Table 8.1 are the product of social desirability bias: parents who received the text messages are more likely to believe that the "correct" response is to attend oversight meetings. There are two reasons to believe this is not the case, however. First, the text messages specifically referred to the meetings of the school meals committee, but not to the *veedurías*. Parents in treated schools participated even more in *veedurías* than in meetings of the school meals committee, however. Second, the presence of such a bias should also yield significant differences between treated and control parents in their responses concerning other types of participation. No such differences are observed.

Tables 8.1 and 8.2 provide evidence of a "bottom-up" effect of the interventions. It is likely, however, that operators also responded to a "top-down" effect, the fear of intervention by national authorities. One indication of this is that, although there were no text messages in control schools to prompt parental participation in oversight activities, these also saw an improvement in menu compliance after a single audit towards the end of the semester (column 1 in Table 5). The increase in menu compliance after the first control audit that was comparable to the effect observed over the first audits in treated schools, just under .25 standard deviations.

Other evidence from online media reports and the websites of the main national government oversight agencies also supports the salience of the "top-down" mechanism. The Fiscalía (Attorney General) and the Contraloría (Auditor General) place the results of all their investigations on their respective websites. In August 2018, the Fiscalía announced that it had been investigating the PAE program in 14 departments.[24] Media reports of PAE corruption were also common prior to the treatment period.[25] Hence, operator concerns regarding prosecutorial interest in the audits were well-founded.

Moreover, there was a significant difference in prosecutorial activity, pre-pilot, across the two departments. Nariño was not mentioned in media reports nor targeted by prosecutors. Cesar, in contrast, was prominent in both respects, precisely because of

---

[24] The 14 departments under investigation were Atlántico, Cesar, Huila, Norte de Santander, Santander, La Guajira, Chocó, Córdoba, Bolívar, Valle del Cauca, Cauca Amazonas, Sucre, Magdalena. https://www.fiscalia.gov.co/colombia/seccionales/contundente-golpe-a-entramado-de-corrupcion-por-plan-de-alimentacion-escolar-pae/ (Accessed 1/24/2019)

[25] Reports emerged as early as September 2011, (http://www.noticiascandela.informe25.com/2011/09/corrupcion-en-el-pae.html), with frequent reports in 2016, the year before the intervention. For example, from April of that year, https://www.semana.com/educacion/articulo/corrupcion-en-el-programa-de-alimentacion-escolar-del-ministerio-de-educacion/468169 https://www.elespectador.com/noticias/educacion/mas-y-mas-medidas-combatir-corrupcion-programa-de-alime-articulo-626586. And reports emerged as well in September 2017, just before the intervention. https://www.vanguardia.com/opinion/editorial/el-pae-y-los-carteles-de-la-corrupcion-JFVL409099 (All accessed 1/24/2019).

the Aguachica scandal discussed earlier.[26] The Contraloría had been examining PAE contracts in Cesar continuously from August 2016 to September 2017; our intervention in Cesar began in November 2017. Towards the end of the treatment period, the Contraloría reported the results of its August 2016–September 2017 investigations in Cesar, naming specific individuals and contractors whom it deemed responsible for improper financial management of the PAE program.[27]

The disproportionate attention paid to Cesar by media and the Contraloría has two empirical implications. First, if top-down oversight has a significant effect on operator behavior, operators in Cesar should have exhibited significantly greater menu compliance at baseline compared to operators in Nariño: neither prosecutors nor the media gave operators in Nariño reason to change their behavior prior to the pilot study. Second, the effects of the treatment on operator behavior should have been greater in Nariño than in Cesar.

Table 9 documents the first implication, reporting evidence that menu compliance at baseline was significantly greater in Cesar. The first panel reports the values of the menu compliance index used in Table 4 for the two departments separately. For both departments, the index accounts for quantity-shaving. For Nariño, the index accounts for unauthorized menu substitutions. In Cesar, substitutions are always authorized. The first audit in Cesar, whether in the treated or control schools, yields a menu compliance index that is several times larger than in Nariño—for example, .75 versus .21 in the first audits of treated schools.

