

## **Voluntary Audits: Experimental Evidence on a New Approach to Monitoring Front-Line Bureaucrats<sup>1</sup>**

Ana L. De La O, Yale University  
ana.delao@yale.edu

Lucas Gonzalez, CONICET/UCA-UNSAM  
lucas\_gonzalez@uca.edu.ar, lgonzalez@unsam.edu.ar

Rebecca Weitz-Shapiro, Brown University  
rbweitz@brown.edu

### **ABSTRACT**

Can opportunities for undergoing voluntary oversight improve bureaucratic motivation and effort? Drawing on insights from the social sciences, we argue that voluntary oversight increases front-line bureaucrats' sense of autonomy and competence, and may therefore increase their motivation and effort. Partnering with a provincial auditing body in Argentina, we implement an encouragement design in which school principals are invited to receive a voluntary audit of a publicly funded school meal program. We employ a two-level randomization, in which regions are first randomly assigned to a higher or lower rate of invitations, and then schools within regions are randomly assigned to treatment or control. We find divergent effects of treatment based on the density of treatment; in the group of regions assigned to the lower rate of invitations, school principals assigned to treatment report increased motivation and a decrease in school closings. In contrast, in the group of regions assigned to the higher rate of invitations, we observe the opposite effect. Drawing on qualitative fieldwork, we speculate that a higher rate of invitations may generate pressure to accept the invitation and thereby undermine any positive effects of volunteering. Our results suggest the promise of voluntary audits as well as the need for further research on the conditions under which voluntary oversight may have differing consequences for bureaucratic effort and motivation.

---

<sup>1</sup> Authors are listed in alphabetical order. Thanks to Don Green, Robert Blair, Matthew S. Winters, Kate Baldwin, Chappell Lawson, Laura Paler, Lily Tsai, Hannah Baron, Jake Bowers, Alexander Coppock, attendees at seminars and meetings at MIT Gov/Lab, EGAP, IAST, Cornell University, and the University of Illinois, for helpful comments and feedback, as well as Anna Nakai, Catlan Reardon, Cecilia Nuñez Raynoldi, Anna Baker, Zoe Ervolino and Nicolas Taccone for excellent research assistance and support and Yi Qi for help with GIS. Brenda Deniz Schneider oversaw the fieldwork and this project would have been impossible without her substantial commitment and contributions. We sincerely thank the Tribunal de Cuentas of the province of Chaco, especially Luis María del Cerro, for their collaboration, as well as Mirta Merlo and her team at the Escuela de Gobierno for their insights into the provincial education system and for carrying out the endline survey. We acknowledge funding from the ISPS Field Experiments Initiative at Yale University and from Brown University. The study's preregistration at EGAP can be found here: <http://egap.org/registration/5832>. This study received and IRB exemption from Brown University (protocol number 1704001741) and Yale University (protocol number 2000021005).

## I. Introduction

Can voluntary forms of oversight improve street-level bureaucrats' motivation and performance? Poor performance of street-level bureaucrats is a common complaint, especially in lower and middle-income countries where public employees often work with limited resources and for low wages (Dasgupta and Kapur 2020; Pepinsky, Pierskalla, and Sacks 2017; Rauch and Evans 2000). In this paper, we ask whether offering these front-line service providers (FLSPs) the opportunity to volunteer to undergo an audit affects motivation and performance. Drawing on a pre-registered randomized field experiment among school principals in the province of Chaco, Argentina, we find that voluntary audits improve motivation and performance when few school principals are invited to volunteer, but they have the opposite effect when many school principals are invited to volunteer. We draw on qualitative fieldwork to interpret these results. Our findings suggest that voluntary audits can work in some circumstances, and that examining the collective dynamics of offering street-level bureaucrats (SLBs) a choice is essential for understanding the potential and limits of voluntary oversight.

Over the past 30 years, bureaucrats on the front lines of service delivery have “become a central plank in the global development agenda” (Pepinsky, Pierskalla, and Sacks 2017).<sup>1</sup> The expansion of direct welfare provision to the poor in much of the global south (e.g., World Bank 2018) means that the choices made by front-line service providers are crucial to the nature and quality of social services that citizens receive. Reflecting the importance of these actors, a growing literature across the social sciences seeks to understand the day-to-day behaviors and beliefs of bureaucrats in lower and middle income countries and the factors that affect their performance in their jobs (see the summaries and calls for further research in Bear and Mathur 2015; Bertelli et al.

2020; Finan, Olken, and Pande 2017; Gans-Morse et al. 2018; Grossman and Slough 2022; Hoag and Hull 2017; Pepinsky, Pierskalla, and Sacks 2017).

Much of this literature draws heavily on the principal-agent framework, with its emphasis on the extent to which bureaucrats' preferences diverge from those of politicians and the public. Empirical work in this tradition focuses on the need to design incentives—whether material, career, or otherwise—to induce effort among individual bureaucrats (see the reviews in Dustan et al. (2018) and Finan et al. (2017)). Other approaches highlight the limitations of the principal-agent framework for explaining bureaucratic behavior, for example by pointing to the emergence of pockets of bureaucratic excellence across or within agencies, even in settings where incentive and oversight structures are fixed (e.g., McDonnell 2020). Even with this diverse and growing literature on bureaucratic effort and motivation in lower and middle-income countries, other sources of motivation continue to receive less attention. In particular, the nature and sources of self-regarding, intrinsic motivation in the public sector, like work satisfaction or enjoyment, remain understudied in both observational and experimental research (Esteve and Schuster 2019).

In this paper, we draw on existing theoretical and empirical work on bureaucratic performance, along with insights from psychology and our field work, to examine whether a novel form of oversight—offering bureaucrats the opportunity to *volunteer* to receive an audit—can affect bureaucratic motivation and effort. We focus on variation across individual bureaucrats while exploring a source of motivation that is not an incentive (whether positive or negative) but instead an opportunity to exercise autonomy and choice. We argue that volunteering for oversight provides bureaucrats with autonomy and a sense of competence, which are associated with greater job satisfaction, effort, and motivation in general, and with greater intrinsic motivation in particular (e.g.,

Cassar and Meier 2018; Esteve and Schuster 2019; Ryan and Deci 2000). We therefore expect positive effects of volunteering on bureaucrats' motivation and effort. To our knowledge, this paper is the first to examine an intervention that explicitly gives FLSPs choice over whether to receive oversight.

We partner with the Provincial Auditing Office (PAO) in Chaco, Argentina, which has authority to oversee public school principals' implementation of a publicly funded meal program. For the purposes of our study, we consider school principals to be SLBs because they interact directly with parents and students around the meal program.<sup>2</sup> Common throughout Argentina, these school meal programs are recognized as crucial for child nutrition, while also suffering from diverse problems in their implementation, many of which are due to the demands they place on school staff beyond their areas of core competency (Diaz Langou et al. 2014). Chaco's PAO typically carries out a very small number of mandatory audits of schools' conduct in the meal program, and many schools report no audits in recent memory.<sup>3</sup>

Our study employs a two-level randomization and an encouragement design on a convenience sample of 188 public pre-K and elementary schools located across 13 geographically-defined regions in Chaco. Within our study group, we first randomly assign each geographically-defined region to either a higher or lower density of treatment group. Within each of these regions, we then randomly assign schools to the treatment group, in which school principals receive an invitation to a voluntary audit of the meal program from the PAO, or to the control group, where schools face a small (but non-zero) likelihood of a mandatory audit. In the group of regions with a higher density of treatment, 50% of schools included in our sample in each region are assigned to treatment and receive the invitation. We refer to this group of regions as the "higher rate of invitation" arm of the study. In regions with a lower density of treatment, 16% of

schools included in our sample in that region are assigned to treatment and receive the invitation; we refer to this group of regions as the “lower rate of invitation” arm of the study. As elaborated below, including this first stage of randomization allows us to investigate possible spillover or peer effects, which we expected to be stronger in the higher rate of invitation group.

We measure the effects of the intervention with a combination of administrative data and a survey of school principals carried out approximately eight months after the intervention.<sup>4</sup> About one-third of school principals assigned to treatment accept the invitation, with higher rates of acceptance in the higher rate of invitation group than in the lower rate of invitation group. All of those who accept the invitation then receive the voluntary audit. We compare outcomes between schools assigned to receive the invitation (regardless of whether they accepted it) and those not assigned to receive the invitation. Focusing on this intent to treat effect ensures that the analysis reflects the randomization process and allows us to estimate the effects of offering voluntary audits at the programmatic level; we expect that any program of voluntary audits would necessarily include a mix of bureaucrats who do and not accept such an invitation.<sup>5</sup>

We find that in the lower rate of invitation arm, school principals who receive an invitation to the voluntary audit report higher levels of motivation and a smaller proportion of school closings after the audits take place. In contrast, we find negative effects of treatment on motivation and effort in the higher rate of invitation arm. Drawing on qualitative field work, we speculate that a higher rate of invitations creates peer pressure to accept the invitation to the audit. This pressure may undermine bureaucrats’ sense of autonomy and competence, the very mechanisms through which we posit that a voluntary oversight process can create positive effects.

The rest of the paper proceeds as follows. We first review existing literature on the oversight of and motivation among FLSP in order to develop our theory of voluntary oversight. We then detail the social and bureaucratic context of our research site. Next, we summarize our experimental intervention, followed by a presentation and discussion of results and their implications for understanding bureaucratic effort and motivation.

