Journal of Development Economics 160 (2023) 102995

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

Motivating bureaucrats with behavioral insights when state capacity is weak: Evidence from large-scale field experiments in Peru☆

Andrew Dustan [a], Juan Manuel Hernandez-Agramonte [b], Stanislao Maldonado [c,*]

[a] Department of Economics, Vanderbilt University, United States of America
[b] Innovations for Poverty Action, United States of America
[c] Faculty of Economics, Universidad del Rosario, Colombia

ARTICLE INFO

JEL classification:
C93
D73
O15

Keywords:
State capacity
Behavioral insights
Civil servants

ABSTRACT

We study how text messages incorporating behavioral insights can be used as a tool to affect civil servant performance when state capacity is weak. By experimentally varying the content of a messaging campaign targeted to civil servants implementing a school maintenance program in Peru, we test the effectiveness of reminders and treatments making salient either monitoring, social norms, the possibility of public disclosure of noncompliance, or audit risk. All messaging treatments improve compliance by similar magnitudes, increasing the probability of submitting a key expense report by an average of 3.9 percentage points over a base of 74%. The inability of this large-scale experiment to detect differential impacts by treatment arm is consistent with timely reminders being the main driver of increased compliance. We explore generalizability across time and populations in two supplemental experiments, confirming the promise of such campaigns to improve civil servant performance when the state lacks enforcement capacity.

1. Introduction

State capacity depends on a well-functioning bureaucracy in which competent civil servants are incentivized to perform their duties (Bertrand et al., 2020; Best et al., 2017). While improving selection into civil service has received significant attention as a mechanism for improving capacity (Dal Bo et al., 2013; Ashraf et al., 2015; Leaver et al., 2021), much remains to be learned about inducing existing bureaucrats to perform. Evidence is particularly sparse regarding the impact of non-monetary incentives, which are often necessary because of the fiscal costs and political economy challenges inherent to implementing monetary incentives at scale.[1] Non-monetary incentives such as social recognition (Ashraf et al., 2014) and location postings (Khan et al., 2019) have shown the potential to affect bureaucrat behavior in

☆ We thank Ernesto Dal Bo, Kitt Carpenter, Claudio Ferraz, Fred Finan, Alvaro Forteza, Federico Gutierrez, Daniel Hicks, Rachid Laajaj, Mounu Prem, Santiago Saavedra, Juan Vargas, Andrew Zeitlin, and seminar participants at Universidad del Pacífico, Universidad del Rosario, Vanderbilt University and participants at the Advances with Field Experiments, Association for Public Policy Analysis and Management, Midwest International Economic Development Conference, North East Universities Development Consortium Conference, Pacific Conference for Development Economics, Peruvian Economic Association, Southern Economic Association, and RIDGE/LACEA-PEG conferences for insightful comments. Field implementation would not have been possible without the support of the MineduLAB and PRONIED teams at the Ministry of Education and the CUNA MAS team at the Ministry of Social Inclusion. We especially thank Jorge Mesinas, Fabiola Caceres, Luis Baiocchi, Dennis Velasco, Rosie Fontinier, Milagros Barrantes, Jorge Fernandez, Karina Tecse, Marcos Bravo, Liz Espinoza, Susan Fiestas, and Martin Blas. Juliana Aragon, Maria Paula Medina, and Cesar Huaroto provided excellent research assistance. Maldonado acknowledges financial support from Colombia Cientifica – Alianza EFI 60185, contract FP44842-220-2018, funded by The World Bank through the Scientific Ecosystems, managed by the Colombian Ministry of Science, Technology, and Innovation (MINCIENCIAS). IRB approval was secured with Vanderbilt University (Study 151334). The experiments were registered with the AEA RCT registry under ID AEARCTR-0003722.

* Corresponding author.
E-mail addresses: andrew.dustan@vanderbilt.edu (A. Dustan), jmhernandez@poverty-action.org (J.M. Hernandez-Agramonte), stanislao.maldonado@urosario.edu.co (S. Maldonado).

[1] Finan et al. (2017) provides an overview of the literature on monetary incentives in developing countries. Evidence for the case of developed countries is extensive and results are mixed. See, for instance, Burgess et al. (2017) for the UK case.

[2] The use of behavioral insights by governments to affect economic and social outcomes is still limited but increasing, especially in developed countries. Examples include the Behavioral Insights Team in United Kingdom and the Social and Behavioral Sciences Team at the White House in United States. Developing countries have begun to consider these issues in their policy-making, but little has been done in addressing civil servants' potential behavioral biases. A recent report by the Behavioral Insights Team (Hallsworth et al., 2018) provides one of the first systematic attempts to address these issues.

https://doi.org/10.1016/j.jdeveco.2022.102995
Received 19 July 2021; Received in revised form 29 August 2022; Accepted 17 October 2022
Available online 23 November 2022
0304-3878/© 2022 Elsevier B.V. All rights reserved.

a flexible, low-cost way, yet few other policies have been rigorously evaluated. In particular, there has been limited use of insights from behavioral economics to design interventions that induce civil servant effort and compliance with public policies.[2]

This paper analyzes the impacts of a randomized, large-scale intervention that leveraged behavioral insights to increase policy compliance. We partnered with Peru's Ministry of Education (MINEDU) to design and execute a text messaging (SMS) campaign, aimed at reducing noncompliance among bureaucrats – typically school principals – in implementing a national school maintenance program. Bureaucrats in this program file a work plan, receive a monetary transfer, execute the plan, and then document their expenditures. Before the SMS campaign, noncompliance was high: 9% filed expense reports late or in an incomplete manner, while 11% did not file at all.[3] Furthermore, 15% of civil servants did not withdraw the designated monetary transfers, representing 10% of the total assigned budget.[4] The SMS campaign randomly assigned bureaucrats to receive a series of messages from MINEDU that targeted one of five specific behavioral mechanisms, discussed in more detail below, or to a pure control group. Treatment status was linked to administrative databases that tracked compliance with program milestones, with a particular focus on withdrawing transfers and filing the culminating expense report.

We find that these messages increase compliance on the targeted outcomes. Assignment to any treatment arm causes a 3.9 percentage point (p.p.) increase in the probability of filing the expense report compared to 74% in the control group, a 15% closing in the compliance gap (the distance between the current levels of compliance and full compliance). The messages also increase by 1.5 p.p. the probability of (nearly) complete withdrawal of the transferred amount compared to 89%, closing the compliance gap by 13%. While the absolute size of these impacts is modest, given the low cost and relative simplicity of implementing the intervention at scale, these results indicate that SMS campaigns are one promising tool for increasing compliance. The magnitudes of these effects are similar to other behaviorally-motivated interventions in Peru that did not specifically target civil servants or their particular activities: letters highlighting tax compliance rates decreased the property tax compliance gap by 20% in Del Carpio (2013), while reminder text messages decreased the rate at which savers failed to meet their savings commitments by about 7% (Karlan et al., 2016).

The scale of the intervention and use of administrative data for measurement result in precise impact estimates, allowing us to test for differential impacts from the treatment arms targeting distinct behavioral mechanisms, introduced briefly here. The most basic treatment arm was simply a *reminder* of necessary program activities and well-known deadlines, aimed at helping to overcome limited attention problems by increasing the salience of program requirements. The other treatment arms also increase salience, while attempting to achieve additional impact by targeting additional behavioral barriers and mechanisms. The *salience of monitoring* treatment told civil servants their level of compliance, increasing the salience of administrative oversight that was already known to be occurring. The dynamic *social norm* treatment emphasized that most bureaucrats were already complying, appealing to people's desire to conform to common behavior among peers while also likely serving as a reminder that noncompliance was visible to supervisors. The *soft shaming* treatment reminded bureaucrats that their names would appear in a public list in the case of noncompliance, while the *salience of auditing* treatment reminded recipients that they would be visited to supervise their activities, increasing the salience of consequences based in the loss of social status and formal discipline, respectively. None of these treatments modify the information civil servants have access to or the structure of the program's monitoring system. At the beginning of the intervention, civil servants know they are being monitored on a regular basis, that they will eventually be visited, and that their names will be published in the case of noncompliance. Responses to these treatments may thus be interpreted as a consequence of the targeted behavioral dimensions.

Interestingly, the pattern of results suggests that the treatment arms were similarly effective in inducing compliance with the targeted outcomes. We cannot reject that all treatments have equal impacts on expense report filing, with point estimates ranging from 3.0 p.p. (reminder treatment) to 4.9 p.p. (salience of monitoring treatment). We are also unable to reject equal impacts on a variety of metrics for transfer withdrawal. For example, equality of impacts cannot be rejected for the outcome of withdrawing 99% of transferred funds, and these impacts range from 1.0 p.p. (social norm and salience of auditing treatments) to 2.1 p.p. (salience of monitoring treatment). There is, however, suggestive evidence that treatments had distinct impacts on the *quality* of expense reports, a dimension which was not targeted by the intervention. The social norms treatment increased the probability of having an expense report approved by the central government by 3.3 p.p., a point estimate greater than any other arm. The overall results are consistent with a conservative interpretation that all treatments acted as reminders, helping civil servants to overcome limited attention problems by making salient the need to comply with program policies, while leaving open the possibility that social norms may have induced differential impacts on implementation quality.

We enrich these primary findings with results from two additional SMS campaign experiments. A follow-up experiment implemented one year later, in 2016, explores the issue of learning and whether this SMS campaign can induce compliance over time. One concern is that civil servants become desensitized to such campaigns, making them ineffective beyond the first iteration. We find no evidence that treatment in 2015 reduces 2016 compliance or responsiveness to the 2016 treatment campaign, suggesting a lack of learning effects and lending additional support to the interpretation that the messages serve as timely reminders. The 2016 data also contain detailed information on civil servants' planned and reported expenditures, allowing us to assess whether the campaign affected the alignment between planned and executed activities.[5]

Finally, we report on a randomized SMS campaign targeting a different national program to learn whether the basic results can be replicated in a population of civil servants with different characteristics. In contrast to the MINEDU-employed bureaucrats characterized by rigid contracts and promotion practices, this experiment targeted field monitors employed by the Ministry of Social Inclusion (MIDIS). These employees have short-term contracts linked to less generous benefits packages than in the education sector.[6] Many field monitors fail to file mandated field reports, affecting planning for future service delivery. We test the role of social norms and the salience of monitoring, two promising interventions identified in the main experiment, in the full population of field monitors. The social norms treatment is ineffective,

[3] As a point of reference, Chaudhury et al. (2006) provide evidence on the poor performance of front-line public good providers. Absenteeism of teachers (19%) and health providers (35%) in developing countries are both high and in line with the rate of noncompliance observed in our context.

[4] Noncompliance in this setting hinders the government's ability to implement its policy preferences and service delivery. It also affects efficiency, since following up with noncompliers is costly. Lack of compliance is directly related to citizen-level outcomes — school infrastructure has a significant impact on educational progress, learning, and long-run outcomes (Akresh et al., 2018; Barret et al., 2019; Duflo, 2001; Hanushek, 1995).

[5] A subset of schools in the 2016 experiment are linked to a rolling school census that includes information on physical infrastructure, but results regarding impacts on infrastructure quality are imprecise.

[6] The existence of a sizable fraction of temporary workers in the public sector is a characteristic that has persisted in the developed and developing world, despite efforts to minimize patronage and political control (see Grindle (2012)). Meyer-Sahling et al. (2018) estimates that the fraction of temporary workers is about 23% in a group of African, Asian, Eastern European, and Latin American countries.

while the monitoring treatment improves compliance significantly. One explanation for this divergence from the main experiment is that field monitors may not expect to keep their posts far into the future, so they give little weight to their peers' perceptions in their compliance decisions. The salience of monitoring treatment is still effective in this setting because officials do care about the information that upstream bureaucrats have regarding their performance.