These differences may be an artefact of the rules used to construct the index, which exclude menu substitutions from entering the calculation of the Cesar index since operators were permitted to make such substitutions. The remaining two panels in Table 9, however, demonstrate that the differences remain when the departmental indices are calculated identically, as in the last two columns of Table 7.2. The second panel of Table 9 compares the Nariño index with a recalculated Cesar index that accounts for menu substitutions. The Cesar index falls from the first panel to the second (e.g., from .75 to .61 for first audits of treated schools), since menu substitutions provide an additional way for operator shirking in Cesar to be revealed. The Cesar index remains several times larger than in Nariño (e.g., .61 versus .21 in first audits of treated schools). The last panel in Table 9 recalculates the Nariño index to exclude from consideration menu

[26] In April 2016, the media reported that Colombia's Attorney General (the Fiscalía) had intervened against the mayor of Aguachica (Cesar) and its PAE contractors (accessed 1/24/2019). https://www.fiscalia.gov.co/colombia/noticias/destacada/corrupcion-en-contratos-para-alimentar-ninos-capturan-a-alcalde-de-aguachica-cesar-contratistas-de-esa-ciudad-y-de-maicao-guajira/

[27] https://elpilon.com.co/irregularidades-pae-abrio-13-procesos-8-518-millones-cesar/. Accessed 1/24/2019.

substitutions and compares that to the original Cesar index. The Nariño index rises compared to the previous panels, since one source of shirking is no longer accounted for. Nevertheless, the differences with Cesar remain large (e.g., .75 in Cesar versus .39 in Nariño in first audits of treated schools).

Table 9: Menu Compliance Index, Cesar and Nariño, Various Measures

Index used in Table 4: Menu substitutions enter Nariño index, not Cesar index; quantity-shaving enters both

	Cesar		**Nariño**	
	Mean	Std. dev.	Mean	Std. dev.
Control				
First audit	**0.90**	0.14	**0.34**	0.32
Last audit	0.82	0.29	0.48	0.31
Treated				
First audit	**0.75**	0.23	**0.21**	0.23
Last audit	0.79	0.26	0.45	0.31

Cesar index calculated as in Nariño: menu substitutions and quantity-shaving enter both indices

	Cesar		**Nariño**	
	Mean	Std. dev.	Mean	Std. dev.
Control				
First audit	**0.80**	0.14	**0.34**	0.32
Last Audit	0.63	0.29	0.48	0.31
Treated				
First audit	**0.61**	0.23	**0.21**	0.23
Last audit	0.65	0.26	0.45	0.31

Nariño index calculated as in Cesar: quantity-shaving enters both indices, menu substitutions enter neither

	Cesar		**Nariño**	
	Mean	Std. dev.	Mean	Std. dev.
Control				
First audit	**0.90**	0.14	**0.51**	0.32
Last audit	0.82	0.29	0.74	0.31
Treated				
First audit	**0.75**	0.23	**0.39**	0.23
Last audit	0.79	0.26	0.65	0.31

The evidence in Table 9 is consistent with the proposition that operators are highly concerned about national enforcement authorities, particularly since high observed menu compliance in Cesar is contrary to all observers' subjective assessments of operator performance in the department and with the behavior of national enforcement authorities, who directed substantial investigative resources to the department. Table 9 is suggestive, moreover, that the pilot intervention should have had little effect in Cesar since operators had already responded to pre-pilot interventions by prosecutorial authorities. Table 10 confirms this intuition. It demonstrates that menu compliance rose significantly in treated relative to control schools in Nariño, but not at all in Cesar.