## **II. The difficulty of motivating front-line bureaucrats**

Under what conditions will bureaucrats be motivated and exert effort in their jobs? In order to answer this question, one research tradition within the social sciences draws heavily from the principal-agent framework (Stiglitz 1989) and postulates that variation in the incentives that bureaucrats face shapes their performance. Empirically, a growing experimental literature uses interventions among FLSPs in the Global South to explore the effects of financial or in-kind incentives (positive or negative), often in combination with mandatory monitoring, on variation in performance across individual bureaucrats (see Finan, Olken and Pande (2017), Pepinsky, Pierskalla and Sacks (2017), and Dustan, Maldonado and Hernandez-Agramonte (2018) for reviews).<sup>6</sup> In certain circumstances, these studies show that these forms of oversight can work—for example, there is evidence they can generate gains in teacher and health worker attendance and student outcomes in contexts including India (i.e., Duflo, Hanna, and Ryan 2012; Muralidharan and Sundararaman 2011) and Guinea (Barrera-Osorio et al. 2022). At the same time, this research program provides evidence of the limitations and unanticipated consequences of this approach to motivating bureaucrats. Gains in bureaucratic performance may be limited to the incentivized metric (e.g., Glewwe, Ilias, and Kremer 2010), or to politically relevant districts (Callen et al. 2016). Monitoring and incentives may even produce perverse effects, wherein front-line bureaucrats collude with their

supervisors (e.g., Banerjee, Duflo, and Glennerster 2008) or citizens (Khan, Khwaja, and Olken 2016) to subvert and undermine monitoring systems.

The limitations of existing models of oversight have prompted some recent experimental studies that provide evidence that novel forms of oversight, including information and non-financial inducements, can increase the effort exerted by bureaucrats and other public-service providers (Ashraf, Bandiera, and Jack, 2014; Bandiera et al. 2020; Dustan, Maldonado, and Hernandez-Agramonte 2018). This shift in the experimental literature resonates with a separate literature that highlights the limitations of the principal-agent approach and places greater emphasis on a diverse range of explanations of bureaucratic motivation and effort. At the macro-level, this work has explored the correlates of the emergence of a bureaucratic “esprit de corps” either across states (Rauch and Evans 2000) or within unique bureaucratic subcultures within a state or even an agency (e.g., Bersch, Praça, and Taylor 2016; McDonnell 2020). While these findings resonate with work from public administration that calls attention to the importance of public sector motivation (e.g., Perry, Hondeghem, and Wise 2010; Perry and Wise 1990), they place less emphasis on variation in individual-level bureaucratic effort.<sup>7</sup> This tradition has also tended to focus on variation in the performance of middle and higher-level bureaucrats located in policy-making bureaucracies rather than among bureaucracies composed of front-line service providers (though see Tandler (1997) and Davis (2004) for important exceptions).

Many open questions remain about the sources of individual-level variation in bureaucratic motivation and effort among FLSP in the Global South. If understanding the nature and sources of bureaucratic motivation remains the “Holy Grail” (Esteve and Schuster 2019) of public management even in the long-standing wealthy democracies,

this is even more true among front-line service providers in lower- and middle- income countries.

### **III. A Theory of Voluntary Oversight**

In this paper, we build on these distinct research approaches to public sector performance. We share with experimental approaches a focus on variation in individual-level bureaucratic performance, and with sociological approaches an emphasis on the responsiveness of public sector employees to non-incentive based forms of motivation (whether remunerative or not). We also rely on insights from psychology that demonstrate the importance of autonomy and competence to individual motivation and effort in diverse settings. We bring these together to develop a theory of how voluntary government audits may increase motivation and effort among front-line public sector workers.

Public audit bodies or oversight agencies are independent governmental organizations charged with monitoring other government actors, including elected politicians and bureaucrats, for the proper use of government funds.<sup>8</sup> These bodies are increasingly prominent in lower and middle-income countries, where they are often characterized as a solution to public sector corruption (Dye and Stapenhurst 1998; Santiso 2006). Importantly, government audits are understood to be mandatory by definition.<sup>9</sup> Voluntary audits do exist in the private sector, although their presumed effectiveness largely relies on their role in market signaling (e.g., Lennox and Pittman 2011), which would not be relevant for public sector actors. In existing theoretical frameworks, public audits' effectiveness derives from their capacity to detect and punish non-compliance and malfeasance, and hence to deter them (e.g., Gans-Morse et al. 2018).<sup>10</sup>

On the whole, existing research suggests a mixed record for government audits. In some cases, these audits can facilitate public accountability (e.g., Ferraz and Finan 2011). Elsewhere there is evidence that audits may fail to have the desired effects or even have perverse effects, due to factors including weak state capacity, collusion between auditors and audited, political pressure on auditors, the displacement of employee effort away from audited activities, or the reduction in staff morale (De La O, Fernandez Vazques, and Martel Garcia 2022; Gonzalez-Lira and Mobarak 2019; Hidalgo, Canello, and Lima-de-Oliveira 2016; Olken 2007; Wang forthcoming).

In existing theoretical frameworks, allowing bureaucrats to select into public audits would seriously undermine their effectiveness, as bureaucrats who believe they would perform poorly would simply not volunteer for oversight. In contrast, we argue that participating in a voluntary public-sector audit can increase motivation and effort through a different route. We define a voluntary public sector audit as one in which a public servant who represents some public body (such as a school, hospital, municipal government, or administrative office) can freely choose whether or not to receive an audit administered by a public auditing or oversight body. How might such a system affect bureaucratic effort?

We know from self-determination theory in psychology that individuals value tasks that provide them with a sense of autonomy, competence, and relatedness to others (Deci and Ryan 1985; Ryan and Deci 2000). There is evidence from across the social sciences that people report greater job satisfaction and exert more effort in jobs that address these three psychological needs (Bandiera et al. 2020; Cassar and Meier 2018; Petter et al. 2002). Greater autonomy may particularly increase intrinsic motivation, generally defined as motivation to do some task because it is “inherently interesting or enjoyable,” as opposed to extrinsic motivation, which refers to the

motivation to carry out a task “because it leads to a separable outcome” (Ryan and Deci 2000, 55).<sup>11</sup> This link between autonomy and intrinsic motivation has been identified with respect to motivation to perform tasks in general settings (Zuckerman et al. 1978) as well as within the workplace (Cassar and Meier 2018), and Esteve and Schuster (2019) expect this logic to hold for public sector workers, as well. We argue that participating in a voluntary audit should help FLSPs exercise autonomy and feel competent in their jobs, and consequently should increase motivation (perhaps particularly intrinsic motivation) and effort.

Beginning with the psychological need to feel autonomous, a voluntary audit creates an opportunity for self-direction and increases control by giving bureaucrats a real choice as to whether to receive oversight or not. This is in sharp contrast with conventional forms of government oversight of public employees (including but not limited to audits), which are obligatory and have long been linked to bureaucratic resentment and efforts to resist such oversight (Gans-Morse et al. 2018; e.g., Lipsky 1980, 19; Prottas 1978). In contexts where autonomy on the job is limited, bureaucrats may be especially responsive to opportunities to exercise some control over the nature of oversight they receive.<sup>12</sup>

Voluntary oversight also offers bureaucrats an opportunity to feel competent and to display that competence to others. The act of volunteering inherently creates opportunities for both personal and social recognition, as FLSPs who volunteer can conceive of themselves as having made a positive choice, and they can also share their decision to volunteer with others. Even if, for example, auditors follow the same procedures whether an audit is mandatory or voluntary, bureaucrats being audited are likely to experience a greater sense of personal satisfaction and have greater possibilities for receiving social recognition when they have volunteered for that audit. Personal and

social recognition for public-serving tasks may be motivating in and of themselves and also, as Ellingsen and Johannesson (2007) argue, such recognition fosters a sense of competence.

Voluntary oversight may also increase FLSPs' felt sense of competence by allowing bureaucrats to offer meaningful feedback to policymakers. Although it has received somewhat less attention than other factors, studies of public sector workers show that they value the opportunity to give input into policy formulation and decision-making (Perry and Wise 1990; Petter et al. 2002; Rasul and Rogger 2018).<sup>13</sup> Accepting an invitation for oversight might present street-level bureaucrats with the opportunity to share their experiences on the front lines of service delivery with those farther up the hierarchy, and thus to contribute their perspective to policymakers. This should be especially appealing to FLSPs given the top-down, hierarchical nature of many modern bureaucracies.

In summary, participating in a voluntary audit offers street-level bureaucrats the opportunity to exercise autonomy and feel competent in their professional roles. By increasing a sense of autonomy and competence, the experience of voluntary oversight can lead to increased bureaucratic effort and motivation. Nonetheless, there are limits to this new approach. Notably, we would not expect corrupt officials to volunteer for audits if they believe this will increase their risk of punishment. Voluntary audits have the most promise in contexts where stakeholders identify variation in motivation and effort (e.g., Dasgupta and Kapur 2020), rather than malfeasance or corruption, as primary concerns.

#### **IV. Oversight and motivation of school principals in Chaco, Argentina**

We study the effects of voluntary audits by conducting an experimental intervention into the oversight of school principals' implementation of a publicly funded

meal program in the province of Chaco, Argentina. Located in the northeast region, Chaco is one of Argentina's poorest provinces, with a GDP per capita less than half the national average (Ministerio de Economía y Finanzas Públicas de la Nación 2015). Both under and over nutrition are a problem in Chaco; it is one of the few Argentine provinces where over 5% of young children were found to be underweight in a Ministry of Health study (Dirección Nacional de Maternidad e Infancia 2006, 52).<sup>14</sup> Free school meals are considered an important policy intervention in Argentina and the government devotes significant resources to nutrition programs; in Chaco, 14% of the province's Ministry of Social Development budget was dedicated to these programs in 2021.<sup>15</sup> Providing meals also places significant demands on school principals: above and beyond their other pedagogical, administrative, and supervisory roles, school principals are responsible for managing foodstuffs, coordinating and overseeing food preparation on site, interacting with students and parents about the menu and food served, ensuring that food is distributed to students during each school day.