This paper relates to an emerging literature about the personnel economics of the state (Finan et al., 2017) and bureaucracies in developing countries (Pepinsky et al., 2017).[7] In particular, we contribute to a growing literature regarding strategies to increase civil servant performance. Existing studies have emphasized the role of monetary incentives in the case of enforcement agents (like tax collectors as in Khan et al. (2016)) and front-line service providers like teachers and health professionals, especially regarding incentives based on outcomes and inputs, with mixed results (Hasnain et al., 2014; Finan et al., 2017; Gilligan et al., 2022).[8] More specifically, we explore the role of non-financial incentives for civil servants. Banerjee et al. (2021) consider the role of transfers for police officers in India, Karachiwalla and Park (2017) study teacher promotion in China, and Khan et al. (2019) study postings for property tax inspectors in Pakistan, with positive results. More in line with the approach followed in this paper, others have found promise in non-financial rewards such as in-kind prizes (Glewwe et al., 2010) and social recognition (Ashraf et al., 2014).

A recent literature explores the role of behavioral biases in civil servants, systematically documenting the existence of present bias (Andreoni et al., 2016), status quo bias (Celhay et al., 2019), confirmation bias, framing effects, inattention bias and optimism bias (Hallsworth et al., 2018; Banuri et al., 2017) among bureaucrats. Rather than documenting the existence of such biases, this paper tests a set of strategies to deal with them. This paper is also related to a literature regarding the use of digital technology to improve policy outcomes. Researchers have used smartphones and other technological devices to monitor bureaucrats' attendance (Callen et al., 2018; Cilliers et al., 2018; Dhaliwal and Hanna, 2017; Duflo et al., 2012), increase accountability (Aker et al., 2017), improve public service delivery (Dal Bo et al., 2018) and minimize corruption (Muralidharan et al., 2016; Lewis-Faupel et al., 2016). This paper contributes an innovative way to use text messaging to induce compliance among civil servants.

Finally, this paper relates to a recent but growing literature about experimentation at scale (Muralidharan and Niehaus, 2017; Davis et al., 2017). Large-scale experiments are important for many reasons, including the search for external validity.[9] Among the strategies used to generalize the results from small-scale experiments to a relevant population are theoretical approaches (Banerjee et al., 2017b) and statistical designs based on reweighting (Andrews and Oster, 2017) or bounds (Kowalski, 2018). This paper's design addresses several concerns regarding external validity. By implementing the interventions with a population of civil servants, concerns about scalability and generalizability of results from small samples are not present. The subsequent follow-up intervention addresses the concern that its impacts are not externally valid due to time-specific aggregate shocks (Rosenzweig and Udry, 2020). Furthermore, by implementing a similar SMS campaign with a different population of civil servants, we address the issue of how generalizable the results are for alternative settings and populations.

The rest of the paper is organized as follows. Section 2 describes the institutional setting. Section 3 introduces the intervention and research design. Section 4 describes the results of the main experiment. Section 5 presents the results of the subsequent two experiments. Section 6 concludes.

2. Institutional setting

The field experiment was run in collaboration with the National Educational Infrastructure Program (PRONIED) at the Ministry of Education (MINEDU) in Peru. The goal of this program is to expand, improve, replace, rehabilitate, and construct public educational infrastructure, including both buildings and furniture. One of its critical functions is the School Infrastructure and Furniture Maintenance Program, which entails the allocation of monetary transfers to a civil servant in charge of regular maintenance activities in each school. The progression of the intervention is summarized in Fig. 1.

The program has a strong participatory component. At the beginning of each school year, a Maintenance Committee is created with the participation of teachers, students and parents. This committee is the unit at the school level in charge of implementing all activities related to the maintenance of basic infrastructure and furnishings. In addition, an Oversight Committee is also formed along the same lines. A coordinator of the Maintenance Committee is selected, which is usually the school principal but occasionally a teacher. This coordinator is required to submit a signed legal document called a commitment act to the Educational Local Management Unit (UGEL, similar to a school district in the US case), declaring that they fully understand the program's policies and procedures and that they promise to execute and declare all expenditures according to stated deadlines.

Once this institutional structure is in place, a technical form is produced with the maintenance investments that are considered relevant for the Maintenance Committee. The technical form establishes not only the expenditure items but also the estimated budget. This form is submitted to the UGEL for approval. An UGEL is similar to a school district in the US case. Once the technical form is approved, maintenance civil servants are able to start carrying out maintenance activities and withdraw maintenance funds from accounts at the National Bank assigned to them exclusively for investments in infrastructure. Typical maintenance activities include refurbishing classrooms, such as repairing floors and walls or replacing windows; repairing and improving bathrooms; upgrading electrical installations; and repairing or replacing classroom furniture.

Once the execution phase is completed, the coordinator is required to write an expense report accompanied by invoices for all expenditures, submitted with a report prepared by the Oversight Committee, which evaluates whether they believe that funds were used for their intended purpose. Unused funds are returned. The final step is the approval of the expense report, an activity performed by an infrastructure specialist at the corresponding UGEL.

The maintenance and reporting activities carried out by the coordinator are in addition to the usual workload for which they are responsible. For school principals, these duties include all aspects of school administration, including day-to-day management of teachers and school activities and compliance with MINEDU norms. That is, maintenance tasks are undertaken alongside many other obligations, creating the potential for inattention to deadlines for implementation and reporting.

It is important to emphasize that the maintenance funds are assigned to a single maintenance civil servant in each school. While

[7] Recent reviews in the political economy literature include Dal Bo and Finan (2016) and Azulai et al. (2014).

[8] Barlevy and Neal (2012) present optimal financial incentives for educators, highlighting context-specific factors such as the possibility that standardized tests used to assess performance may change over time.

[9] Researchers have explored several dimensions of external validity, in particular the scalability of interventions, the existence of market equilibrium effects and externalities, site selection and piloting bias, the effect of treatments on different populations, the effect of treatment in the same population under different circumstances, and the effect of different, but related, technologies (Allcott, 2015; Banerjee et al., 2017a; Al-Ubaydli et al., 2017; Banerjee et al., 2017b). See Peters et al. (2018) for an overview of these issues in the development literature.

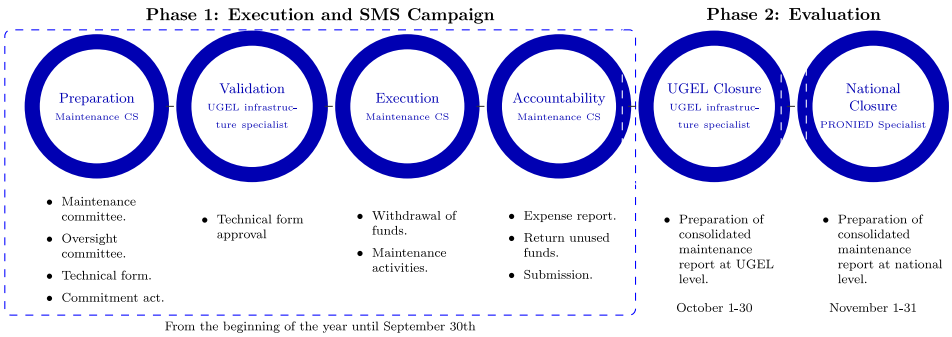


Fig. 1. Intervention cycle for maintenance program (2015). Note: Authors' construction based on administrative reports. Each circle represents a step during the intervention cycle. Maintenance CS stands for maintenance civil servant. UGEL infrastructure specialist is the official in charge of overseeing compliance with maintenance activities at the school district level. PRONIED specialist is the official in charge of overseeing compliance with maintenance activities at the national level. Specific activities during a particular step are described in bullet points.

Maintenance Committees are composed of teachers, students, and parents, only the Committee coordinator is formally assigned the role of receiving the maintenance funds via an account at the National Bank. For this reason, the number of maintenance civil servants is the same as the number of schools that were part of these experiments.

3. Intervention and research design

We partnered with MineduLAB, an innovation lab that exists inside of MINEDU, to design a cost-effective strategy to address civil servants' noncompliance with PRONIED's rules.[10] We implement this strategy in a large-scale field experiment with the population of civil servants in charge of the school maintenance program for which a cellphone number is recorded in MINEDU administrative records. Based on conversations with the MineduLAB and PRONIED teams, we focused on two critical variables: withdrawal of maintenance transfers and reporting of expenses. As discussed above, these two variables are the most relevant in the context of this intervention. In particular, these two outcomes are those for which the maintenance civil servant is most directly responsible (as opposed to oversight committees or UGEL officials) and correspond most closely to observable actions toward carrying out the required maintenance (in the case of withdrawals) and accounting for expenses incurred in performing the maintenance (in the case of submitting the expense report). We will consider additional outcome variables that are also available in the administrative data, but these two will be of particular interest.

3.1. Treatment and conceptual framework

Here, we describe the 2015 SMS campaign and connect it to the behavioral mechanisms being targeted. Each SMS contains a fixed and a variable component. The fixed component includes the bureaucrat's first name and the deadline for task compliance. These fixed elements are rooted in behavioral insights. The use of personalized messages has been shown to be an effective strategy (Karlan et al., 2012). On the other hand, the use of exogenous deadlines has been proven to be useful when agents suffer from procrastination (Ariely and Wertenbroch, 2002). The variable component, described below, is the main behavioral lever that we use to induce a change in bureaucrats' behavior.

Maintenance bureaucrats are assigned to one of six groups. Bureaucrats in the control group receive no SMS. The remaining bureaucrats receive an SMS with behavioral content at fixed points during the intervention cycle. In total, each bureaucrat in any of the treatment groups receives up to five SMS. These SMS share the same behavioral content over the cycle but vary in terms of the type of maintenance activity that is emphasized. For instance, near the beginning of the intervention cycle, bureaucrats receive SMS emphasizing the withdrawal of maintenance funds, whereas near the end of the cycle, SMS emphasize the filing of expense reports. Civil servants only receive a particular message if they have not complied with the activity being emphasized in that SMS. The full set of messages is presented in Online Appendix Table S1.

The variable components can be related to DellaVigna (2009), who classifies behavioral science into three main subfields according to the source of behavioral biases: (1) non-standard preferences, (2) non-standard beliefs, and (3) a non-standard decision-making. Civil servants may behave suboptimally due to deviations from rationality in the way they form their preferences, including time-inconsistent, reference-dependent, and other-regarding preferences. Civil servants could also make mistakes when forming beliefs regarding their own and others' behaviors.[11] Finally, they may fail to optimize their performance at work or consider factors besides optimality in their choices at work. Most of the interventions employed here address failures in optimal decision-making, particularly by helping bureaucrats to deal with information they already possess.

Bureaucrats in the *reminder/warning treatment* receive SMS with an alert and the URL of the PRONIED website where the bureaucrat can obtain more information. Reminders are one of the most popular tools used in behavioral science to influence behavior and the inclusion of an alert is motivated by the need to prime a sense of urgency to comply with maintenance activities. Reminders increase the salience of the highlighted task and are motivated by the existence of limited attention problems and are tools that can potentially change the intertemporal allocation of mental resources to enforce compliance. Karlan et al. (2016) propose a theoretical model to justify the use of reminders in the context of saving based on the idea that individuals misunderstand the value of future consumption and then under-save or under-borrow. From an empirical point of view, reminders have been proven to be useful in inducing donations (Sonntag and Zizzo, 2015), take-up of social benefits (Bhargava and Manoli, 2015), gym attendance (Calzolari and Nardotto, 2017), electricity consumption (Allcott and Rogers, 2014), savings (Karlan et al., 2016), and adherence to medical treatments (Vervloet et al., 2012).

Bureaucrats in the *salience of monitoring treatment* receive SMS with information regarding the amount of transfers not yet withdrawn from the bank or not yet declared on the expense report, depending on the timing of the message in the intervention cycle. In addition to serving as a reminder (as with all other treatment arms), this information makes

[10] MineduLAB is an innovation lab for education policy in Peru, created by Innovations for Poverty Action (IPA), the Abdul Latif Jameel Poverty Action Lab (J-PAL) and the Ministry of Education of Peru. For more details, see https://www.poverty-action.org/minedulab.