Table 10: Treatment effects, Nariño and Cesar

	Nariño		Cesar	
	(1)	(2)	(3)	(4)
Last 3 treated visits vs 1st control	0.130***	0.190***	-0.023	-0.042
	(0.049)	(0.050)	(0.044)	(0.034)
Balance Controls	No	Yes	No	Yes
Observations	254	246	165	165
R2	0.030	0.326	0.002	0.360
Municipality FE	No	Yes	No	Yes
Operator FE	No	Yes	No	Yes

Notes: Balance controls are type of meal, mode of preparation, number of meals in the contract and the total students of the school. Changes in observations responds to the subsamples conditions and singleton observations are dropped following Correia (2015) when municipality fixed effects are included. Cluster standard errors at the school level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11 offers additional insight about mechanisms, reinforcing the importance of the "top-down" element in controlling corruption. If parental engagement were the critical factor in motivating a change in operator behavior, we would expect correspondingly weaker effects of the interventions on parental meeting attendance in Cesar compared to Nariño. In Cesar, where rates of compliance were already high, parental attendance at meetings should already have been high in both treatment and control schools, offering little scope for the intervention to increase it. In Nariño, in contrast, rates of non-compliance were high at baseline and fell significantly over the course of the treatment. If parental engagement were the sole reason for this increase, treatment effects on parental engagement should have been larger in Nariño.

Table 11: Treatment Effects on Parental Participation, Nariño and Cesar

	(1)	(2)	(3)	(4)
	Did you attend a meeting of the school meals committee this semester?		Did you attend an oversight meeting (veeduría) this semester where school meals were discussed?	
	Cesar	Nariño	Cesar	Nariño
Treated	0.100***	0.031	0.104***	0.132***
	(0.026)	(0.049)	(0.023)	(0.045)
Constant	0.140***	0.249***	0.194***	0.198***
	(0.018)	(0.031)	(0.014)	(0.028)
Controls	No	Yes	No	Yes
Municipality FE	Yes	Yes	Yes	Yes
Observations	2,657	1,547	2,657	1,547
R-squared	0.039	0.042	0.041	0.058

Notes: Ordinary least squares, Cesar and Nariño subsamples. Controls included are: economic stratum; respondent relationship to the household children; gender of respondent; if respondent is the mother or knows how to read and write; marital status, education level, hours worked if employed, earnings at work; participation in community/volunteer work in the previous three months; if respondent has a phone or smartphone; if respondent knows how to use a computer; the frequency of text messages their school received; if the respondent’s residence has electricity or water and sewerage connections; if the family belongs to "Familias en Acción"; the mode of transportation and travel time to school; and if the respondent feels part of the community, neighborhood, and school's parent community. Clustered standard errors at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 11 results indicate that this is not the case. The fraction of parents in treated schools in Cesar who attended both types of meetings was 10 percentage points higher than in control schools. In Nariño, treatment effects on attendance at school meals committee meetings were insignificant. The fraction of parents who attended *veedurías* was 13 percentage points higher in treated schools. However, even in control schools, the constant coefficients in Table 11 indicate that the fraction who attended meetings was as high or higher in Nariño, where baseline menu compliance was low, than in Cesar, where it was high.

Conclusions

Corruption by a few, influential actors is a notorious problem but rigorous empirical analysis of interventions to curb it are rare. The most extensively researched interventions are related to transparency and the literature offers ample evidence of their efficacy, particularly audits. However, this evidence emerges in the context of corruption undertaken by many actors with limited ability to manipulate or prevent oversight of their behavior, which is in turn relatively easy to measure.

This paper makes three contributions to the study of transparency and corruption. First, it highlights the obstacles that in theory should limit the efficacy of transparency interventions in the context of significant corruption by a few. Second, it identifies and documents empirically the methodological obstacles to studying the impact of transparency interventions in the context of significant corruption by a few. Third, it produces suggestive evidence that, despite these obstacles, transparency interventions—informal audits and communication with beneficiaries—can change behavior. The evidence also sheds light on mechanisms and is consistent with both bottom-up pressures from parents and operator concerns about top-down enforcement actions by national oversight agencies.

Efforts to combat grand corruption almost always focus on institutional, judicial, and fiduciary remedies. This is also true in Colombia's oversight of the School Meals Program. Problems with the program have prompted successive governments to shift responsibility for the program to different ministries and, for the contracting of operators, from the national level to the departmental level. Fiduciary and prosecutorial attention has been a constant feature of the program.