In this context, field work carried out both prior to and after our intervention suggested there is notable variation in how much work school principals invest in the meal program. In interviews, many school principals contrasted their own self-reported significant effort with lower levels of motivation and effort exerted by their peers. According to one principal, for example, "there are directors for whom the meal program is a concern, and others for whom it is not." Another communicated that, with respect to the meal program, "unfortunately not all the schools and all the principals did things right."<sup>16</sup> Although self-reports of this kind should be treated with caution, they offer useful qualitative evidence that principals believe motivation and effort vary amongst their peers. These interviews, as well as background discussions with our field partner, the PAO, also suggest that although corruption occurs in this program, it is believed to be

uncommon, and the main dimension of variation is between more or less effort, rather than the presence or absence of malfeasance.<sup>17</sup> This is the type of setting in which voluntary audits might be effective.

### *Existing oversight*

Chaco's educational bureaucracy is extremely hierarchical, with limited rewards or recognition for good performance and established, if infrequently applied, threats of punitive action for poor performance. In interviews, school principals repeatedly emphasized that their performance on the job is not part of any advancement process and that there are no incentives or rewards, or even public recognition, for particularly good performance.<sup>18</sup> The provincial Ministry of Education does have standard processes for reviewing accusations against school principals for malfeasance. In qualitative interviews, all principals we interviewed were aware of these procedures, although financial and human resource constraints within the provincial government mean that the frequency of oversight that might result in punishment remains quite low. For example, of the 126 schools in the control group in our study, the provincial auditing body selected four, or about 3%, for a surprise, mandatory audit in 2018.<sup>19</sup> In qualitative interviews, principals repeatedly shared that they rarely, if ever, receive any type of feedback (positive or negative) from the provincial Ministry of Education.<sup>20</sup>

### *Self-Perceptions of Motivation and Effort*

While the institutional design of Chaco's educational bureaucracy privileges top-down supervision, our field work suggests that motivation among school principals might be more responsive to a different logic. In interviews, school principals emphasized the pro-social nature of their work and the inherent value of the tasks they completed, especially in a context of significant needs. In our endline survey, described in more detail below, we ask principals to rate the importance of various types of motivation in

explaining the effort they invest in the meal program and in supervising teachers. For both tasks, intrinsic motivation significantly exceeds extrinsic motivation. For example, average intrinsic motivation for supervising teachers is 3.5 on a 4-point scale, as compared to an average of 2.5 for extrinsic motivation.<sup>21</sup> These data also show significant variation in motivation across individuals. For example, for intrinsic motivation in the meal program, the bottom 25% of principals report an average motivation level of 2.8 or below, while the top 25% report an average motivation level of 3.4 or above.

Local observers emphasized that school principals are often highly respected leaders in their neighborhoods and communities, and this embeddedness is also reflected in school principals' own assessment of their most important audiences. In spite of the top-down bureaucratic structure of formal oversight, principals are more attentive to the opinions of parents, students, and teachers than to their professional superiors. For example, our endline survey asked principals, "At the time of being held accountable for your general performance in the school, whose opinion matters to you most?" When asked to name the single most important audience, 45% of principals named parents, followed by 22% who named students, and 15% who named teachers. 16% name their direct supervisors, and only 3% named the provincial Ministry of Education. The provincial audit body received no mentions at all. School principals do not appear to be primarily motivated by the institutional forms of hierarchical oversight that are in place.

Finally, the potential resonance of forms of oversight that appeal to other sources of motivation was also evident in the expressed desire on the part of many school principals to share the knowledge gleaned from their "on the ground" experiences with politicians and policymakers. Many principals expressed frustration that higher level officials rarely visited schools and apparently failed to understand the difficulty of effectively serving students with the resources available to them.<sup>22</sup> Other principals noted

they had sent written requests for more assistance for the meal program, but that these went unanswered by the provincial government.

In summary, then, our field research suggests a significant tension between existing forms of oversight and the nature of principals' motivation. While existing oversight is hierarchical, mandatory, and infrequent, school principals emphasized the pro-social nature of their jobs and being motivated by the inherent value of their work. In Chaco, existing forms of oversight do not appeal to school principals' pro-social motivations or give them an opportunity to exercise choice or share their expertise. This suggests an opportunity for new approaches to oversight that will better appeal to important components of school principal motivation.

#### **IV. The Intervention**

In this context, we test our theory about the effects of voluntary oversight through an experimental encouragement design in which some randomly selected schools received an invitation to undergo a voluntary audit from Chaco's Provincial Auditing Office (PAO), which is legally empowered to oversee the implementation of the school meal program. We study a convenience sample of 188 urban and semi-urban schools across 13 regions in the province. The 13 regions include 7 municipalities and the provincial capital of Resistencia, which we divided into six geographically compact areas for the purpose of our study. The regions included in our sample were determined by our government partner's capacity to travel from its offices in the provincial capital.<sup>23</sup> The intervention was carried out jointly by the PAO and the research team. The research team implemented the random assignment to treatment and control and offered logistical support to the PAO, while the PAO mailed the letters of invitation and carried out the audits. The PAO did not include any mention of the research team while carrying out

these tasks. The study had a two-stage design: first, we randomly assigned each region to be a part of either the higher or lower rate of invitation group with a probability of .5. Then, within each region, we randomly assigned schools to treatment and control. In the higher rate of invitation group, the probability of treatment was .5; in the lower rate of invitation group, it was .16. A total of 126 schools were assigned to the control group and the remaining 62 assigned to the treatment group.

The inclusion of one arm with a higher rate of invitation and one arm with a lower rate of invitation in the intervention allows for the examination of possible spillover and/or peer effects. School principals interviewed during our fieldwork shared that they are often in communication with one another, primarily via WhatsApp groups. To the extent this communication affects our intervention, we expect that would be in one of two ways. The first would be a conventional spillover effect, wherein some school principals in the control group learn about the treatment and therefore behave as if they were treated. We believe this is unlikely because of a key feature of our intervention—only those schools that received the intervention could actually volunteer for an audit from the PAO.<sup>24</sup> Still, it is possible that some school principals in the control group might change their behavior in anticipation of receiving an invitation to an audit at some point in the future.

The second possibility is that the treatment generates differential peer pressure *among* those principals assigned to treatment in areas with a high versus low rate of invitation. In areas with a high rate of invitations sent, principals should be exposed to more messages from their peers about the audit invitation. This increased communication about the voluntary audits could generate peer pressure to accept the invitation, which would lead to higher take-up rates of the invitation to the voluntary audit in these areas. We explore some of the empirical implications of these two

possibilities—conventional spillovers, on the one hand, and peer-pressure among the treated, on the other hand—in more detail below.

In both arms, the treatment group was encouraged to volunteer for an audit through a written invitation, which the PAO sent on official letterhead via postal mail in July 2018.<sup>25</sup> The encouragement was effective: 24 out of the 62 invited school principals agreed to volunteer. In areas with a lower rate of invitation, 25% of invited principals accepted the audit; in higher rate of invitation areas, 43% accepted. Given common assumptions that bureaucrats shirk oversight whenever possible, we view this as a relatively high rate of acceptance.<sup>26</sup> All of those who volunteered received an audit, at a time coordinated between the principal and the PAO staff in advance. To formally test whether assignment to treatment lead to a voluntary audit, we specified a separate regression for each arm of the study where the dependent variable takes the value of 1 if the school principal received a voluntary audit, and zero otherwise. The independent variables included assignment to treatment and region fixed effects. These models confirm that assignment to treatment is a significant predictor of receiving an audit; schools assigned to treatment were 22 percentage points more likely to receive a voluntary audit compared to control schools in the lower rate of invitation areas, and 43 percentage points more like in the higher rate of invitation areas. Both coefficients are statistically significant at the .05 level. Table A8 in the SI file reports these regressions.

All audits took place between August and October 2018. During the audits, PAO auditors reviewed and collected information on spending and receipts and administered a short interview and questionnaire with the school principal. The questionnaire covered a variety of topics, including the frequency of unanticipated school closings, the menu served over the past five days, and the principal's knowledge of any past audits by the PAO. Upon the completion of the visit, the audit team also delivered an official document

(*Acta de Cierre*) indicating that the school had volunteered to receive the audit team, describing the accounts that were audited, and noting that the audit had been completed.<sup>27</sup>

Schools in the control group were subject to the PAO's default system of oversight; using its own criteria, the PAO additionally audited 4 of the 126 schools in this group. These audits were mandatory and carried out by surprise, and the PAO did not leave an official document from the audit team at the end of these audits.<sup>28</sup> We have no evidence of crossover from control to treatment—no school in the control group received the letter of invitation or a voluntary audit arranged in advance. In addition, no school assigned to treatment received a mandatory, surprise audit.

As noted above, this is an encouragement design wherein schools assigned to treatment are encouraged to participate in a voluntary audit. From a theoretical perspective, we can think of the treatment as having two components: the invitation to participate in the audit and the actual experience of the audit for those who accepted the invitation. We believe the latter is likely to be more consequential, as it involved a sustained interaction with members of the PAO, with whom school principals rarely, if ever, interact. As discussed below, the time elapsed between invitation and audit, along with the temporal nature of some of our endline outcomes, allows us some insight into this question and supports our intuition about the relative importance of the actual audit experience.