[11] Alternatively, they may form beliefs rationally but lack information.

salient to civil servants that their actions are being monitored. This treatment should not be surprising for a fully rational agent since it is perfect knowledge for all the civil servants that the government is able to observe funds withdrawal and expense reporting. Therefore, by making salient a fact that is common knowledge among civil servants, it is possible to re-create some critical dimensions of monitoring systems in a cost-effective way.[12] However, it is important to emphasize that we are not experimentally manipulating who is monitored by the government because all the civil servants in our experiment are being monitored on a regular basis. Rather, we are making this fact salient to a subset of civil servants without changing the information they have about it at the beginning of the experiment.

Bureaucrats in the *social norm treatment* receive SMS with a message emphasizing that most bureaucrats are complying in their reference group (UGEL). Social norms are understood in this paper as a set of informal rules and unwritten codes that establish what we expect of others and what others expect from us Young (2015).[13] Through this lens, the treatment makes salient that complying with program requirements is the norm, inducing compliance by evoking a desire to conform to that norm. The underlying mechanism driving this conformity may be an overt preference for conformity, or it may reflect a desire not to be seen as performing poorly compared to peers. Following Cialdini and Goldstein (2004), it is possible to establish a useful distinction between norms that inform us about what is typically done (descriptive norms) and norms that inform us about what is typically approved or disapproved (injunctive norms).[14] We used a trending qualitative norm, a variation of dynamic social norms (Sparkman, 2021), chosen with the goal of minimizing the risk of backlash effects.[15] Using dynamic social norms, instead of the standard descriptive norms commonly used in the literature, exploits the human tendency to pre-conform with behaviors that are believed to become more common in the near future (Mortensen et al., 2019; Sparkman and Walton, 2017). To the best of our knowledge, this is one of the first papers to use dynamic norms to induce compliance among civil servants.

Bureaucrats in the *soft-shaming treatment* receive SMS reminding them of the possible publication of a list with the names of those bureaucrats who fail to comply with the reporting of expenses. By making this already-known potential consequence more salient, the goal is to induce concern regarding potential reputational loss in order to motivate compliance (Eyal, 2014). This treatment arm is based on a large body of evidence indicating that people are more likely to comply when their behaviors are observed (Rogers et al., 2018). This insight has been used in health interventions to induce behavioral change of unhealthy behaviors such as open defecation (Gertler et al., 2015) and smoking (Voigt, 2013), and a number of other settings as well.[16]

Finally, bureaucrats in the *salience of auditing treatment* receive SMS with a soft threat of auditing. Specifically, they are reminded that they will be visited for supervision of their maintenance activities. Schools are already visited on a regular basis by UGEL representatives for several matters, including the development of maintenance activities. The intervention is simply making salient an event that civil servants will face over the course of the academic year. However, given the scale of the intervention, the probability of facing a visit is low at any given moment of time. We take advantage of this fact to induce compliance among civil servants by reminding them about the fact that they will be visited by UGEL officials. There is a growing body of evidence about the effectiveness of audits. Most papers in this literature have explored the role of audits in making politicians accountable (Ferraz and Finan, 2008) and reducing political corruption (Bobonis et al., 2016). Audits have been also used to induce legal compliance by firms (Duflo et al., 2013) and to induce better public service provision among local politicians (De La O and Martel, 2015). We are not aware of studies using soft forms of auditing with civil servants as used in this paper.

Fig. 2 summarizes the treatments and provides examples of the content of the SMS messages. Online Appendix Table S1 and Figure S1 present, respectively, the detailed contents of all SMS delivered over the intervention cycle in 2015 and the details about critical dates during the SMS campaign. We note that all treatments emphasized the "quantity" dimension of compliance: withdrawing funds and filing the necessary report. They did not incentivize the "quality" dimension, such as maximizing the positive impact of expended funds or producing meticulous expense reports. This raises the question of whether incentivizing quantity came at the cost of quality (Holmstrom and Milgrom, 1991), a possibility that we explore empirically.

3.2. Data

This section discusses the data sources and variables used in the experiment. We combine different administrative records for implementing the research design and to evaluate the impact of the intervention. These data are complemented with surveys that are typically carried out by MINEDU for other purposes. As mentioned above, we consider information for all schools with a maintenance civil servant for which a cellphone number is available in MINEDU's administrative records. Fig. A.1 presents a map with the location of these schools across the country.

3.2.1. Data sources

To implement and evaluate the quality of our randomization strategy, we exploited the School Census that is carried out annually by the Educational Statistics Unit at MINEDU. This census collects information from all public and private schools in the country and includes information on enrollment, students' performance (progress, promotion, repetition, drop-out rates, etc.), teacher and school characteristics. We used the 2014 Census in the design.

Information on outcomes was obtained from two main administrative data sources. To monitor compliance with the maintenance activities, we used PRONIED's school maintenance management system. This system was designed for maintenance civil servants to record and update information related with all maintenance activities, including the uploading of maintenance technical files and expense reports.

[12] It has been shown that monitoring is an effective strategy for improving performance (Callen et al., 2014) and controlling corruption (Olken, 2007), but it is costly and can be captured by corrupt officials (Finan et al., 2017). We are less aware of the use of salience of monitoring to induce compliance among civil servants. There is a large literature about the use of monitoring mechanisms by the government and citizens. See Molina et al. (2017) for a review.

[13] There is a large literature on social norms in economics, sociology, psychology, philosophy, legal studies, political science and anthropology. Given the behavioral approach used in this paper, a psychological approach is emphasized. For an overview of this literature, see Mackie et al. (2015) and the references therein.

[14] Social norms have been proven to be effective in inducing behavioral change in domains such as recycling (e.g. Schultz (1999); but also see Chong et al. (2013)), energy consumption (Alcott and Rogers, 2014), water use (Ferraro and Miranda, 2013), smoking and drinking (Foxcroft et al., 2015; Hansen and Graham, 1991), sexual practices (Lynch et al., 2004), domestic violence (WHO, 2009), female labor supply (Bursztyn et al., 2018), voting (Gerber and Rogers, 2009), charitable giving (Frey and Meier, 2004), and tax compliance (Hallsworth et al., 2017).

[15] A large body of evidence suggests that providing actual levels of conformity with a social norm can induce more people to deviate from it if their baseline expectations regarding conformity with the norm were higher. For a discussion about backlash effects, see Miller and Prentice (2016).

[16] Among other topics, scholars have explored the role of shame on voting (Gerber et al., 2010), environmentally friendly behaviors (Delmas and Lessem, 2014), and charitable giving (Karlan and McConnell, 2014).

Treatment name	Mechanisms targeted	Example content
Reminder/Warning	Salience of deadlines	YRMA: ALERT! Declare maintenance expenses before September 30th. For more details, visit www.pronied.gob.pe.
Salience of Monitoring	Salience of routine monitoring by supervisors	LUCILA: Declare maintenance expenses before September 30th. You have S/.3507 undeclared in the Wasichay system.
Social Norm	Desire to conform with norms, information about norms	BENJAMIN: Declare maintenance expenses before September 30th. The rest of schools in your UGEL are advancing. You are behind.*
Soft-shaming	Salience of effect of noncompliance on local reputation	ADRIAN: Declare maintenance expenses before September 30th. We will publish the names of schools and maintenance civil servants that do not comply.
Salience of Auditing	Salience of future audit by supervisors	KARINA: Declare maintenance expenses before September 30th. We will visit your school to supervise your activities.

Fig. 2. Treatment arm descriptions and example contents. Note: Authors' elaboration. Each message includes the person's name and the deadline to comply with the activity. The rest of the content varies according to the behavioral principle to be emphasized. This example corresponds to the 3rd message delivered during the SMS campaign. All of the messages delivered are described in Online Appendix Table S2.
* Direct translation of the final sentence is difficult. The Spanish phrase is "falta usted", which literally means "you are missing".

To analyze the impact of our intervention on PRONIED funds withdrawal, we had access to balance information for all of the accounts created by PRONIED on behalf of the maintenance civil servants at the National Bank of Peru. These administrative records have the advantages of allowing access to detailed information during different parts of the process and the minimizing attrition problems that are typical of many field experiments.[17]

3.2.2. Main variables

The main outcome variables are constructed based on compliance with the infrastructure maintenance policies. Online Appendix Table S2 contains a full list of all the variable definitions. We construct a set of dummy variables for compliance with each step of the maintenance cycle. The most important variable is a dummy for whether the maintenance civil servant submitted the expense report. This is the one for which maintenance civil servants are accountable. This is a relevant outcome because failure in filling this report introduces administrative and legal costs for the organization due to higher monitoring costs as well as delays in budget execution and planning.

We also consider completion of the oversight report and the approval of the expense report. These two later outcomes are not directly under the control of the maintenance civil servants, but they provide some measure of the quality of their performance since it is expected that these reports are more likely to be approved when the maintenance activities are performed correctly. This is an imperfect measure of quality, however, since imperfect compliance by UGEL monitors in reviewing these reports on time can affect their approval.[18] Because the final step that the maintenance civil servant carries out is the submission of the expense report, we limited the experimental sample to the universe of civil servants who had not already completed this step at the time that the SMS campaign began.

We also create variables for different levels of compliance with the withdrawal of maintenance funds at the National Bank. We consider the proportion of allocated funds withdrawn and dummy variables for the withdrawal of any positive amount, the withdrawal of at least 50%, 95% and 99% of the transferred funds.

We use these administrative records to estimate the impact of the SMS campaign for different periods of time. We present estimates for different points in time during the SMS campaign as well as for time periods after the reporting deadline, when PRONIED officials engaged in costly individualized follow-up with non-compliant bureaucrats to try to induce reporting.

3.2.3. Summary statistics

Table 1 reports summary statistics for the experimental sample. Panel A presents information about pre-treatment outcomes. At week 20 of the PRONIED project cycle (with the end of week 30 representing the end of the cycle and the deadline for filing expense reports), a large proportion of bureaucrats have already submitted the forms necessary to receive their transfers and begin maintenance activities. For example, 86% have gone through the relevant steps with the oversight committee and 67% have filed their commitment act, which signals that they intend to carry out an approved plan of work. By construction, nobody in this sample has submitted an expense report. Before the start of the intervention, National Bank balances were on average PEN 2700 (close to US$ 820).

Panel B presents outcomes at week 30, the official close of the project cycle. As expected, compliance outcomes are better. More than 76% comply with the required submission of the expense report. Somewhat higher compliance levels are found for submission of the commitment act (84%). Outcomes under the control of UGEL officials show relatively low levels of compliance, as expected.

Panel C presents information about bureaucrats' characteristics. Forty-five percent of the maintenance civil servants are male, with an average age of 46 years. Close to 30% of them are appointed civil servants and receive about PEN 7700 (more than US$ 2100) as funds for maintenance activities.

Finally, Panels D and E present school- and district-level characteristics. The average number of classrooms is 5 (with a standard deviation of 6.2), which suggests that schools in Peru are typically small. This is consistent with the average of 128 students. Infrastructure quality is relatively poor: most schools do not have bathrooms connected to a public drainage system, while leaks and water infiltration are common. These schools are located in districts that are mostly rural, although a high proportion of them are connected to electric service. Access to internet and bank branches in the district is also low on average.

3.3. Randomization

Assignment to treatment was randomized at the school level. To implement this design, we exploited school census data and other administrative records to evaluate randomization balance. We proposed

[17] For the 2016 follow-up experiment, we complement these data with the SEMAFORO survey. This is a rolling census that covers all public schools during the academic year to collect information about the provision of educational services. Each month, a random group of schools is visited by MINEDU monitors until all public schools are eventually covered at the end of the academic year. In the Online Appendix, we exploit the random component in the selection of schools to be surveyed over the academic year to compare schools in the control and treatment groups to assess whether treatment affected the quality of infrastructure, one of the components evaluated in the SEMAFORO survey.