Such strategies bear fruit occasionally—the Odebrecht scandal shook Latin America—although we have little understanding of the conditions under which prosecutors act and when their actions are enough to bring down powerful actors with important political connections. At the same time, corrupt actors are aware of the threats posed by fiduciary oversight and negotiate contracts that raise the costs of oversight beyond what governments are generally willing to pay. The results of the pilot study of the School Meals Program point to a complementary approach and provide important insights into the future design of evaluations of interventions to curb corruption by a few influential actors.

Substantively, the pilot results are suggestive that transparency interventions can influence behavior, particularly if they are likely to prompt action by both victims and high-level enforcement agencies. The most important methodological insights are two. First, spillovers from treatment to control beneficiaries served by the same operator are significant, but operator self-interest is sufficient to allow for treatment effects to be observed: operators are reluctant to lose money in control schools by curbing rent-seeking too much, too soon. Second, operators respond rapidly once they observe an intervention. Hence, the direct measurement of services provided to control group beneficiaries at end-line should be based on only a single observation.

One priority for future research is to replicate these results in a fully-fledged evaluation. A broader priority, however, is to disentangle more definitively the extent to which transparency interventions aimed at curbing grand corruption depend on the presence of higher-level agencies willing to investigate and take enforcement actions against the influential actors behind the corrupt activity.

References

Alderman, Harold and Donald Bundy (2012). "School Feeding Programs and Development: Are we Framing the Question Correctly?" *The World Bank Research Observer* 27(2): 204-221.

Avis, Eric, Claudio Ferraz and Frederico Finan (2016). "Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians." NBER Working Paper No. 22443 (July).

Banerjee, Abhijit, Sendhil Mullainathan and Rema Hanna (2012). "Corruption." NBER Working Paper 17968 (April).

Becker, Gary S. and George J. Stigler (1974). "Law Enforcement, Malfeasance, and Compensation of Enforcers." *Journal of Legal Studies* 3:1.

Bobonis, Gustavo J., Luis R. Cámara Fuertes, and Rainer Schwabe (2016). "Monitoring Corruptible Politicians." *American Economic Review* 106(8): 2371-2405.

Broms, Rasmus, Carl Dahlström, and Mihaly Fazekas (2017). "Procurement and Competition in Swedish Municipalities." Quality of Government (Qog) Institute Working Paper 2017:5.

Colonnelli, Emanuele and Mounu Prem (2017). "Corruption and Firms: Evidence from Randomized Audits in Brazil." Stanford Institute for Economic Policy Research (SIEPR) Working Paper No. 18-060 (November).

Costas-Pérez, Elena, Albert Solé-Ollé and Pilar Sorribas-Navarro (2012). "Corruption scandals, voter information, and accountability." *European Journal of Political Economy* 28(4): 469-484.

Cruz, Cesi, Philip Keefer and Julien Labonne (2018). "Buying Informed Voters: New Effects of Information on Voters and Candidates." Mimeo.

Dávid-Barrett, Elizabeth, and Mihály Fazekas. "Grand corruption and government change: an analysis of partisan favoritism in public procurement." *European Journal on Criminal Policy and Research* (2019) (online).

de la Cadena, Santiago, Aura Cifuentes, Javier Guillot, Philip Keefer, Ruth Persian, Luke Ravenscroft, Benjamin Roseth and Mónica Wills-Silva (2020). "Easy as PAE: Optimizing Text Messages to Encourage Parental Engagement with the School Meals Programme in Colombia." Inter-American Development Bank.

De Vries, Catherine E. and Hector Solaz (2017). "The Electoral Consequences of Corruption." *Annual Review of Political Science* 20:391-408.

Di Tella, Rafael and Ernesto Schargrodsky (2003). “The Role of Wages and Auditing During a Crackdown on Corruption in the City of Buenos Aires.” *Journal of Law and Economics* 46:269-300.

Enikolopov, Ruben, Maria Petrova and Konstantin Sonin (2018). “Social Media and Corruption.” *American Economic Journal: Applied Economics* 10(1): 150-74.

Ferraz, Claudio and Frederico Finan (2008). “Exposing corrupt politicians: the effects of Brazil’s publicly released audits on electoral outcomes.” *The Quarterly Journal of Economics* 123(2): 703-745.