#### *Balance Tests and Endline Data*

We rely on a mix of administrative and researcher-collected data for our analyses. Administrative data was collected from the provincial Ministry of Education prior to commencing the study and includes data from 2016 including school enrollment, the nature of meals served at the school, the type of school (preschool or elementary

school), and its geographic coordinates (used to calculate each school's distance from the provincial capital). Additional original data was collected during an endline survey of school principals carried out in June-July 2019. This was nearly a year after the intervention started—letters of invitation to the treatment group were sent in July 2018—and eight months after the intervention was completed (the final voluntary audits were carried out in October 2018).<sup>29</sup> The endline survey was administered by a team from an independent public academic institution (the Escuela de Gobierno of the province of Chaco) with the approval of the provincial Ministry of Education. The survey team had no relationship with the PAO team that implemented the intervention, and the survey covered a large number of topics so as not to call attention to questions about the meal program. Only 5 of the 188 school principals in our sample refused to answer the endline survey: 3 in control and 2 in treatment. Attrition is statistically indistinguishable across treatment and control in both the lower and higher rate of invitation groups. We also find no evidence that item non-response is systematically related to assignment to treatment.<sup>30</sup> The survey also collected information on school attendance on a randomly selected set of dates before and after the intervention. School principals in Chaco usually record attendance in a standard notebook, which they consulted during the endline survey in order to answer questions on attendance.<sup>31</sup> Thus, although this information was collected during the endline survey, it reflects contemporaneous record-keeping.

Before turning to the main outcomes of theoretical interest, we conduct a series of balance tests to compare the treatment and control groups prior to the intervention on a variety of school-level characteristics. We check for balance across 12 variables measured at the school level. These include 2016 school enrollment, the nature of meals served at the school, the type of school, and whether the school is within a 100 or 200km radius from the provincial capital. Others were collected via our endline survey but reflect

pre-treatment information, including school closings and school attendance prior to the intervention, an index that measures how embedded the school principal is in the school's community (i.e. years working in the school, neighborhood of residence, own children attend the school, among others), and an index of the sociodemographic characteristics of the school neighborhood, which combines the assessment of school demographics by the enumerator and survey respondent.<sup>32</sup> In Table 1, we compare values for these variables between treatment and control schools within the lower and higher rate of invitation groups, respectively. The table shows that randomization produced balance in baseline characteristics across the treatment and control groups for both groups. The final row shows the result of an F-test in which we regress assignment to treatment within the lower and higher rate of invitation groups with all the baseline variables and region fixed effects. The final column shows the result of F-tests from weighted least squares regressions in which we regress each baseline variable on indicators for three of the four experimental groups and a constant, with observations weighted by inverse probability of treatment because probability of treatment is different in the lower and higher rate of invitation groups. Not surprisingly given the small sample size, there are some differences across columns, but these tests indicate that they are not statistically significant. As a robustness check, we also run all the analyses presented below with control variables selected via lasso, as specified in our pre-analysis plan. Results are substantively similar in all cases and are presented in Supplementary Information file.<sup>33</sup>

**\*\*Insert Table 1 about here\*\***

## **V. Measuring Effects on Motivation, Effort, and School-level Outcomes**

The study allows us to draw causal inferences on our main theoretical question of interest: whether voluntary audits generate a direct effect on school principals'

motivation and effort, and possibly downstream effects on school-level outcomes. As outlined above, we expect that principals in the treatment group will be more motivated and exert more effort in the implementation of the school meal program compared to the control group which, in turn, could affect downstream outcomes for the children attending and receiving food at these schools.<sup>34</sup>

We present results for three broad categories of outcomes. The first captures school principal's self-reported intrinsic and extrinsic motivation, both in the school meal program and, as a placebo test, in supervising teachers. We expect that the treatment, which focused explicitly on the meal program, will increase principal motivation with respect to the meal program, but it will have no effect on principal motivation to supervise teachers. The second category examines school principals' self-reported behavior, captured by their self-report of hours worked, both in the school meal program and in other areas. Finally, the third category measures unanticipated school closings, which, as we explain below, can serve as a proxy for school principals' effort.<sup>35</sup> While we do not measure child learning directly, school closings may have negative effects on learning, along with the broader social roles that school serve (e.g., Huebner et al. 2014).

All the analyses below follow the same basic structure. For all outcomes of interest, we estimate and report the intent to treat effect, which compares schools randomly assigned to treatment, regardless of whether or not they accepted the audit invitation, with those assigned to control. Because randomization occurs at the assignment to treatment stage, this ensures our results are not biased if school principals who accept the invitation are different from those who do not.<sup>36</sup> For any such program, we would expect incomplete take-up of an invitation to volunteer for oversight, and thus the effects of such a program should account for the fact that many of those invited will not avail themselves of the opportunity presented to them.

For each outcome, we estimate the intent to treat effect separately for the higher and lower rate of invitation areas. For each set of schools, we run OLS regressions with an indicator for assignment to treatment, region fixed effects and robust standard errors.<sup>37</sup> The region fixed effects take into account our blocked randomization. The bottom of each column notes the mean value of the outcome variable in control schools in either low or high density regions, depending on the regression. As a robustness test, in the last row, we also report the randomization inference p-value associated with treatment, which does not rely on assumptions about the distribution of the outcomes or the size of the sample.<sup>38</sup>

In the SI file, we present the raw differences in means between each subgroup (T and C in areas with a higher and lower rate of invitation, respectively), along with additional pre-registered analysis including OLS regressions with lasso-selected controls and weighted least square regressions with pooled data.<sup>39</sup> We also include in the SI file the complier average causal effect, which we estimate with two-stage least square regressions in which we use assignment to treatment as an instrument for actual treatment. We also carried out additional analyses to test for the possibility of spillover effects, which we discuss in more detail below.

### *Motivation*

We measure principals' motivation in two separate tasks: working on the meal program and supervising teachers. We expect effort in the meal program to be affected by the voluntary audit invitation, as the PAO is charged specifically with auditing how schools carry out the meal program. In contrast, and as noted in our pre-analysis plan, we treat effort in supervising teachers as a placebo test; the provincial Ministry of Education supervises school principals in this realm and the PAO has no authority here. We examine

intrinsic and extrinsic motivation separately in order to explore whether the voluntary audit program may have differential effects on these two types of motivation.

In order to create measures of intrinsic and extrinsic motivation for each task (working in the meal program and supervising teachers) from the endline survey, the enumerator asked the school principal to consider why she dedicated time to that task. The enumerator then read the principal a list of 15 reasons, one at a time. The principal responded on a 4 point scale, indicating the extent to which that given reason was true for her personally (ranging from “nada cierta” to “muy cierta” in Spanish).<sup>40</sup> The scales were constructed drawing on existing work in psychology (Ryan and Connell 1989; Ryan and Deci 2000) and each reason was pre-classified as corresponding to extrinsic or intrinsic motivation. So, for example, for the school meal program, statements that the respondent works on the program because “those are the rules” or because “I want parents to appreciate the work I do” are coded as capturing extrinsic motivation. In contrast, statements that the respondent works on the program because “I like to do it” or because “I’m interested in understanding how the canteen works” are coded as capturing intrinsic motivation. For each set of items, we calculate the mean response to create a corresponding index of motivation, and higher numbers indicate greater motivation in this area. The scales were independent; responses were not constrained such that higher levels of intrinsic motivation would decrease extrinsic motivation or vice versa.

**\*\*Insert Table 2 about here\*\***

Table 2 presents the results. The results when we pool the lower and higher rate of invitation groups, which show no overall effect, are available in the Supplementary Information file. The null effect with the pooled data masks distinct effects of treatment within the higher versus lower rate of invitation areas, which are presented here. Columns 1-4 present the results for motivation within the meal program (extrinsic motivation in columns 1-2 and intrinsic motivation in columns 3-4), while columns 5-8 present the placebo tests, measuring extrinsic and intrinsic motivation to supervise teachers.

Beginning with motivation to exert effort in the lunch program, the results show a contrasting pattern in terms of the effects of treatment in lower versus higher rate of invitation areas. In the former, where a small number of school principals received the audit invitation, assignment to treatment increases intrinsic motivation significantly (p-value= 0.009). This increase in motivation of .34 points is also substantively significant on the four-point scale we employ. Although not statistically significant, extrinsic motivation also increases in the treatment group in these areas (p-value= 0.275). In contrast, assignment to treatment in higher rate of invitation areas has the opposite effect. In these regions, it decreases school principals' internal motivation by .20 points (p-value= 0.113 ) and extrinsic motivation by .25 points (p-value= 0.042).<sup>41</sup> These results offer evidence that intrinsic and extrinsic motivation to work in the meal program are not substitutes.

Columns 5-8 present our placebo tests. They show, as expected, no result of the treatment on either internal or external motivation to supervise teachers. In low density areas, the coefficients on extrinsic and intrinsic motivation are signed in opposite directions, and the coefficients on assignment to treatment are smaller than those observed for motivation in the school meal program in all four regressions.

*Principal effort: Hours worked*

Next, we examine school principal behavior, measured through the number of hours the school principal reports working total and on a variety of specific tasks. To improve data validity, the endline survey asked the school principal to think about the week prior to the survey when answering these questions.<sup>42</sup> Principals were asked to choose from a range of hours, and then prompted to state, within this range, the specific total number of hours they had worked on “activities related to the school” in the previous week. They were then asked how many hours they spent working on the meal program, along with on a series of other tasks, including supervising teachers, speaking to parents, and working on curricular or pedagogical tasks.<sup>43</sup> Finally, they were asked how much of their working time they considered to be “extra” hours. For the present analysis, we examine the effects of treatment on total hours worked, hours worked on the meal program, “extra” hours, and hours worked in other activities. As indicated in the PAP, we expected a positive effect of treatment on hours worked overall and in the meal program, and we had no clear expectations on how treatment might affect hours worked in other activities.