[18] Monitors at the UGEL level are in charge of evaluating a large number of expense reports. This introduces delays in completing the evaluation. Therefore, whether a report is approved by the time that we are considering in the analysis is partly due to these delays and not only caused by the inability of the maintenance civil servant to produce reports of sufficient quality.

Table 1
Descriptive statistics.

Variables	Mean	Std. Dev.	N
Panel A: Pre-treatment outcomes			
Submitted to maintenance committee at week 20	0.859	0.348	24,268
Submitted to oversight committee at week 20	0.857	0.350	24,268
Submitted technical form at week 20	0.707	0.455	24,268
Submitted commitment act at week 20	0.674	0.469	24,268
Submitted expense report at week 20	0	0	24,268
Submitted oversight report at week 20	0	0	24,268
With approved expense report at week 20	0	0	24,268
Bank balance at 26/06/2015	2685	5321	20,899
Panel B: Outcomes			
Submitted commitment act at week 30	0.839	0.367	24,268
Submitted expense report at week 30	0.769	0.421	24,268
Submitted oversight report at week 30	0.102	0.302	24,268
With approved expense report at week 30	0.277	0.448	24,268
Proportion withdrawn	0.927	0.249	21,023
Withdrew something	0.997	0.052	21,023
Withdrew 50%	0.929	0.256	21,023
Withdrew 95%	0.905	0.293	21,023
Withdrew 99%	0.898	0.303	21,023
Panel C: Maintenance CS characteristics			
Sex (%Men)	0.455	0.498	24,268
Age	46.17	8.113	24,268
Appointed maintenance	0.276	0.447	24,268
Hired maintenance CS	0.108	0.311	24,268
Allocation transfer	7733	7972	24,268
Panel D: School characteristics			
Classrooms	4.967	6.159	24,268
Students	128.1	592.7	24,268
Bathroom connected to public drainage system	0.379	0.485	24,268
Bathroom connected to septic tank	0.240	0.427	24,268
Bathroom connected to a black well	0.262	0.440	24,268
Bathroom connected to river, ditch or canal	0.0279	0.165	24,268
No bathroom	0.0652	0.247	24,268
Total land area	6746	37,833	24,268
Fully fenced	0.327	0.469	24,268
Partially fenced	0.251	0.434	24,268
Not fenced	0.396	0.489	24,268
Number of educ-admin spaces	8.104	9.420	24,268
Number of buildings	2.211	2.286	24,268
Average leaks in pavilions	1.163	1.507	24,268
Average leaks	1.063	1.413	24,268
Panel E: District characteristics			
Altitude	1639	1557	24,268
Area (%Rural)	0.603	0.489	19,365
Electricity	0.836	0.370	23,650
Public drinking water network	0.669	0.470	23,650
Public drainage network	0.398	0.489	23,646
Internet cafe	0.245	0.430	23,651
Bank branch	0.123	0.328	23,651

Note: Authors' elaboration based on MINEDU's administrative records. The table reports the means, standard deviations, and the sample size. Sample includes all maintenance civil servants who had not submitted their expense report at the beginning of the SMS campaign.

a simple randomized design to PRONIED for two reasons. On one hand, the sample size of our experiment limits the potential gains from more elaborate randomization methods. As discussed in Bruhn and McKenzie (2009), all randomization methods for sample sizes higher than 300 units deliver very similar results. With more than 24,000 schools, we are clearly beyond this threshold.[19]

The second reason is the role of spillovers. We do not expect spillovers to be an issue in this setting. Bureaucrats from MINEDU receive SMS on a regular basis for various matters unrelated to this experiment. We do not expect them to share their SMS with their colleagues in other treatment arms, given that receiving these messages is not notable in itself.[20] Even if we were concerned about this issue, there is little reason to believe that a cluster-randomized design would have successfully mitigated this problem. Cluster designs are useful for spillovers that depend on physical distance, which in this setting may not be the relevant dimension. Given the characteristics of the bureaucracy at MINEDU, with long-term horizons and with regular rotations across the same UGEL or region, there is a high chance of some level of contamination. If such spillovers do exist, they likely bias the estimated treatment effects toward zero.[21]

Table A.1 reports the means and standard deviations for pre-treatment characteristics by treatment status. We consider a large set of variables and find that all treatment groups are balanced on pre-treatment characteristics, with two exceptions. We find evidence of imbalance at the 10% level in terms of the proportion submitting the commitment act by week 20, although the magnitude of this potential imbalance is very small, and in district altitude. We show that results are robust after controlling for exogenous covariates.

3.4. Estimation

To evaluate the effect of the different messages on civil servant compliance, we estimate:

$$y_{smu} = \alpha + \sum_{j=1}^{5} \beta_j treat_s^j + X_{sm}\delta + \varepsilon_{smu} \quad (1)$$

where y_{smu} measures, for school-maintenance civil servant s in municipality m and UGEL u, the outcomes of interest in terms of compliance at different stages of the maintenance cycle as well as the withdrawal of funds from the National Bank accounts. The term $treat_s^j$ denotes the five treatment groups and X_{sm} is a vector of school/civil servant s and municipality m characteristics that may be correlated with compliance. These characteristics include location, number of classrooms, number of students, allocated funds, gender, age, and other school and municipality-level characteristics. Standard errors are clustered at the UGEL level.[22] We account for multiple hypothesis testing by controlling for the false discovery rate within the full set of outcomes (both expense report and bank balances) and presenting the corresponding sharpened q-values (Benjamini et al., 2006).

4. Results

We begin by estimating Eq. (1) for expense report outcomes and maintenance transfer withdrawal. We then provide evidence on the dynamic behavior of the treatment effects during and after the intervention.

[19] In the follow-up and external validity experiments we considered block-randomized designs to stratify on prior treatment status (a dimension of heterogeneity that we test) and to account for a lower number of civil servants in the sample, respectively.

[20] At the time we implemented these experiments, civil servants from MINEDU were regularly exposed to the use of SMS as part of MINEDU's communication policy.

[21] Moreover, Savje et al. (2017) concludes that, for scenarios with limited or even moderate spillovers, standard estimators are "good enough" to recover causal effects as long as the sample size is very large, a condition fulfilled in our experiments. Savje et al. (2017) propose a new parameter robust to the presence of spillover effects and show that standard estimators for the average treatment effect converge to this new parameter when the sample size converges to infinity.

[22] The standard practice consists of clustering standard errors at the treatment level (school in this case). We follow a more conservative approach of clustering the standard errors at a higher level because some of the study outcomes are determined by officials at the UGEL level.

Table 2
Effect of SMS campaign on expense report outcomes.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Expense report	Oversight report	Approved expense report	Approved, cond. on submission	Expense report	Oversight report	Approved expense report	Approved, cond. on submission
SMS	3.855***	0.765*	1.602**	0.286	4.160***	0.732*	1.677***	0.308
	(0.710)	(0.440)	(0.650)	(0.792)	(0.684)	(0.427)	(0.622)	(0.769)
	[0.001]	[0.091]	[0.035]		[0.001]	[0.062]	[0.018]	
Reminder/warning	3.029***	1.165*	1.837*	0.976	3.179***	1.104	1.838*	0.987
	(0.856)	(0.693)	(0.975)	(1.171)	(0.851)	(0.691)	(0.945)	(1.126)
Social norm	4.266***	1.884***	3.326***	2.295**	4.675***	1.857***	3.373***	2.335**
	(1.053)	(0.675)	(0.975)	(1.137)	(1.016)	(0.666)	(0.944)	(1.124)
Salience of monitoring	4.862***	−0.064	0.960	−0.987	5.115***	−0.131	1.014	−0.976
	(0.975)	(0.479)	(0.898)	(1.097)	(0.931)	(0.459)	(0.871)	(1.052)
Soft-shaming	3.601***	0.395	1.005	−0.364	3.771***	0.252	0.979	−0.412
	(0.813)	(0.626)	(0.874)	(1.088)	(0.812)	(0.596)	(0.849)	(1.068)
Salience of auditing	3.479***	0.457	0.886	−0.462	4.016***	0.587	1.180	−0.366
	(1.049)	(0.584)	(0.884)	(1.083)	(0.970)	(0.576)	(0.839)	(1.044)
Control mean	74.15	9.624	26.54	35.79	74.15	9.624	26.54	35.79
p-value, = treatment effects	0.397	0.0830	0.0720	0.0314	0.352	0.0720	0.0799	0.0223
Controls	No	No	No	No	Yes	Yes	Yes	Yes
Observations	24,257	24,257	24,257	18,663	24,257	24,257	24,257	18,663

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 5 to 8 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Online Appendix Table S2 contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. FDR sharpened q-values corresponding to hypothesis tests for treatment effects for all indicated reporting and withdrawal outcomes are in brackets. p-values for equality of treatment effects are from a test that all five treatments have equal effects on the given outcome.

Table 3
Effect of SMS campaign on withdrawal of maintenance funds.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Proportion withdrawn	Withdraw something	Withdraw 50%	Withdraw 95%	Withdraw 99%	Proportion withdrawn	Withdraw something	Withdraw 50%	Withdraw 95%	Withdraw 99%
SMS	0.623	0.043	0.534	1.048**	1.458**	0.746*	0.057	0.664	1.132**	1.527***
	(0.445)	(0.081)	(0.463)	(0.523)	(0.572)	(0.394)	(0.076)	(0.409)	(0.481)	(0.531)
	[0.122]	[0.278]	[0.154]	[0.062]	[0.035]	[0.051]	[0.132]	[0.065]	[0.026]	[0.016]
Reminder/warning	0.517	0.071	0.360	1.030	1.445**	0.582	0.086	0.424	1.096*	1.511**
	(0.565)	(0.099)	(0.595)	(0.660)	(0.715)	(0.521)	(0.097)	(0.547)	(0.631)	(0.682)
Social norm	0.171	0.111	0.160	0.569	1.036	0.465	0.142	0.475	0.773	1.204*
	(0.635)	(0.126)	(0.651)	(0.726)	(0.766)	(0.560)	(0.121)	(0.575)	(0.661)	(0.711)
Salience of monitoring	0.998*	0.046	0.952*	1.624***	2.125***	1.104**	0.057	1.061**	1.683***	2.172***
	(0.530)	(0.135)	(0.563)	(0.607)	(0.660)	(0.489)	(0.133)	(0.524)	(0.576)	(0.627)
Soft-shaming	0.779	−0.058	0.647	1.152*	1.644**	0.833	−0.051	0.706	1.168*	1.648**
	(0.572)	(0.128)	(0.594)	(0.677)	(0.726)	(0.537)	(0.123)	(0.558)	(0.639)	(0.686)
Salience of auditing	0.649	0.046	0.547	0.867	1.044	0.741	0.050	0.648	0.941	1.101
	(0.587)	(0.102)	(0.622)	(0.709)	(0.762)	(0.534)	(0.096)	(0.566)	(0.661)	(0.715)
Control mean	92.24	99.69	92.55	89.78	88.74	92.24	99.69	92.55	89.78	88.74
p-value, = treatment effects	0.694	0.837	0.736	0.664	0.544	0.793	0.736	0.813	0.710	0.549
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012	21,012

Note: Treatment effects and means are reported in percentage points. "SMS" pools all treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 6 to 10 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Online Appendix Table S2 contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. FDR sharpened q-values corresponding to hypothesis tests for treatment effects for all indicated reporting and withdrawal outcomes are in brackets. p-values for equality of treatment effects are from a test that all five treatments have equal effects on the given outcome.