Fried, Brian J., Paul Lagunes and Atheendar Venkataramani (2010). “Corruption and inequality at the crossroad: a multimethod study of bribery and discrimination in Latin America”. *Latin American Research Review* 45(1): 76-97.

Fung, A., Graham, M. and Weil, D., 2007. *Full disclosure: The perils and promise of transparency*. Cambridge University Press.

Gallego, Jorge A., Mounu Prem, and Juan F. Vargas. "Corruption in the Times of Pandemia." Available at SSRN 3600572 (2020).

Jomaa, Lamis H., Elaine McDonnell and Claudia Probart (2011). “School feeding programs in developing countries: impacts on children’s health and educational outcomes.” *Nutrition Reviews* 69(2): 83-98.

Knack, Stephen, and Philip Keefer (1995). "Institutions and economic performance: cross-country tests using alternative institutional measures." *Economics & Politics* 7(3): 207-227.

Lagunes, Paul (2018). “Guardians of Accountability: A Field Experiment on Corruption and Inefficiency in Local Public Works.” Mimeo, Columbia University.

Lui, Francis T. (1986). “A Dynamic Model of Corruption Deterrence.” *Journal of Public Economics* 31:2, 215-236.

Mauro, Paolo. "Corruption and growth." *The Quarterly Journal of Economics* 110.3 (1995): 681-712.

Mironov, Maxim and Ekaterina Zhuravskaya (2016). “Corruption in Procurement and the Political Cycle in Tunneling: Evidence from Financial Transactions Data.” *American Economic Journal: Economic Policy* 8(2): 287-321.

Muralidharan, Karthik, Paul Niehaus, Sandip Sukhtankar and Jeffrey Weaver (2019). “Improving Last-Mile Service Delivery using Phone-Based Monitoring”. Mimeo, University of Southern California.

Olken, Benjamin A. (2007). “Monitoring Corruption: Evidence from a Field Experiment in Indonesia.” *Journal of Political Economy* 115:2 (April): 200-249.

_____ (2009). "Corruption perceptions vs. corruption reality." *Journal of Public Economics* 93 (7-8): 950-964.

_____ and Rohini Pande (2012). “Corruption in Developing Countries.” *Annual Review of Economics* 4: 479-409.

Persson, Anna, Bo Rothstein, and Jan Teorell (20). “Why Anticorruption Reforms Fail – Systemic Corruption as a Collective Action Problem.” *Governance* 26:3, 449-471.

Purroy, Miguel (2021). “Corruption, Transparency, and Natural Resources.” Unpublished manuscript, Universidad del Rosario, Colombia.

Reinikka, Ritva and Jakob Svensson (2005). “Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda.” *Journal of the European Economic Association* 3(2-3): 259-257.

Rossi, Martín, Antonia Vazquez and Juan Cruz Vieyra (2020). “Information Disclosure and the Performance of Public Investment.” Inter-American Development Bank. Discussion Paper No. IDB-DP-795.

Treisman, Daniel (2000). "The causes of corruption: a cross-national study." *Journal of Public Economics* 76(3): 399-457.

Weitz-Shapiro, Rebecca and Matthew S. Winters (2017). “Can Citizens Discern? Information Credibility, Political Sophistication, and the Punishment of Corruption in Brazil.” *The Journal of Politics* 79(1): 60-74.

Winters, Matthew S., Paul Testa and Mark M. Fredrickson (2012). “Using Field Experiments to Understand Information as an Antidote to Corruption.” Chapter 8 in Danila Serra and Leonard Wantchekon (ed.), *New Advances in Experimental Research on Corruption (Research in Experimental Economics, Vol. 15)*, 213-246. Emerald Group Publishing Limited.

Zamboni, Yves and Stephan Litschig (2018). “Audit risk and rent extraction: Evidence from a randomized evaluation in Brazil.” *Journal of Development Economics* 134 (September): 133-149.

Appendix: Text Messages

In both departments, parents were asked in the first week if they would be willing to receive text messages regarding the PAE and were given an option to opt out of receiving further text messages, in accordance with Colombian law.[28] The exact message content and timing was determined based on earlier randomized controlled trials that we used to test alternative messages in eight out-of-sample schools in the two departments (see de la Cadena, et al. 2020).