**\*\*Insert Table 3 about here\*\***

Table 3 presents the results. As the summary statistics for the comparison group mean shows, principals in the control group report having worked about 46-47 hours in the week prior to survey implementation. They report spending on average about five hours on the administration of the meal program, and 9-10 extra hours. Here, in contrast to our results for motivation, we find no statistically significant effect of treatment on hours worked in either high or low density areas.<sup>44</sup> When we look specifically at hours worked in the meal program, columns 3 and 4 show that the direction of results are consistent with those of Table 2. School principals of treatment schools in low density areas report working about 1.4 hours more on the meal program as compared to principals of control schools in low density areas. Within high density areas, in contrast, assignment to treatment is associated with a reduction of .19 hours worked in the meal program. Nonetheless, these coefficients are not estimated

precisely. Given the long time elapsed between the experimental intervention and the endline survey, we cannot evaluate if the treatment had short term effects on hours worked. However, we can say that we find no evidence of a sustained effect on hours worked.

### *School-level outcomes*

Finally, we examine whether the treatment had any downstream effect on school-level outcomes. Here, we focus on unanticipated school closings—that is, days a school is expected to be open but nonetheless reports zero attendance in the contemporaneous record books. Data from qualitative work suggests a number of ways in which school principals exercise discretion over unanticipated closures. In situations in which parents are uncertain about whether or not to send their children to school, notably during days of inclement weather and teacher strikes, school principals' choices affect whether a school effectively stays open or not. Families in Chaco are often reluctant to send students to school on days of heavy rain, due to the poor conditions of unpaved roads. Principals who choose to communicate to families that the school will not serve meals on those days effectively close the school. Similarly, principals' decisions during teacher strikes affect whether a school remains open to students or not. During the period under study, teacher strikes in Chaco were carried out with teacher presence in the schools—that is, teachers who adhered to a strike would come to the school building but would not lead instruction or take attendance. Facing a full or partial strike, principals could choose not to serve meals, leading parents to keep children at home.<sup>45</sup> Alternately, principals could choose to cover classrooms without teachers and provide meals, in which case at least some students would attend.

We collect data on “zero attendance” days during our endline survey using contemporaneous record books that school principals maintain to track attendance (see Figure B3 in the SI file for a photo). During the endline survey, we asked school principals to report attendance for 16 dates—1 randomly

selected day for each of 16 weeks, beginning with the 4 weeks prior to the date when the voluntary audit invitations were mailed, and continuing with 12 subsequent weeks.<sup>46</sup>

**\*\*Insert Table 4 about here\*\***

In Table 4, we examine the effect of assignment to treatment on unanticipated school closings. For our analyses, the outcome takes on the value of the proportion of dates for which a school reported zero attendance.<sup>47</sup> In Columns 1 and 2, we examine closures for all weeks after PAO conducted the voluntary audits. For the lower rate of invitation group, schools assigned to treatment report somewhat fewer closings, although this difference is not statistically significant (p-value= 0.146). In the case of the higher rate of invitation group (column 2), assignment to treatment increased unanticipated school closings by 2 percentage points (p-value= 0.013).<sup>48</sup>

As noted above, our intervention can be considered to have two components: the invitation to the audit and the audit itself. Data on unanticipated school closings allows us to investigate which of these components is more consequential. In Table 4, columns 3 through 6 we examine the effects of treatment in the two weeks after the audit (columns 3 and 4), and in the two weeks after the invitation to the voluntary audit is received (columns 5 and 6). Columns 3 and 4 again show different effects of treatment in low versus high rates of invitation areas. In the low-rate of invitation arm, schools assigned to the treatment condition have about 4 percentage point fewer closures (p-value=0.035).<sup>49</sup> In contrast, among schools assigned to the high-rate of invitation arm, those assigned to treatment report about a 4 percentage point increase in school closures immediately after the audit (p-value=0.041). On the other hand, we see no significant effects on unanticipated closings in either the higher or lower-rate of invitation groups in the two weeks after the invitation. One caveat to keep in mind is that there are very few school closings in the two weeks after the audit and after the invitation. We interpret these results with some caution to mean that the experience of the voluntary audit itself, rather than the written invitation to participate in the audit, was the stronger component of treatment. Finally, the comparative

strength of these results immediately after the audit offers some tentative evidence that effects of the intervention may diminish with time.

## **VI. Discussion**

Of the results reported in the section above, perhaps most striking is the finding that assignment to treatment has divergent results across areas with a higher versus lower rate of invitations. With respect to motivation to work in the meal program and unanticipated school closings, the treatment had significant effects in the expected direction in the lower rate of invitation group only. In these areas, principals of schools assigned to treatment report higher levels of intrinsic motivation to work in the meal program—a statistically and substantively significant effect that we detect in a survey conducted fully eight months after the intervention. Contemporary records also show that these treated schools report fewer unanticipated closings in the weeks immediately after the audit. Furthermore, although not statistically significant, principals in treated schools in the lower rate of invitation group also report working more hours on the meal program. In sum, in areas with a lower rate of invitation, the effects of the voluntary audit encouragement are consistent with our theoretical expectations.

In contrast, the effects of treatment in areas with a higher rate of invitation to the voluntary audit appear to be the opposite of the anticipated. In these regions, principals of treatment schools report lower motivation to work in the meal program eight months after the intervention. Furthermore, contemporaneous record keeping shows that these schools experienced more unanticipated closings. Finally, although not statistically significant, principals in treated schools report working fewer hours on the meal program than principals in control schools in the higher rate of invitation group.

Why would the intervention have the anticipated effect in the group where relatively few schools were assigned to treatment, but the opposite effect in the group where many schools were assigned to treatment? We consider a number of possible explanations for these divergent results.<sup>50</sup> As noted above, we have no evidence of crossover from treatment to control—that is, no school assigned

to control receive a voluntary audit. Nonetheless, it is possible that some control schools learned about the invitations and changed their behavior as a result, perhaps anticipating that they would receive such an invitation soon. If this occurred, it should be more likely in areas with a higher rate of invitation, which might explain the differential effects of assignment to treatment in the group with a higher versus lower rate of invitation. We conduct a number of analyses to explore this possibility. First, we compare the values of our outcome variables across schools assigned to control in the higher versus lower rate of invitation groups. This analysis may offer some evidence of spillovers if values across the two groups are statistically different. The difference in means between control groups, reported in the SI file (Table A10), show no differences for motivation or hours worked, but they provide some evidence that school closings were higher in the control group in lower versus higher rate of invitation.<sup>51</sup> To investigate this further, we use baseline data on school closings and run an additional analysis where we use the difference in school closings (post-audit minus baseline) as our dependent variable. These results, presented in Table A10b, show no significant differences between the control groups.

In the absence of compelling evidence of a conventional spillover effect, we focus instead on the possibility of what we call “peer” effects among treatment schools in areas with a higher rate of invitation. It is possible that in the group where many schools were assigned to treatment, communication among principals created pressure for school principals to accept the invitation. Unfortunately, we do not have direct evidence that speaks to this possibility. However, we do note that schools assigned to the treatment group in the high invitation rate group accepted the invitation at a higher rate than those assigned to treatment in the low invitation group. In the former, 43 percent of school principals assigned to treatment accepted the voluntary audit, whereas in the latter 25 percent of school principals assigned to treatment accepted the voluntary audit.

A peer pressure effect could lead to negative effects of treatment in areas with a higher rate of invitation areas in a number of ways. The first would be via a direct effect. We assumed in designing our study that principals would treat the invitation to a voluntary audit as, in fact, voluntary. However, it

might be the case that principals were skeptical of the voluntary nature of the invitation and believed there was an expectation that they would accept the invitation. Presumably this belief would be more widespread in the higher rate of invitation group, where principals would have seen and received many messages about the invitation. If this is the case, then receiving the invitation to participate in a voluntary audit in this group would undermine the sense of autonomy our intervention was designed to foster.

The existence of peer pressure to accept the invitation in regions with a higher density of invitation could also generate a negative effect of treatment in other ways. If peer pressure led some school principals in higher rates of invitation areas to volunteer only reluctantly, these individuals may experience and respond to the audit and its aftermath differently than school principals who freely accepted the audit invitation. Qualitative interviews with school principals (from both control and treatment) conducted after the endline survey suggest that responses to the invitation and voluntary audit were mixed.<sup>52</sup> Some school principals said they welcomed the opportunity to share their experiences with oversight officials, anticipated that these voluntary audits would be different from regular forms of oversight, and expressed satisfaction with the audit. Others expressed the view that the audits might take time away from other tasks and ultimately make little difference on levels of program support received from higher levels of government. If school principals who accepted the invitation due to peer pressure (due to the higher density of invitations), were more likely to experience disappointment and annoyance with the audit, this might explain the negative effect of treatment in high density areas. Relatedly, some school principals might decrease their effort after an audit, knowing that they have just received oversight and are therefore unlikely to experience another round of oversight in the near future. Bobonis et al. (2011) provide evidence of this type of dynamic immediately after elections. If principals who accept because of peer pressure are more likely to reduce their effort after experiencing the audit, this could account for the negative effect of treatment in the higher rate of invitation group.