4.1. Expense report filing

Table 2 presents impacts on expense report filing. Column 1 gives the effect of the campaign on the submission of the expense report, omitting municipality, school and bureaucrat covariates. Reported coefficients correspond to percentage point changes (that is, the dummy dependent variables are multiplied by 100). The top row reports the estimated average treatment effect, pooling all arms into a single treatment indicator. Treatment increases the probability of submitting an on-time expense report by 3.86 p.p., compared to a 74% compliance rate in the control group. This result is strongly statistically significant and represents a reduction of 15% in the compliance gap. Estimating separate treatment effects for each message type, we find that all treatments are statistically significant, with coefficients of similar magnitudes. We fail to reject equality of treatment effects across treatments ($p = 0.397$). This is consistent with the messages serving primarily as reminders, or with civil servants responding to similar degrees to the different behavioral principles targeted by each treatment arm.

The intervention incentivized report submission, not quality. To test whether this feature led to changes in average report quality, we first estimate treatment effects on submission of the oversight report and the approved expense report. As discussed above, the approval of these

reports is performed by UGEL officials who are not directly exposed to the intervention. Columns 2 and 3 present the results. We find positive impacts on both outcomes. Treatment increases the probability that UGEL officials have approved an expense report by 1.60 p.p and this impact is statistically significant ($q = 0.035$). This indicates that treated civil servants not only complied with the policy at higher rates, but did so in a way that increases the (unconditional) probability of having a report approved by their UGEL.

We then estimate, in column 4, impacts on report approval *conditional* on submission, to assess whether average report quality is affected. The sample composition for this conditional outcome regression is endogenous with respect to the treatment, so it captures both average quality for those induced to report by the treatment and changes in report quality for "always-reporters".[23] The point estimate is 0.29 p.p. and is not statistically different from zero. We thus conclude that the intervention increased the targeted quantity without appreciable impacts on quality.

While we cannot reject equality of treatment effects on the targeted reporting outcome across treatment arms, we note (with caution) that the social norms treatment was estimated to be the second-most effective in inducing submission (4.27 p.p., second only to the salience of monitoring treatment), while also having the largest estimated impacts on the untargeted reporting outcomes. The reasons that appealing to social norms triggers more compliance but also better report quality are ambiguous, but it may function by making clear that noncompliance is outside of the norm and thus consistent with poor performance. We stress, however, that given the large number of treatment arms and relatively small absolute differences in treatment effects across arms, that this focus on social norms as a particularly effective tool is speculative.

Columns 5 to 8 report the same outcomes as Columns 1 to 4, adding controls to the basic specification. These controls include pre-treatment bureaucrat and school characteristics as previously described in detail in Section 3.2.3. Standard errors decline somewhat, as expected, and point estimates are similar but are slightly larger in most cases.

4.2. Transfer withdrawal

Table 3 presents results on the withdrawal of funds from National Bank accounts assigned to each maintenance civil servant. This allows us to analyze whether the intervention induces maintenance civil servants to withdraw monetary transfers to be invested in school infrastructure maintenance.

Column 1 presents the results for the proportion of allocated funds that were withdrawn, from a specification excluding covariates. We find no effect of the interventions on this proportion. Column 2 presents the results for whether maintenance civil servants withdraw any positive amount of funds. Again, there is no evidence of an effect on withdrawal. This is not unexpected since almost all civil servants withdraw at least something (the control mean is 99.7%). A similar result is found for withdrawing at least 50% of the transferred funds (Column 3). Columns 4 and 5 show a positive effect of the SMS campaign on the probability of withdrawing a high proportion of funds. The campaign caused a 1.05 and 1.46 p.p. increase in the probability of withdrawing at least 95% and 99% of allocated funds, respectively, compared to 90% compliance rates in the control group for both outcomes. These effects represent a reduction in the compliance gap of 10% and 13% with respect to the control group, respectively. As with the expense support submission, we are unable to distinguish statistically between the effects of any of the treatments.

Columns 6 to 10 present results for the specification with controls. Again, point estimates are slightly larger and standard errors smaller.

[23] Because of the endogenous sample composition, we exclude this outcome from the family of "clean" outcomes included in the computation of q-values.

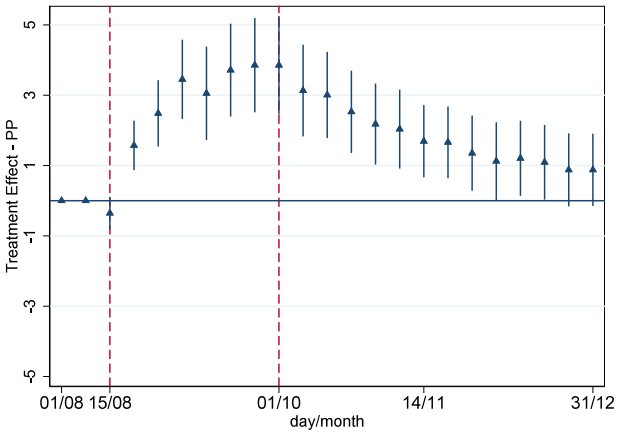


Fig. 3. Treatment effect on expense report submission, by week. Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Eq. (1) with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

Qualitative results are essentially the same. National Bank account data were not recoverable for 13.4% of maintenance civil servants. To verify that this source of attrition is not driving our results, we implement the bounding strategy proposed by Lee (2009). Results are reported in Online Appendix Table S3. Our results are robust to this form of attrition.[24]

4.3. Dynamics of effects

The detailed administrative data allow us to explore the dynamics of the SMS campaign treatment effects. For brevity, we focus on the average effect of receiving a SMS regardless of its behavioral content. We also restrict the analysis to two outcomes: submission of the expense report and withdrawal of at least 99% of funds. The Online Appendix presents the results for the other outcomes.[25]

Fig. 3 shows the dynamics of treatment effects for submission of the expense report. The horizontal axis shows the date for which the treatment effect was estimated. The dashed lines indicate the SMS campaign period from August 15th to October 1st. Pre-treatment data covers the two weeks before the start of the SMS campaign and post-treatment data includes several weeks until December 31st. The vertical axis reports the effect of the SMS campaign in p.p. Before August 15th, we observe no differences between treatment and control groups. We do observe differences after the beginning of the intervention, nearing their peak after three weeks and remaining similar until the deadline. The effect persists even several weeks after the end of the SMS campaign. The 95% confidence intervals do not contain zero until

[24] We had no prior beliefs about particular dimensions of heterogeneous impacts and did not stratify the design accordingly. In an exploratory exercise, we report heterogeneous impacts of the SMS campaign on expense report submission in Online Appendix Table S4 and the withdrawal of 99% of maintenance funds in Online Appendix Table S5. We find no evidence of heterogeneity with respect to school size, the size of the assigned budget, assigned budget per capita, or rurality of the school. There is weak evidence of heterogeneity with respect to gender of the civil servant, suggesting that males experience a 2.5 p.p. higher treatment effect on expense report submission (significant at the 10% level). This finding is robust to including covariates.

[25] Online Appendix Figure S2 presents the results for the oversight report and Online Appendix Figure S3 the results for the approved expense report. Results for the different levels of the withdrawal of funds are reported in Online Appendix Figures S4, S5, and S6.

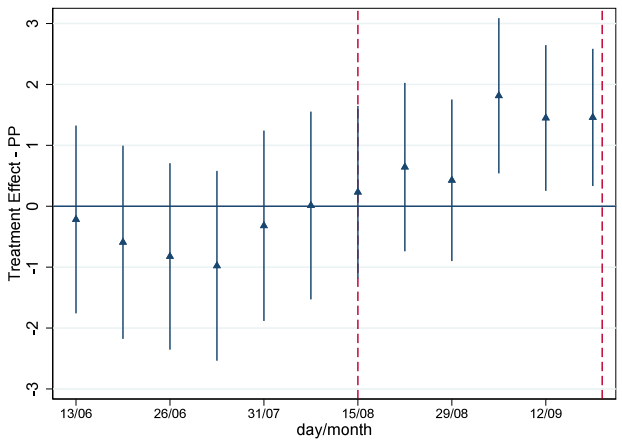


Fig. 4. Treatment effect on withdrawal of 99% of bank balance, by week. Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Eq. (1) with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

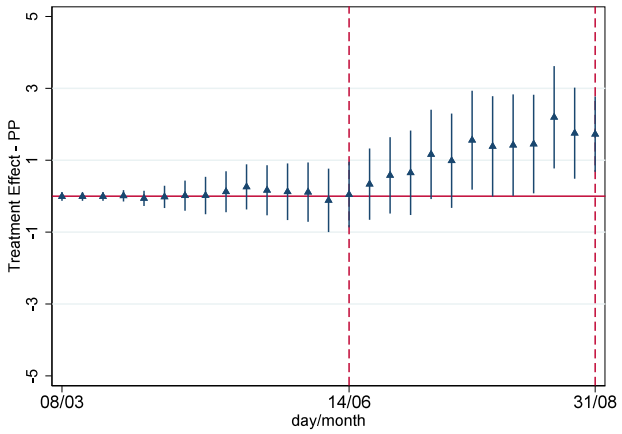


Fig. 5. Treatment effect on expense report submission, by week, follow-up experiment. Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Eq. (1) with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

December.[26] There are two reasons for the declining treatment effect after the deadline. First, some subset of bureaucrats was induced by treatment to comply on time rather than late, as opposed to never complying at all. Second, PRONIED engages in costly follow-ups with non-compliers after the deadline by using a centralized call center and other means. Thus the slowly narrowing gap between treatment and control groups is, at least in part, a reflection of expensive manual efforts that the SMS campaign mitigates. Fig. A.2 in Appendix A presents the results for each treatment arm.

Fig. 4 shows the dynamic effects for the case of withdrawal of funds. We restrict the analysis to the case of withdrawal of 99% of maintenance funds. All effects before the start of the SMS campaign are indistinguishable from zero. We see evidence of positive impacts weeks after the beginning of the campaign. Due to restrictions on obtaining data from the National Bank after the end of the intervention, we are not able to evaluate whether the effect of the SMS campaign persists after the end of the intervention. However, results are in line with the estimates from the previous section. Fig. A.3 presents the results for each treatment arm.

5. Additional experiments

This section briefly presents the results of two additional experiments that explore the persistence and generalizability of the initial intervention: a modified version of the experiment, run in the same population the following year, and a related study in a distinct population of civil servants.

5.1. Follow-up experiment

In 2016, one year after the initial experiment, we implemented and experimentally evaluated two related SMS-based interventions within the same maintenance program. The goals were to further explore the role of the social norm treatment and to evaluate messages that appealed to civil servants' sense of prosociality. While the details and impacts of these later interventions will be explored in future work, here we use this follow-up experiment to explore treatment persistence and impacts on the composition of maintenance expenditures and infrastructure quality.

Given the soft nature of our incentives, we might expect a rational agent to update his or her beliefs to realize that no penalty is enforceable to punish lack of compliance, making subsequent campaigns ineffective. On the other hand, if bureaucrats are forgetful or if SMS serve to reduce problems of limited attention, effects should persist. We test this by comparing performance of civil servants according to their treatment status in the initial experiment to evaluate whether those who were previously exposed to the intervention in 2015 respond to the 2016 intervention.

We used a factorial block design, varying message content and the duration of the campaign. The experiment was stratified on two dimensions: treatment status in the initial experiment (including a stratum for maintenance civil servants who are new to the sample) and the region. Regression specifications include stratum fixed effects. The Online Appendix provides the details of the interventions, sample size, and descriptive statistics for the follow-up experiment. Online Appendix Figure S7 presents the treatment arms, SMS content, and sample size for all treatment variants and Online Appendix Table S6 provides descriptive statistics.