The second week, parents were informed that their school should have a School Meals Committee; that the committee should be comprised of the principal, three students and three parents; that the committee is empowered to review the performance of the contractor and to request a meeting with the contractor to discuss issues. This message sequence ended with a text asking parents if they could identify the parents serving on the committee. If the parents said they could identify them, they received a concluding message encouraging them to talk with those parent representatives if they had any issues. If they responded that they could not identify them, parents were referred to the principal for more information.

In the third week, parents were informed of the food items their children should have received on the day the text was sent and were asked if the children's meal in fact included those items. If parents responded "yes", they received a concluding text expressing satisfaction about this outcome; if they responded "no", they were asked to consult with their child about whether the child received a substitute food item.

In the fourth week, parents were sent a text that told them how many families in the school had indicated the week before that their children had not received the required food items. They were then asked if they were going to talk to the School Meals Committee about this. If they said yes, they were sent a message to prepare for the meeting and reminded of the importance of participating. If they said no, they were reminded that it is their right to speak with the committee or principal regarding any PAE concerns they might have. encourage them to follow through.

Messages differed between the two departments in the fifth week. In Nariño, the text message shared information with the parents about the results of the audit results from the previous week, focusing on whether the number of meals served corresponded to the number contracted. Parents were asked whether they would speak with the School Meals Committee about this result, with the same follow up messages as before if they responded yes or no. Those results were not available in Cesar, so instead in Cesar

[28] The complete stream of messages and response options are available from the authors.

parents again received a message about the food items their children should have received, as in week three.

In week six of the treatment, parents in Cesar received the audit results, while parents in Nariño were once again encouraged to share the information they had received in the previous two weeks regarding contractor compliance with the principal or the departmental Secretary of Education. The text offered specific guidance about the content of any possible letter they might write: that their children did not receive the appropriate food items on a specific date, asking the authorities what they would do about this, and indicating the name of the school.

In week seven, texts informed parents in Cesar how many parents indicated their children had not received the designated food items, just as in week four. In Nariño, parents were once again told what food items their children should have received. The texts followed the same format as in week three. Similarly, in week eight in Nariño, texts informed parents about what other parents had said the week before, following the model of week four texts. Parents in Cesar were told in week eight about how they could complain about problems with the PAE. In the last two weeks of the semester, in all schools in all departments, treatment and control, the same messages were sent out asking parents about the food their children had received.

Table A1: Treatment Effects on Mean of Last Three Audits of Menu Compliance (substitutions and quantity-shaving)

	Mean menu compliance last 3 audits			
	(1)	(2)	(3)	(4)
Treatment vs. first audit control school	0.0670 (0.0456)	0.0831** (0.0375)	0.0715** (0.0345)	0.0774** (0.0368)
Balance controls	No	No	No	Yes
Observations	209	196	209	196
R^2	0.010	0.65	0.45	0.69
Municipality FE	No	Yes	No	Yes
Operator FE	No	No	Yes	Yes

Notes: Balance controls are type of meal, mode of preparation, number of meals in the contract and the total students of the school. Changes in observations responds to the subsamples conditions and singleton observations are dropped following Correia (2015) when municipality fixed effects are included. Cluster standard errors at the school level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Week Correspondence, Any of Last Three Treatment Audits vs. First Control Audit

Week	No. control schools receiving first audit	No. treatment schools receiving any of last three audits	Menu compliance			*p*-value, significance of difference
			Mean control	Mean treatment	Difference	
5	1	0	0	-	-	-
7	0	1	-	0.4708	-	-
8	0	1	-	0.5000	-	-
9	0	11	-	0.4901	-	-
10	52	56	0.3207	0.4439	-0.1231	0.048
11	7	51	0.3561	0.5400	-0.1840	0.169
12	2	55	0.8370	0.5478	0.2892	0.186
13	10	77	0.9305	0.6678	0.2627	0.020
14	26	28	0.8758	0.8504	0.0254	0.601
15	5	36	0.5624	0.7782	-0.2158	0.135