## VII. Conclusions

Street-level, front-line service providers exercise enormous influence on how public policies are implemented on the ground and thus on the quality of public services citizens enjoy and citizen perceptions of government performance. Traditional models of oversight use the promise of rewards or the threat of punishment in an attempt to deter malfeasance and incentivize better bureaucratic performance. In this project, we suggest that an alternative model of oversight that provides bureaucrats with autonomy and choice might lead to improvements in bureaucratic motivation and effort. We collect data through an intervention carried out in conjunction with the provincial auditing office (PAO) in the province of Chaco, Argentina. Our intervention randomly assigned some school principals to be invited to undergo a voluntary audit of an important school meal program.

This project helps meet scholars' recent calls for more research into the daily work on FLSP and the range of factors and circumstances that affect the performance of those front-line service providers and street-level bureaucrats who are the face of the state to most citizens (Finan, Olken, and Pande 2017; Gans-Morse et al. 2018; Grossman and Slough 2022; Hoag and Hull 2017; Nathan and White 2021; Pepinsky, Pierskalla, and Sacks 2017). To our knowledge, we report results from the first experimental intervention that examines the effects of voluntary oversight from a government audit body on street-level bureaucrats' motivation and effort, as well as on downstream measures of the services provided to citizens. Our intervention finds the expected, positive effects of the voluntary audit for one group of school principals—those in areas assigned to receive a lower rate of invitations to an audit. The fact that we are able to detect effects on self-reported motivation 8 months after the intervention suggest that effects on motivation could be quite long-lasting.<sup>53</sup> Our qualitative work suggests that the opportunity to be recognized for their knowledge and to give feedback to other actors in government that formed part of the voluntary audit process was attractive to many school principals. While some scholars include the opportunity for bureaucratic input into policy formulation and implementation into indices of autonomy

(Rasul and Rogger 2018), our study suggests that the consequences of offering such opportunities merits further study.

At the same time, the unexpected negative effects of treatment on motivation and effort in areas with a higher rate of invitations points to the challenges of scaling up a policy that relies on volunteerism. In a context where regular forms of oversight are extremely hierarchical and offer bureaucrats no choice, a program of oversight that appears to offer autonomy may not be understood as such by bureaucrats who are unused to such opportunities. We would expect this to be especially likely in contexts of frequent communication between those receiving an invitation. We speculate that our results stem from the negative effects of peer pressure effects on bureaucrats' autonomy or the emergence of heterogeneous treatment effects when some bureaucrats volunteer only reluctantly.

More research is needed to examine the possibility and nature of such peer pressure effects. Our study did not include direct questions on school principals' sense of autonomy, but future work could do so. Our findings also suggest the importance of explicitly incorporating subgroups of different density of invitations into experimental encouragement designs, as we have done here, to test if the divergent results across areas that we document here carry over to other settings. Other areas for future research include whether treatment effects may depend on gender (as shown by Barrera-Osorio, et. al (2022) in the case of teachers in Guinea). As our population of school principals is 95% female, we cannot examine that here. More broadly, our results suggest the value of more research that explores the conditions under which oversight is more likely to have the anticipated versus unexpected effects.

Finally, this paper contributes to ongoing discussions about the relationship between social science experiments and the policy process (see, for example, the recent special issue in *World Development*, including contributions by Baird et al. (2020) and Das (2020)). With respect to the relative advantages and drawbacks of working directly with a government partner without significantly supporting their implementation capacity (e.g., Peters, Langbein, and Roberts 2018), our implementation relied very heavily on the existing capacity of our government partner, the PAO. In

particular, the PAO staffed the audits entirely with their existing staff levels, which were the major constraint on the number of schools assigned to treatment in the experiment. While these staffing restrictions reduced the experiment's sample size, the fact that the PAO was able to carry out the voluntary audits supports the external validity of the results. Our cooperation with the PAO also offers some evidence that oversight policy is indeed "manipulable" (Samii 2020): audit bodies designed to carry out mandatory audits can indeed offer novel forms of oversight. The nature of this partnership also offers some reassurance that such an intervention could be offered by oversight bodies elsewhere, thus allowing scholars to explore further the effects of voluntary audit systems in the public sector.

### Works Cited

- Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge. 2017. *When Should You Adjust Standard Errors for Clustering?* National Bureau of Economic Research. Working Paper. <https://www.nber.org/papers/w24003> (June 8, 2022).
- Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee. 2020. "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services." *American Economic Review* 110(5): 1355–94.
- Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack,. 2014. "No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery." *Journal of Public Economics* 120: 1–17.
- Baird, Sarah, Joan Hamory Hicks, and Owen Ozier. 2020. "Randomized Control Trial as Social Observatory: A Case Study." *World Development* 127: 104787.
- Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat. 2020. *The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats*. National Bureau of Economic Research. <https://www.nber.org/papers/w26733> (October 29, 2020).
- Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster. 2008. "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System." *Journal of the European Economic Association* 6(2–3): 487–500.
- Banuri, Sheheryar, and Philip Keefer. 2016. "Pro-Social Motivation, Effort and the Call to Public Service." *European Economic Review* 83: 139–64.
- Barrera-Osorio, Felipe, Jacobus Cilliers, Marie-Hélène Cloutier, and Deon Filmer. 2022. "Heterogenous Teacher Effects of Two Incentive Schemes: Evidence from a Low-Income Country." *Journal of Development Economics* 156: 102820.

- Bear, Laura, and Nayanika Mathur. 2015. "Introduction: Remaking the Public Good." *The Cambridge Journal of Anthropology* 33(1): 18–34.
- Bersch, Katherine, Sérgio Praça, and Matthew M. Taylor. 2016. "State Capacity, Bureaucratic Politicization, and Corruption in the Brazilian State." *Governance* 30(1). <http://onlinelibrary.wiley.com/doi/10.1111/gove.12196/abstract> (August 27, 2016).
- Bertelli, Anthony M. et al. 2020. "An Agenda for the Study of Public Administration in Developing Countries." *Governance* 33(4): 735–48.
- Bobonis, Gustavo J., Luis R. Camara Fuertes, and Rainer Schwabe. 2011. "The Dynamic Effects of Information on Political Corruption: Theory and Evidence from Puerto Rico."
- Boris-Schacter, Sheryl, and Sondra Langer. 2006. *Balanced Leadership: How Effective Principals Manage Their Work*. Teachers College Press.
- Britos, Sergio et al. 2016. *Lineamientos Para Una Política Federal de Alimentación Escolar*. Buenos Aires: CIPPEC.
- Callen, Michael, Saad Gulzar, Syed Ali Hasanain, and Yasir Khan. 2016. *The Political Economy of Public Sector Absence: Experimental Evidence from Pakistan*. National Bureau of Economic Research. Working Paper. <http://www.nber.org/papers/w22340>.
- Cassar, Lea, and Stephan Meier. 2018. "Nonmonetary Incentives and the Implications of Work as a Source of Meaning." *Journal of Economic Perspectives* 32(3): 215–38.
- Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *The Quarterly Journal of Economics* 128(3): 1169–1218.
- Das, Sabyasachi. 2020. "(Don't) Leave Politics out of It: Reflections on Public Policies, Experiments, and Interventions." *World Development* 127: 104792.
- Dasgupta, Aditya, and Devesh Kapur. 2020. "The Political Economy of Bureaucratic Overload: Evidence from Rural Development Officials in India." *American Political Science Review* 114(4): 1316–34.
- Davis, Jennifer. 2004. "Corruption in Public Service Delivery: Experience from South Asia's Water and Sanitation Sector." *World Development* 32(1): 53–71.
- De La O, Ana, Pablo Fernandez Vazques, and Fernando Martel Garcia. 2022. "Federal and State Audits Do Not Increase Compliance with a Grant Program to Improve Municipal Infrastructure: A Pre-Registered Field Experiment." *Yale University Working Paper*.
- Deci, Edward L., and Richard M. Ryan. 1985. "The General Causality Orientations Scale: Self-Determination in Personality." *Journal of Research in Personality* 19(2): 109–34.
- Diaz Langou, Gala et al. 2014. "Los Modelos de Gestión de Los Servicios de Comedores Escolares En Argentina."

- Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-Telling by Third Party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics*: 1–47.
- Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102(4): 1241–78.
- Dustan, Andrew, Stanislao Maldonado, and Juan Manuel Hernandez-Agramonte. 2018. "Motivating Bureaucrats With Non-Monetary Incentives When State Capacity Is Weak: Evidence From Large-Scale Field Experiments in Peru." *SSRN Electronic Journal*.  
<https://www.ssrn.com/abstract=3307140> (July 30, 2019).
- Dye, Kenneth M., and Rick Stapenhurst. 1998. *Pillars of Integrity: The Importance of Supreme Audit Institutions in Curbing Corruption*. Economic Development Institute of the World Bank.  
<http://info.worldbank.org/etools/docs/library/18120/pillars.pdf> (August 6, 2015).
- Ellingsen, Tore, and Magnus Johannesson. 2007. "Paying Respect." *Journal of Economic Perspectives* 21(4): 135–50.
- Esteve, Marc, and Christian Schuster. 2019. *Motivating Public Employees*. 1st ed. Cambridge University Press. <https://www.cambridge.org/core/product/identifier/9781108559720/type/element> (July 19, 2020).
- Ferraz, Claudio, and Frederico Finan. 2011. "Electoral Accountability and Corruption: Evidence from the Audits of Local Governments." *American Economic Review* 101(4): 1274–1311.
- Finan, F., B. A. Olken, and R. Pande. 2017. "The Personnel Economics of the Developing State." *Handbook of Economic Field Experiments* 2: 467–514.
- Gans-Morse, Jordan et al. 2018. "Reducing Bureaucratic Corruption: Interdisciplinary Perspectives on What Works." *World Development* 105: 171–88.
- Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2(3): 205–27.
- Gonzalez-Lira, Andres, and Ahmed Mushfiq Mobarak. 2019. "Slippery Fish: Enforcing Regulation under Subversive Adaptation." <https://www.iza.org/publications/dp/12179/slippery-fish-enforcing-regulation-under-subversive-adaptation> (December 1, 2021).
- Grossman, Guy, and Tara Slough. 2022. "Government Responsiveness in Developing Countries." *Annual Review of Political Science* 25(1). <https://doi.org/10.1146/annurev-polisci-051120-112501> (December 16, 2021).
- Heß, Simon. 2017. "Randomization Inference with Stata: A Guide and Software." *The Stata Journal* 17(3): 630–51.
- Hidalgo, F. Daniel, Júlio Canello, and Renato Lima-de-Oliveira. 2016. "Can Politicians Police Themselves? Natural Experimental Evidence From Brazil's Audit Courts." *Comparative Political Studies*: 0010414015626436.