We focus the analysis on the expense reporting outcomes, beginning with the full experimental sample. As shown in Table 4, treatment has a positive and statistically significant impact on submission of the expense report (Column 1). The average SMS effect estimate is 1.71 p.p, compared to 81% compliance in the control group. This represents a reduction of about 9% in the compliance gap. The estimated effects of receiving a SMS are also positive for submission of the oversight report and the approval of the expense report, although effects are smaller and weaker in terms of statistical significance and are insignificant when considering the q-values. Similar to the initial intervention, conditional on submitting an expense report, the estimated impact of treatment on report approval is close to zero and statistically insignificant. Results are largely unaffected after controlling for pre-treatment characteristics (Columns 5 through 8). Fig. 5 replicates the dynamic analysis of effects for expense report submission. We find no difference between treated and control civil servants exists before the beginning of the SMS campaign. Estimated impacts increase after the campaign starts.[27]

[26] Standard errors are clustered at the UGEL level, which accounts for the mechanical within-school correlation in outcomes over time.

[27] Online Appendix Table S7 presents the results for funds withdrawal from the National Bank accounts and Online Appendix Figure S8 shows dynamics. We find, at most, weak evidence of an effect on withdrawals. The 0.9 p.p. effect on withdrawal of 99% of funds (Column 5) has a q-value of 0.13. Results are robust to controlling for pre-treatment characteristics (Columns 6 to 10).

Table 4
Effect of SMS campaign on expense report outcomes, follow-up experiment.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Expense report	Oversight report	Approved expense report	Approved, cond. on submission	Expense report	Oversight report	Approved expense report	Approved, cond. on submission
SMS	1.709***	1.081**	1.194*	0.393	1.702***	1.076**	1.208*	0.442
	(0.524)	(0.495)	(0.635)	(0.714)	(0.525)	(0.497)	(0.626)	(0.711)
	[0.011]	[0.117]	[0.127]		[0.006]	[0.094]	[0.094]	
Control mean	80.62	17.37	39.19	48.59	80.62	17.37	39.19	48.59
Controls	No	No	No	No	Yes	Yes	Yes	Yes
Observations	31,947	31,947	31,947	26,217	31,947	31,947	31,947	26,217

Note: Treatment effects and means are reported in percentage points. All specifications include stratum fixed effects. Columns 4 to 6 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Online Appendix Table S2 contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. FDR sharpened q-values corresponding to hypothesis tests for treatment effects for all indicated reporting and withdrawal outcomes are in brackets.

Table 5
Persistence of treatment effects between main and follow-up experiments.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Expense report	Oversight report	Approved expense report	Approved, cond. on submission	Expense report	Oversight report	Approved expense report	Approved, cond. on submission
Control 2015 and Treated 2016	1.723	−0.234	−0.618	−1.816	1.588	−0.425	−0.543	−1.459
	(1.541)	(1.205)	(1.834)	(2.023)	(1.527)	(1.184)	(1.788)	(1.968)
Treated both years	2.839**	0.762	0.148	−1.508	2.843**	0.631	0.320	−1.070
	(1.430)	(1.113)	(1.708)	(1.888)	(1.412)	(1.101)	(1.684)	(1.868)
Treated 2015 and Control 2016	−0.486	0.195	−0.961	−0.938	−0.403	0.244	−0.517	−0.231
	(1.794)	(1.331)	(1.915)	(2.156)	(1.748)	(1.310)	(1.860)	(2.103)
Control mean	77.96	15.49	37.53	48.14	77.96	15.49	37.53	48.14
p-value for Control/Treat = Treat/Treat	0.0902	0.158	0.417	0.777	0.0506	0.135	0.351	0.715
Controls	No	No	No	No	Yes	Yes	Yes	Yes
Observations	17,533	17,533	17,533	14,026	17,533	17,533	17,533	14,026

Note: Sample includes civil servants who were in both main (2015) and follow-up (2016) experiment samples. Control group was in the control group in 2015 and 2016. Columns 5 to 8 include controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. p-values are from a test that the average treatment effect in the 2016 was equal regardless of treatment status in 2015.

Several factors could explain the smaller treatment effects in this cycle, some of which we can consider as time-varying aggregate shocks as discussed in Rosenzweig and Udry (2020). These aggregate-level shocks imply that some elements of the context may introduce random variation in terms of treatment effect sizes across different cycles. Under this interpretation, the reduction in the effect size does not necessarily represent a decrease in terms of treatment effectiveness. Among other differences, compliance in the control group was 6.5 p.p. higher than in the previous cycle, leaving less room to improve compliance. The maintenance cycle was also shorter, with an earlier deadline (August 31 instead of September 30), although the campaign covered approximately the same span of time as the initial experiment. The population of bureaucrats is substantially larger in the follow-up experiment, in part because MINEDU provided an additional database of cellphone numbers. Finally, although we are interested in the overall effect of receiving a behavioral-motivated text message, the messages in the 2016 cycle are different from those used in 2015, making both campaigns not fully comparable. All these differences suggest that the results need to be interpreted with caution.

We now evaluate whether being exposed to treatment in the initial experiment affects civil servants' response in the follow-up experiment, again focusing on expense report submission. We compare the effects of receiving a SMS conditional on treatment status in the initial experiment. Results are presented in Table 5. We compare three groups to the control group, which was in the control group in both years. In the first row, we estimate the impact of the SMS campaign for those who were initially part of the control group but become part of a treatment group in the follow-up. There is a statistically insignificant positive impact of 1.72 p.p. for the case of submission of the expense report (Column 1). In the second row, we present the treatment effect (again vs. pure control) for those exposed to the intervention in both years, finding a statistically significant positive impact of 2.84 p.p. of the SMS campaign at follow-up. The effect for those treated in both years is marginally higher than for those only treated in the follow-up (p = 0.090), suggesting that previous exposure to the SMS campaign does not hinder the ability of the program to induce compliance. Finally, we consider the case of those initially exposed to treatment but assigned to the control group in the follow-up. If treatment effects persist over time, we should expect for this group to have a higher compliance rate than those in the pure control group. We find no evidence in favor of this hypothesis. Estimates are negative, close to zero, and statistically insignificant.

Columns 2 through 4 present the results for submission of the oversight and approved expense report, as well as approval conditional on submission. All estimated impacts are statistically insignificant. Results are robust to controlling for pre-treatment characteristics in columns 5 through 8. We thus have no evidence that the effects of the SMS campaign persist over time, while it does appear that this tool can be used to influence compliance on a regular basis. This is consistent with messages serving as reminders, helping to overcome limited attention.[28]

[28] We show in Online Appendix Table S8 that, while there was significant attrition between the initial round of the experiment and the follow-up, attrition is uncorrelated with initial treatment status. The 28% attrition between rounds is likely due to a variety of reasons, among them that not all schools scheduled

Table 6
Difference between expense report and technical form, follow-up experiment.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Total	Ceilings	Sanitary facilities	Floors	Walls	Doors	Windows	Electrical installations	Furniture repairs	Paint	Furniture replacement	School supplies
	Panel A. Without controls											
SMS	33.051***	6.675	−7.561	6.504	−8.181	−1.011	3.992	1.098	−0.877	11.861	13.994	5.168*
	(10.024)	(7.753)	(9.248)	(8.258)	(8.267)	(5.535)	(3.880)	(4.038)	(8.559)	(7.681)	(10.010)	(2.942)
	Panel B. With controls											
SMS	32.560***	6.512	−8.237	6.527	−8.305	−0.937	3.932	0.839	−0.777	12.132	14.476	5.187*
	(9.956)	(7.674)	(9.267)	(8.277)	(8.301)	(5.548)	(3.881)	(4.054)	(8.572)	(7.694)	(10.036)	(2.950)
Control mean	−77.37	−29.71	−6.920	−1.947	19.46	−7.057	−17.34	−2.744	−10.09	17.87	−19.80	−12.74
Observations	28,171	28,171	28,171	28,171	28,171	28,171	28,171	28,171	28,171	28,171	28,171	28,171

Note: Dependent variables are difference between the expenditure amount declared in the expense report and the amount dedicated to that area in the technical form. All specifications include stratum fixed effects. Panel B includes controls for personal characteristics (age, gender, type of contract), school characteristics (number of classrooms, number of buildings, land area, number of students, bathroom characteristics, distance to UGEL) and municipality characteristics (altitude, access to electricity, access to drinking water network, access to internet, availability of a bank branch), including dummies for missing observations. Online Appendix Table S2 contains the outcome, treatment and control variables' full definitions. Robust standard errors clustered at the UGEL level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table 7
SMS effects on field monitor compliance, external validity experiment.

Variables	(1)	(2)	(3)	(4)
	Compliance	Compliance	Compliance	Compliance
SMS	2.840	3.016	3.440*	4.446**
	(2.261)	(2.112)	(1.908)	(1.905)
Social norm	0.721	1.167	1.269	2.352
	(2.717)	(2.515)	(2.286)	(2.318)
Salience of monitoring	4.828*	4.741**	5.471***	6.404***
	(2.523)	(2.379)	(2.108)	(2.070)
Control mean	68.27	68.27	68.27	68.00
p-value for social norm = salience of monitoring	0.122	0.149	0.0544	0.0629
Observations	5373	5373	5368	3220
Controls	No	Yes	Yes	Yes
Baseline	No	No	Yes	Yes
October and December	Yes	Yes	Yes	No

Note: Compliance is the percentage of home visits for which the field monitor has submitted an online report. The program target is 100%. "SMS" pools both treatment arms. Rows below are from specifications that estimate separate effects for each treatment arm. Columns 3 and 4 include controls for gender, length of tenure on the job, and fixed effects for the regional office (Territorial Unit) overseeing the civil servant. Standard errors clustered at the civil servant level in parentheses. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. p-values are for equality of treatment effects between social norm and salience of monitoring treatments.

Next, we explore whether the intervention induces civil servants to use maintenance funds to invest in the infrastructure categories prioritized in the technical form they have filed. Recall that, before the intervention, the maintenance civil servant – along with other members of the maintenance committee – define the investment priorities and prepare a technical form with a budget to be approved by the UGEL. We want to test whether the intervention affected the way funds are spent across investment categories. To do this, we compare the planned expenditures (as registered in the technical form) against the executed expenditures (as reported in the expense report) in the follow-up experiment, for which we have the appropriate administrative data. Significant departures from the planned expenditures can be interpreted as a signal that the SMS campaign affected the way civil servants used the maintenance funds in ways that were inconsistent with the wishes of the school community. Although this is not necessary or sufficient evidence of malfeasance or corruption, it could be consistent with them.

Table 6 reports the results of an empirical exercise where the dependent variable is the difference between the executed and planned investment categories, measured in PEN. We consider the most common investment categories, including repair of ceilings, floors, sanitary facilities, walls, doors, and windows. We also consider electrical installations along with repair and replacement of furniture and other school supplies. The average spending gap in the control group is PEN −77 (column 1). Although this amount is small (about $20 USD), this means that they spent less than planned. On the other hand, those in the treatment group spent PEN 33 more than the control group, implying that the intervention is inducing higher expenditures, but not enough to arrive at a positive net amount. Therefore, it does not seem that the intervention is causing a large deviation from planned expenditures.

Results are similar when we look into investment categories. Estimated effects are not only statistically insignificant in most cases, but also small in magnitude. In the case of school supplies, there is a positive effect (PEN 5, significant at the 10% level), but small and not enough to bring average actual expenditures in line with the average planned expenses. Taking into account the fact that the expense report has to be backed up with invoices, receipts, and similar evidence, the possibility that the intervention is inducing patterns of investment consistent with corruption seems unlikely.[29]

to participate in the maintenance program in 2015 were also scheduled to participate in 2016, but it is possible that some schools were eligible and did not initiate any program activities. Attrition among the control group was only 0.7 p.p. higher in the control group than the treatment group ($SE = 0.07$ p.p., statistically insignificant at conventional levels), indicating that these factors were not related to the treatment.

[29] As a final exercise, in the Online Appendix, we explore whether the campaign affects infrastructure quality using a rolling infrastructure census. There is little evidence of impact, but estimates are sufficiently imprecise that we cannot rule out a wide range of positive or negative effects.

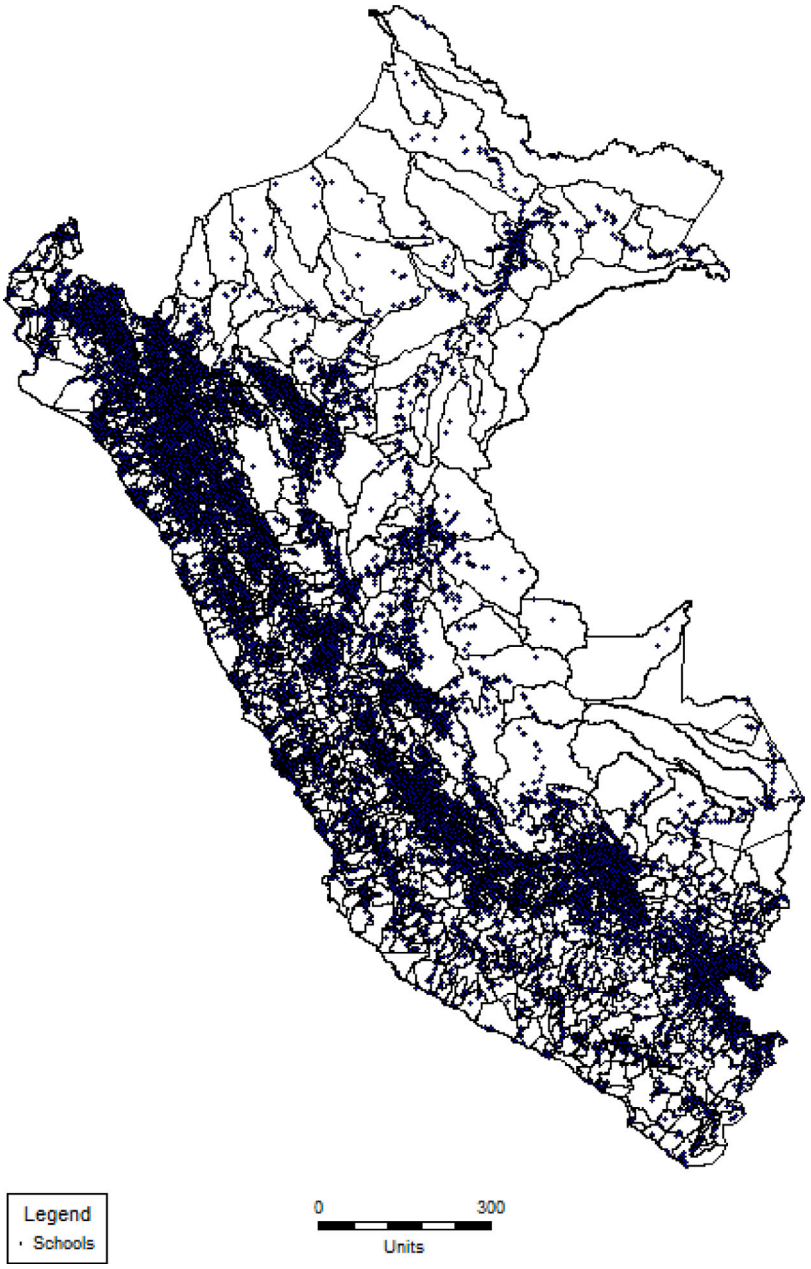


Fig. A.1. Universe of Participating Schools. Note: Authors' elaboration. Each dot represents a school in the experiment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2. External validity experiment

To shed light on the external validity of our findings to other settings, we report on a field experiment in a different population of civil servants. We partnered with the National Program CUNA MAS, an early childhood development program, to implement a SMS campaign to motivate compliance among bureaucrats in charge of a family support service. This campaign was designed to incorporate the lessons learned from the MINEDU experiments. Given the effects found with the social norm and monitoring treatments, we designed a SMS campaign based on these two behavioral contents.[30]

The outcome of interest in this external validity experiment is compliance with reporting of service delivery by field monitors. CUNA MAS requires updated information about the delivery of services (home visits) as well as program beneficiaries' progress on a monthly basis. Field monitors are assigned to work in a geographic area and receive the roster of beneficiaries residing in this area, on whom they must report, directly from centralized CUNA MAS administrators. The reported information is used to plan service delivery for the next period, as well as to update the beneficiary list to incorporate new families. Noncompliance with the submission of this information makes it difficult for this program to respond to its beneficiaries' needs.

The Online Appendix presents the technical details of the implementation of this external validity experiment and a set of analyses to

[30] We consider a qualitative descriptive social norm using the Territorial Unit as reference group. This resembles the social norm treatment used for the initial experiment in 2015. See the Online Appendix for more details.

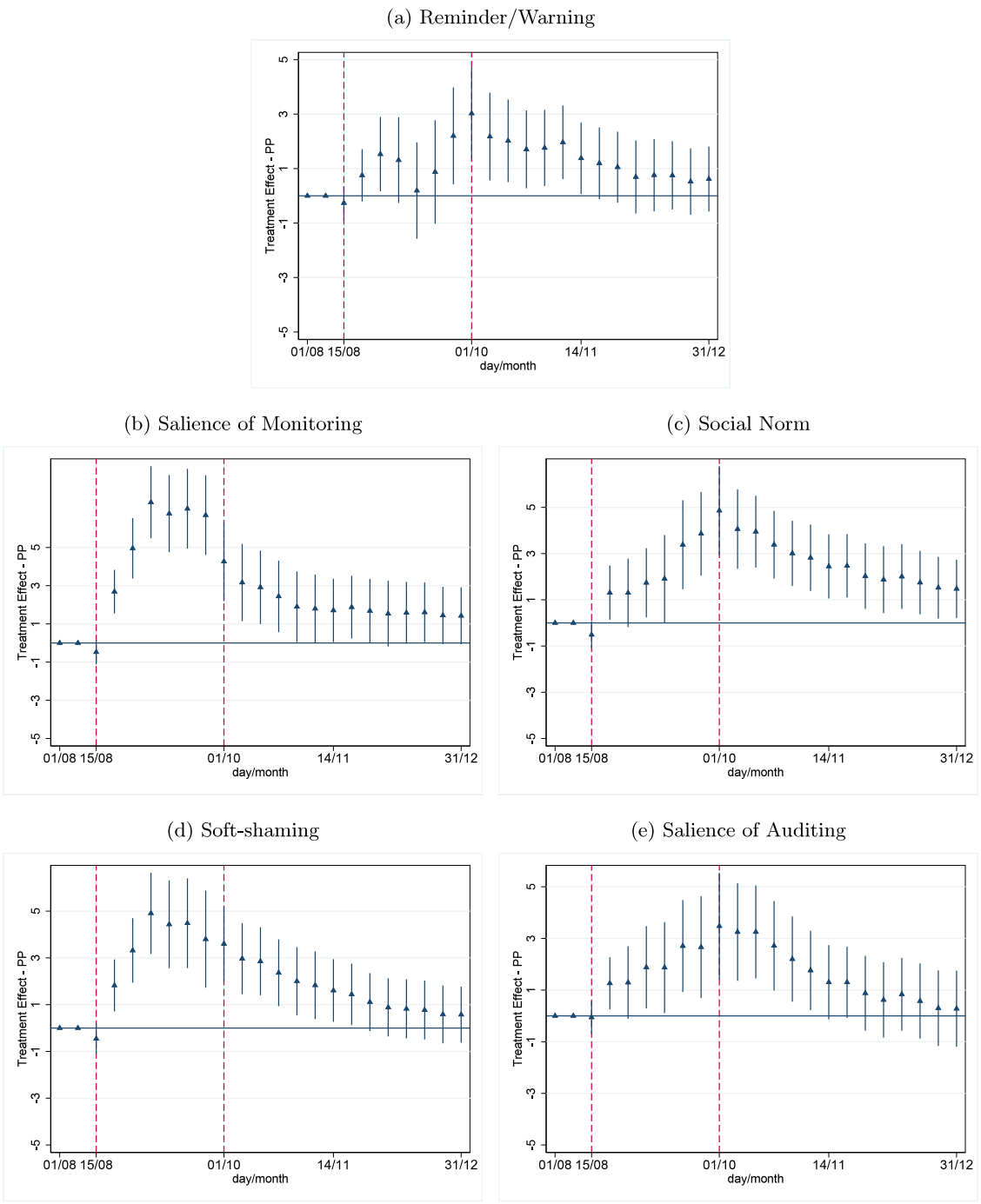


Fig. A.2. Treatment effect on expense report submission, by week, separated by treatment arm. Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Eq. (1) with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

evaluate the internal validity of our research design.[31] We focus here on the basic results of the intervention, presented in Table 7. We aggregate the results for the five months (from September 2016 to January 2017) in which the intervention was in place. All field monitors with access to a tablet assigned by the program were exposed to the intervention (1093 across the country). The necessary information to evaluate the impact of the intervention was obtained from CUNA MAS administrative records. The outcome of interest is the percentage of scheduled field visits that were actually reported by the field monitor by the monthly deadline.

The results suggest that the message that makes monitoring salient is the most effective tool to induce compliance in this population. For instance, in Column 1, the estimated effect is 4.83 p.p. and is significant at the 10% level. Considering a compliance level of 70% for the control group, this estimate represents a 16% reduction in the compliance gap. Point estimates remain unchanged after including control variables

[31] Online Appendix Figure S9 describes the timing of the intervention and Online Appendix Figure S10 provides the details of the treatments, sample size and examples of the text messages delivered in the intervention. Online Appendix Table S9 presents the contents of the text messages delivered during the SMS campaign. Online Appendix Table S10 provides the descriptive statistics and Online Appendix Table S11 evaluates the randomization balance.