- Hoag, Colin, and Matthew S. Hull. 2017. *A Review of the Anthropological Literature on the Civil Service*. Rochester, NY: World Bank. World Bank Policy Research Working Paper No. 8081.
- Huebner, E. Scott et al. 2014. "Schooling and Children's Subjective Well-Being." *Handbook of child well-being 2*: 797–819.
- Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken. 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors." *The Quarterly Journal of Economics* 131(1): 219–71.
- Lennox, Clive S, and Jeffrey A Pittman. 2011. "Voluntary Audits versus Mandatory Audits." *The Accounting Review* 86(5): 1655–78.
- Lin, Winston, Donald P. Green, and Alexander Coppock. 2016. *Standard Operating Procedures for Don Green's Lab at Columbia*. Columbia University. [https://alexandercoppock.com/Green-Lab-SOP/Green\\_Lab\\_SOP.pdf](https://alexandercoppock.com/Green-Lab-SOP/Green_Lab_SOP.pdf).
- Lipsky, Michael. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. Russell Sage Foundation.
- McDonnell, Erin. 2020. *Patchwork Leviathan: Pockets of Bureaucratic Effectiveness in Developing States*. Princeton University Press.
- Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119(1): 39–77.
- Nathan, Noah L., and Ariel White. 2021. "Experiments on and with Street-Level Bureaucrats." In *Advances in Experimental Political Science*, eds. James N. Druckman and Donald Green. Cambridge University Press.
- Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(2): 200–249.
- Pepinsky, Thomas B., Jan H. Pierskalla, and Audrey Sacks. 2017. "Bureaucracy and Service Delivery." *Annual Review of Political Science* 20(1): 249–68.
- Perry, James L., Annie Hondeghem, and Lois Recascino Wise. 2010. "Revisiting the Motivational Bases of Public Service: Twenty Years of Research and an Agenda for the Future." *Public Administration Review* 70(5): 681–90.
- Perry, James L., and Lois Recascino Wise. 1990. "The Motivational Bases of Public Service." *Public Administration Review* 50(3): 367–73.
- Peters, Jörg, Jörg Langbein, and Gareth Roberts. 2018. "Generalization in the Tropics – Development Policy, Randomized Controlled Trials, and External Validity." *The World Bank Research Observer* 33(1): 34–64.
- Petter, John et al. 2002. "Dimensions and Patterns in Employee Empowerment: Assessing What Matters to Street-Level Bureaucrats." *Journal of Public Administration Research and Theory* 12(3): 377–400.

- Pfaff, Steven, Charles Crabtree, Holger L. Kern, and John B. Holbein. 2021. "Do Street-Level Bureaucrats Discriminate Based on Religion? A Large-Scale Correspondence Experiment among American Public School Principals." *Public Administration Review* 81(2): 244–59.
- Prottas, Jeffrey Manditch. 1978. "The Power of the Street-Level Bureaucrat in Public Service Bureaucracies." *Urban Affairs Quarterly* 13(3): 285–312.
- Rasul, Imran, and Daniel Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *The Economic Journal* 128(608): 413–46.
- Rauch, James E, and Peter B Evans. 2000. "Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries." *Journal of Public Economics* 75(1): 49–71.
- Ryan, Richard M., and James P. Connell. 1989. "Perceived Locus of Causality and Internalization: Examining Reasons for Acting in Two Domains." *Journal of Personality and Social Psychology* 57(5): 749–61.
- Ryan, Richard M., and Edward L. Deci. 2000. "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions." *Contemporary Educational Psychology* 25(1): 54–67.
- Samii, Cyrus. 2020. "Reasons for Policy Experimentation That Have Nothing to Do with Selection Bias." *World Development* 127: 104825.
- Santiso, Carlos. 2006. "Improving Fiscal Governance and Curbing Corruption: How Relevant Are Autonomous Audit Agencies?" *International Public Management Review* 7(2): 97–108.
- Stiglitz, Joseph E. 1989. "Principal and Agent." In *Allocation, Information and Markets*, The New Palgrave, eds. John Eatwell, Murray Milgate, and Peter Newman. London: Palgrave Macmillan UK, 241–53. [https://doi.org/10.1007/978-1-349-20215-7\\_25](https://doi.org/10.1007/978-1-349-20215-7_25) (July 22, 2021).
- Tendler, Judith. 1997. *Good Governance in the Tropics*. The Johns Hopkins University Press.
- Wang, Erik H. forthcoming. "Frightened Mandarins: The Adverse Effects of Fighting Corruption on Local Bureaucracy." *Comparative Political Studies*.
- World Bank. 2018. *The State of Social Safety Nets 2018*. Washington D.C.: World Bank.
- Zamboni, Yves, and Stephan Litschig. 2018. "Audit Risk and Rent Extraction: Evidence from a Randomized Evaluation in Brazil." *Journal of Development Economics* 134: 133–49.

---

<sup>1</sup> Street-level bureaucrats (Lipsky 1980) are those public-sector employees working at the front lines, providing services to citizens in schools, health centers, and welfare offices. Throughout this paper, we refer interchangeably to "street-level bureaucrats" (SLBs) and "frontline service providers" (FLSPs).

---

<sup>2</sup> The existing literature sometimes characterizes school principals as managers within street-level bureaucracies (Lipsky 1980), sometimes as street-level bureaucrats because they meet the criteria for SLBs that Lipsky himself offers (e.g. Pfaff et al. 2021), and sometimes as both (Boris-Schacter and Langer 2006, 15–16).

<sup>3</sup> In contrast, some other provinces have more robust approaches to oversight with regular ongoing contact with schools (Diaz Langou et al. 2014, 22).

<sup>4</sup> Hypotheses were pre-registered with EGAP after the intervention while endline data collection was ongoing and before the PIs had access to any of the endline data.

<sup>5</sup> Separately, we also estimate the effect of the invitation on those who actually accepted it (the complier average causal effect). See note 36 below for more details.

<sup>6</sup> Using the framework of Esteve and Schuster (2019, figs. 1, 2), most of these interventions assume an extrinsic, self-regarding conceptualization of bureaucratic motivation. A separate literature explores the conditions under which government actors, especially politicians, actually want to monitor bureaucrats (see Grossman and Slough (2022) for a summary). For the purposes of this paper, we treat the PAO as an apolitical body that genuinely seeks to oversee bureaucratic performance; the sources of politician interest in independent auditing bodies is a promising area for further research (see, for example, Hidalgo et al. (2016)).

<sup>7</sup> Within the experimental tradition, a related research agenda on public sector motivation in the Global South has focused on understanding selection into the public sector (e.g., Ashraf et al. 2020; Banuri and Keefer 2016; Dal Bó, Finan, and Rossi 2013) rather than variation in motivation among existing FLSPs.

<sup>8</sup> These are distinct from third party auditors who may be contracted to carry out government-mandated auditing of firms in the private sector to check compliance with government environmental or labor standards (e.g., Duflo et al. 2013).

<sup>9</sup> In existing literature on government audits, their mandatory nature is understood to be so obvious that it is often not even mentioned.

<sup>10</sup> Interventions that inform politicians or bureaucrats that they face an increased likelihood of a mandatory audit (e.g., Olken 2007; Zamboni and Litschig 2018) rely on the same logic as mandatory audits more generally—that the desire to avoid punishment induces compliant behavior.

---

<sup>11</sup> Ryan and Deci (2000, 55) do note that extrinsic motivation is not necessarily an “impoverished” form of motivation and can some types of extrinsic motivation involve a feeling of agency and choice.

<sup>12</sup> Note that a system of voluntary audits inherently offers bureaucrats the opportunity to choose *not* to accept an audit. Choosing not to accept an audit may also increase bureaucrats’ sense of autonomy, although we expect that experience to be amplified in the audit itself, and only actually engaging in the audit should increase feelings of competence, as described in the following paragraphs. In the empirical section, some of our data allows us to discern whether the audit invitation or the process of experiencing a voluntary audit is responsible for our results, with the evidence favoring the importance of the latter.

<sup>13</sup> In a study of street-level bureaucrats in a state-level human services agency, Petter et al. report that “[w]hen participants talked about what they like and disliked in their jobs, one of the first things they noted was wanting to be allowed input into agency or office decision making” (Petter et al. 2002, 398).

<sup>14</sup> See Diaz-Langou et al. (2014) and Britos et al. (2016) for more details on the history of school meal programs and the phenomena of over and under nutrition in Argentina.

<sup>15</sup> Data on Chaco’s budget is available here: <https://economia.chaco.gov.ar/secciones/6> and here: <https://ipecd.chaco.gob.ar/economia/sector-publico/gastos-publicos>.

<sup>16</sup> Principal interviews, June 2020. See the Qualitative information document in the SI file for more details on the qualitative interviews referenced here.

<sup>17</sup> Quantitative data from our endline survey corroborates interview reports of variable motivation and effort; see Table A1 in the SI file for descriptive statistics.

<sup>18</sup> As one principal related, “we give our lives for the institutions and there’s no type of recognition” (Preschool principal, June 2020).

<sup>19</sup> As far as we are aware, none of those “normal” audits or any of those conducted as part of the experiment described here resulted in any citation or action by the Ministry of Education with respect to any of those audited.

---

<sup>20</sup> In interviews with current and former school principals and vice-principals in the province in February-March 2018, many shared that they had never received information on their performance in the implementation of the meal program we study.

<sup>21</sup> For these figures from the endline survey, we report responses from the control group only to avoid the possibility that responses are influenced by assignment to treatment; results for treated schools are very similar.

<sup>22</sup> As one principal expressed this, she wanted to “have those in charge gain an understanding of how we manage with what they give us” (Interview, preschool principal, June 2020).

<sup>23</sup> Within regions deemed accessible by the PAO, we eliminated from the sample schools located within dispersed rural areas (due to similar concerns about accessibility), schools with fewer than 40 students (the mean enrollment of schools in dispersed rural areas), and schools with missing or duplicate information on the school principal in administrative data.

<sup>24</sup> To this point, to our knowledge, no control schools called the PAO seeking a voluntary audit.

<sup>25</sup> See the SI file, Figure B1, for the full text of the letter. We are able to confirm that 58 of the 62 schools in treatment received the invitation to a voluntary audit, and no schools in control received the invitation. We have no information on the four remaining schools. We consider a letter to have been received if we have a postal confirmation, if the school contacted the PAO in response to the letter, or the letter was hand delivered (our local field representative hand delivered letters which were returned to the PAO by the postal service).

<sup>26</sup> We are unaware of any similar intervention that would allow us to compare rates of acceptance of voluntary audits across contexts.

<sup>27</sup> See the SI file for the full text of the document.

<sup>28</sup> Although not its normal practice, the audit team did administer the same short questionnaire to schools in control who received a mandatory audit.

<sup>29</sup> Our original intention was to complete the endline survey in November 2018, but this was delayed due to bureaucratic hurdles.

<sup>30</sup> See Tables A1b and A1c in the SI file.

<sup>31</sup> See the SI file, Figure B3, for a photo of one of these notebooks.

---

<sup>32</sup> Neighborhood characteristics should not be affected by the intervention. We did not conduct a baseline survey of school principals, as we were concerned this might affect responses to treatment and hence the external validity of the study.

<sup>33</sup> Table 1 uses all observations for which we have baseline characteristics. However, in our main analysis we drop observations for which the dependent variable is missing. To check that item non-response does not create imbalance in baseline characteristics, we include in the SI file, Tables A9-A9c, balance tests like Table 1 using the samples in the main results section. In cases where the value of a baseline covariate was missing, we imputed the overall mean as specified in our pre-analysis plan. This procedure also follows Lin, Green, and Coppock (2016).

<sup>34</sup> Appendix B presents the hypotheses and corresponding variables included in the pre-analysis plan. Note that in the PAP, we anticipated possible effects of treatment on the nature and diversity of foods served in the schools, and, as a result, on attendance. An administrative decision made immediately before our intervention (June 2018) changed the system of food provision from a decentralized one, wherein principals were responsible for procurement, to a centralized one, where principals receive food purchased centrally. Although we were aware of that policy change at the time of writing our PAP, we were not aware of the extent to which this would diminish principal discretion over what foods were served in school canteens. Results for attendance and food composition is reported in Table A6. Within the higher rate of invitation group, we see a small negative effect of assignment to treatment on the proportion of meals served that contain fruits or vegetables. As greater provision of fruits and vegetables is a plausible measure of effort, this is consistent with our other reported findings that assignment to treatment decreased effort in the higher rate of invitation group, although this is significant only at the .1 level. We do not find any effect of treatment on either the proportion of meals provided (the vast majority of schools provide all of the meals they are supposed to) or on student attendance.

<sup>35</sup> We exclude from this measure all school closings that are a result of national or regional holidays.

<sup>36</sup> See Tables A7, A7b, and A7c for a comparison of baseline characteristics of schools within the treatment group who accepted versus did not accept the invitation to a voluntary audit. These tables suggest somewhat smaller schools—those with lower numbers of enrollees and lower

---

attendance—might be somewhat more likely to accept the invitation to an audit; these differences are significant at the .1 percent level in the pooled sample and in the higher rate of invitation group (for attendance only). Within the lower rate of invitation group, there is evidence that schools that accepted the invitation were from higher SES groups (difference is significant at the .05 percent level). Apart from these, there are no significant differences between these groups. As a reminder, in the text we present the intent to treat effects, which are not biased by observable or unobservable differences between principals who accept the invitation and those who do not. In the SI file, we also present the Complier Average Causal Effect, which we estimate using assignment to treatment as an instrument for receiving the voluntary audit.

<sup>37</sup> Since assignment to treatment is at the school level, and there is one survey respondent per school, we compute robust standard errors when analyzing the lower and higher rate of invitation groups separately. For this decision, we follow Abadie et al. (2017) who recommend clustering standard errors only when the experimental design involves clustered randomization.

<sup>38</sup> We use the STATA package `ritest` (Heß 2017). Following our original randomization, we specified 500 random permutations to be performed within each region, for the lower and higher rate of invitation groups, separately.

<sup>39</sup> In the SI file, we present the results for the analyses outlined in the PAP. In those analyses, we first estimate an overall effect for the intervention, pooling low and high density areas, and then we include an indicator variable that takes on the value of one for high-density areas and an interaction term that takes on the value of one for schools assigned to treatment within high-density areas. We also include the p-value for the sum of assignment to treatment and the interaction term (assignment to treatment X high density)—thus indicating whether the effect of treatment is statistically significant in high density regions. These results are substantively similar to those in the main text. They are in Tables A2b, A3b, and A4b of the SI file, corresponding to the same numbered tables in the main text. There, we also present results for each table for the same set of regressions with the inclusion of control variables selected via lasso (in Tables A2, A3, and A4).

---

<sup>40</sup> In each scale, 10 items corresponded to extrinsic motivation and 5 items corresponded to intrinsic motivation. Items were modified as needed so that they described the appropriate task, either teacher supervision or meal supervision. The full list of items is provided in Tables B2 and B2b in the SI file. The order of items was assigned at random for each of the two tasks, and the items were presented in the same order for each task for all respondents.

<sup>41</sup> Randomization inference p-values in the last row of Table 2 confirm the results from conventional t-tests. Moreover, with randomization inference the decrease in intrinsic motivation for the high rate if invitation group is more precisely estimated (p-value=0.098).

<sup>42</sup> Given the large time gap between the intervention and the survey, the only differences we will be able to detect will be long-term, persistent changes in hours worked.

<sup>43</sup> The survey was administered in paper, and so principals were not constrained in any way to report hours in the varied tasks that summed to the total number of hours they reported. A few school principals reported very high values across these outcomes. For example, one principal reported 80 hours working in the lunch program and 80 hours in other activities. We recoded to missing observations that are three standard deviations or more away from the mean for this group of outcomes.

<sup>44</sup> The same is true of the pooled sample; results are reported in the SI file.

<sup>45</sup> Teachers belong to a number of different unions in Chaco and not all unions call for strikes on the same days, so some strikes may involve only a subset of a school's instructors.

<sup>46</sup> We collected attendance for one day per week because looking through old attendance records was time consuming, and we were concerned principals would not have patience to look through multiple days for week over the entire period of interest. When selecting dates, we excluded any dates of known closures (provincial or national holidays). In a few cases, we later learned about municipality-specific holidays. School reports of zero attendance during these days were recoded to missing. The 12 subsequent weeks were not immediately after mailing—letters of invitation went out in early July, right before the two-weeks winter holiday. Our “baseline” closing data thus comes from the 4 immediately prior weeks in June. Our “post-intervention” data on school closing begins in the second week of August. We did not collect attendance data for any weeks in July or early August because school reopening dates after the holidays are often changed at the last minute, and we expect school attendance data to be noisier immediately after holidays.

---

<sup>47</sup> For all schools in the sample, we see unanticipated closings about 1.5% of the days for which we collect data.

This rate, averaged over a 180day school year, would mean schools experience unanticipated closings 2-3 days a year (again, excluding the periods immediately after the return from winter and summer holidays, which might experience additional unanticipated closings).

<sup>48</sup> Randomization inference p-values confirm the results in columns 1 and 2 of Table 4.

<sup>49</sup> The results in columns 3 and 4 should be interpreted with caution; although the conventional p-values indicate significant differences, the randomization inference p-values do not. The dependent variables in these columns are not normally distributed. Instead, they are skewed because there are very few school closings in the two-weeks period after the audit, and so, t-tests and RI p-values lead to different conclusions.

<sup>50</sup> We thank an anonymous reviewer for encouraging us to use various empirical approaches to test for spillovers.

<sup>51</sup> Table A10b also presents a series of analyses using clustered standard errors at the region level, which provides evidence that schools closings may be different between control schools in areas with higher versus lower rates of invitation. However, these differences are not robust to the alternative specification with first differences as the dependent variable, as we explain in the text.

<sup>52</sup> We asked principals in treatment about their experiences with the invitation and (if applicable) audit and all principals about how they and their peers might view such an invitation.

<sup>53</sup> Future work might usefully collect data at varied intervals to allow for more fine-grained estimates of the endurance of any effects of treatment.