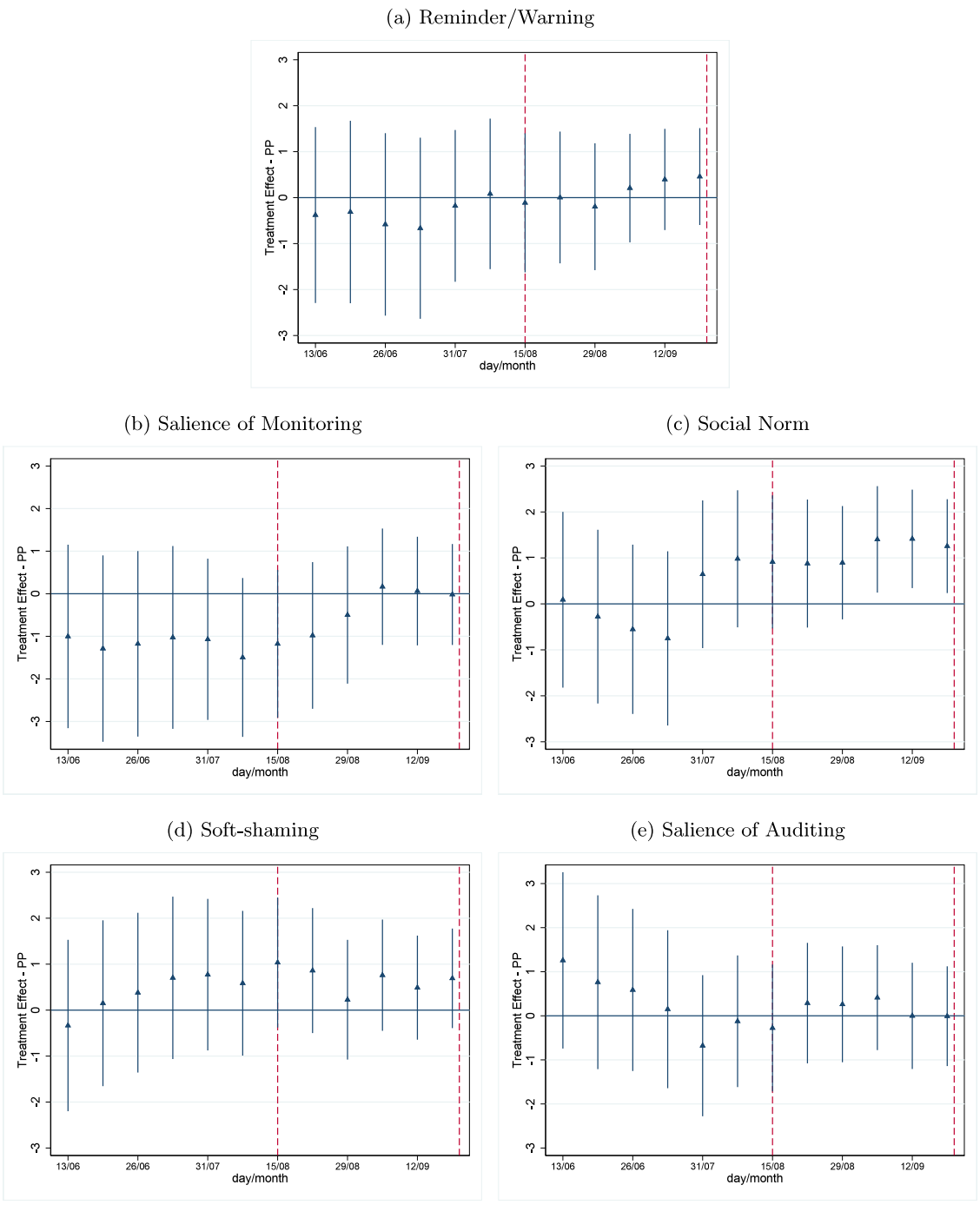


Fig. A.3. Treatment effect on withdrawal of 99% of bank balance, by week, separated by treatment arm. Note: Horizontal axis gives date at which the outcome was measured. Vertical axis is the treatment effect in percentage points, estimated by pooling data from all outcome periods, estimating Eq. (1) with period dummy variables and one treatment dummy per period. Vertical bars represent 95% confidence intervals, where standard errors are clustered at the UGEL level. Dashed vertical lines indicate dates that SMS campaign began and ended.

in the basic specification (Column 2) but standard errors are lower and the effect is significant at the 5% level. Further controlling for baseline outcomes increases the point estimates (5.47 p.p., significant at the 1% level), representing a reduction in the compliance gap of 18% (Column 3). An additional specification drops the October and December months from the sample due to implementation issues.[32] Estimated coefficients slightly increase (Column 4). In the final two specifications, we can reject equality of treatment effects for the social norm and the salience of monitoring treatments at the 6% and 7% significance levels, respectively.[33]

We interpret these results as suggestive evidence that civil servants' characteristics matter for understanding the applicability of the lessons

[32] October was dropped because tablets' operating systems were updated and civil servants' reporting duties were not enforced by program administrators. December was excluded because SMS were not sent due to the holiday season.

[33] Online Appendix Table S12 shows small and statistically insignificant impacts of treatment on the number of visits that must be reported (the denominator of the compliance measure).

Table A.1
Randomization balance analysis.

Variables	Control	Reminder/ warning	Social norm	Salience of monitoring	Soft-shaming	Salience of auditing	Joint Hyp. p-val.
Panel A: Pre-treatment outcomes							
Submitted to maintenance committee at week 20	0.862	0.859	0.855	0.855	0.860	0.859	0.884
	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	
Submitted to oversight committee at week 20	0.860	0.858	0.853	0.852	0.858	0.858	0.861
	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	
Submitted technical form at week 20	0.717	0.697	0.698	0.700	0.715	0.703	0.122
	(0.005)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Submitted commitment act at week 20	0.688	0.666	0.666	0.664	0.677	0.671	0.071
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Submitted expense report at week 20	0.000	0.000	0.000	0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Submitted oversight report to week 20	0.000	0.000	0.000	0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
With approved expense report at week 20	0.000	0.000	0.000	0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Bank balance at 26/06	2600.5	2704.7	2714.4	2764.4	2688.7	2714.5	0.795
	(67.448)	(99.288)	(96.709)	(99.677)	(97.203)	(97.595)	
Panel B: Maintenance CS characteristics							
Sex (% Men)	0.459	0.46	0.448	0.456	0.445	0.462	0.606
	(0.006)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	
Age	46.1	46.2	46.1	46.1	46.2	46.4	0.344
	(0.099)	(0.139)	(0.138)	(0.135)	(0.139)	(0.133)	
Appointed maintenance CS	0.276	0.284	0.274	0.277	0.264	0.285	0.438
	(0.005)	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)	
Hired maintenance CS	0.109	0.108	0.111	0.109	0.108	0.104	0.968
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	
Allocation transfer	7751.1	7850.8	7566.2	7663.7	7697.7	7858.1	0.617
	(97.008)	(138.838)	(131.899)	(133.225)	(134.130)	(135.477)	
Panel C: School characteristics							
Classrooms	5.0	5.1	4.8	4.9	4.9	5.1	0.181
	(0.076)	(0.107)	(0.100)	(0.099)	(0.102)	(0.109)	
Students	121.2	134.3	118.5	141.8	122.2	137.1	0.385
	(2.901)	(8.367)	(5.233)	(22.287)	(4.262)	(6.320)	
Bathroom connected to public drainage system	0.38	0.378	0.37	0.382	0.387	0.374	0.744
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Bathroom connected to septic tank	0.243	0.244	0.235	0.239	0.231	0.247	0.560
	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	
Bathroom connected to a black well	0.26	0.267	0.27	0.26	0.264	0.256	0.799
	(0.005)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	
Bathroom connected to river, ditch or canal	0.029	0.027	0.029	0.026	0.028	0.028	0.983
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	
No bathroom	0.064	0.059	0.069	0.068	0.066	0.067	0.556
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	
Total Land Area	7103.5	7494.5	5820.3	5991.9	6713.3	7059.2	0.335
	(453.294)	(743.779)	(448.382)	(573.630)	(687.226)	(723.130)	
Fully fenced	0.329	0.33	0.314	0.323	0.339	0.322	0.298
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Partially fenced	0.249	0.25	0.264	0.253	0.244	0.25	0.465
	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	
Not fenced	0.398	0.394	0.394	0.398	0.392	0.4	0.986
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Number educ-admin spaces	8.145	8.243	7.902	8.016	8.1	8.186	0.688
	(0.114)	(0.162)	(0.154)	(0.156)	(0.161)	(0.164)	
Number of buildings	2.249	2.206	2.172	2.159	2.213	2.23	0.421
	(0.030)	(0.039)	(0.038)	(0.036)	(0.035)	(0.039)	
Average leaks in pavilions	1.17	1.181	1.178	1.147	1.142	1.154	0.822
	(0.018)	(0.028)	(0.026)	(0.025)	(0.024)	(0.026)	
Average leaks	1.068	1.081	1.082	1.06	1.031	1.051	0.646
	(0.017)	(0.025)	(0.024)	(0.023)	(0.022)	(0.026)	

(*continued on next page*)

of this set of interventions. In particular, we propose the hypothesis that tenure differences might explain the lack of evidence in favor of social norms. Whereas public officials at MINEDU typically have long-term contracts and are highly unionized, CUNA MAS field monitors are hired using a variety of short-term contracts. Since the activities they need to perform do not require of a specific set of skills, they tend to have lower qualifications than officials at MINEDU and higher levels of turnover. Consequently, they are less sensitive to the perceptions that their colleagues may have regarding their work. This may explain why appealing to social norms is ineffective for them. On the other hand, making monitoring salient works better because it is perceived as a tool that can affect their chances of keeping their jobs.

6. Conclusion

This paper provides experimental evidence on the potential effectiveness of messages based on behavioral insights in inducing compliance among civil servants in a setting where the government lacks the capacity to monitor and punish them. We find that these messages are an effective strategy to induce compliance. A wide variety of

Table A.1 (*continued*).

Variables	Control	Reminder/ warning	Social norm	Monitoring	Shaming	Auditing threat	Joint hypothesis
Panel D: District characteristics							
Altitude	1620.8	1658.7	1631.7	1672.8	1682.1	1587.4	0.081
	(18.874)	(26.750)	(26.243)	(26.218)	(26.426)	(26.063)	
Area (% Rural)	0.598	0.603	0.601	0.609	0.606	0.605	0.960
	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	
Electricity	0.842	0.833	0.836	0.835	0.833	0.831	0.749
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	
Public drinking water network	0.674	0.661	0.666	0.668	0.673	0.67	0.840
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Public drainage network	0.398	0.392	0.391	0.401	0.405	0.396	0.832
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	
Internet cafe	0.25	0.242	0.24	0.244	0.246	0.242	0.909
	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	
Bank branch	0.129	0.121	0.115	0.127	0.115	0.123	0.268
	(0.004)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	
Observations	6723	3406	3543	3551	3499	3546	

Note: Sample includes all maintenance civil servants who had not submitted their expense report at the beginning of the SMS campaign. For each treatment arm, means and standard errors are reported for each pre-treatment variable. Final column is the *p*-value for the test of equality of means across all groups. Online Appendix Table S2 contains the variables' full definitions.

message contents were found to be effective, reducing the compliance gap significantly despite the intervention's simplicity and low cost. The apparent invariance of impact with respect to message content is consistent with the intervention increasing salience to overcome limited attention by civil servants. Furthermore, the intervention was found to be effective even when civil servants had already been exposed to it in the previous year, indicating that sending SMS reminders as a matter of policy may have benefits over a course of many years.

We also explored the effects of a similar intervention with civil servants working in a national early childhood program, a population with different characteristics, in particular differences in tenure. Working with CUNA MAS, an early childhood development program, we find that the monitoring treatment is relevant whereas social norms no longer play a role. We interpret this result as potential evidence of the role of the type of labor contract under which the civil servant is governed, since maintenance civil servants are typically hired under long-term contracts but CUNA MAS employees have temporary contracts. These results show that the institutional setting of the intervention can interact with the type of non-monetary incentives being implemented, such that policy designers need to be familiar with the civil servants they are targeting and the environment in which they operate.

Further research is needed to evaluate the role of other types of behavioral insights to increase civil servant compliance. Our treatments mostly focus on limited attention and social norms, but these are certainly not exhaustive of the potential behavioral barriers that one could target. One might also explore alternative means to deliver the interventions beyond text messages. Furthermore, there is substantial room to improve targeting of interventions toward civil servants who are predicted to be non-compliant or for whom treatment is predicted to be most effective. Such approaches, assisted by machine learning techniques made possible given the scale of large interventions, may both save monetary resources and avoid taxing the time and attention of civil servants who were going to comply anyway. Despite this, we believe that the most important result of this paper is to show that using behavioral insights via SMS campaigns can be a powerful, cost-effective, scalable tool to induce compliance among civil servants.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Appendix A. Figures and tables

See Figs. A.1–A.3 and Table A.1.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2022.102995.

References

Aker, J., Collier, P., Vicente, P., 2017. Is information power? Using mobile phones and free newspapers during an election in Mozambique. Rev. Econ. Stat. 99 (2), 185–200.

Akresh, R., Halim, D., Kleemans, M., 2018. Long-Term and Intergenerational Effects of Education: Evidence from School Construction in Indonesia. Technical report, National Bureau of Economic Research.

Al-Ubaydli, O., List, J., Suskind, D., 2017. What can we learn from experiments? Understanding the threats to the scalability of experimental results. Amer. Econ. Rev.: Pap. Proc. 107 (5), 282–286..

Alcott, H., Rogers, T., 2014. The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. Amer. Econ. Rev. 104 (10), 3003–3037.

Allcott, H., 2015. Site selection bias in program evaluation. Q. J. Econ. 130 (3), 1117–1165.

Andreoni, J., Callen, M., Khan, Y., Jaffar, K., Sprenger, C., 2016. Using Preference Estimates to Customize Incentives: An Application to Polio Vaccination Drives in Pakistan. NBER Working Paper No. 22019.

Andrews, I., Oster, E., 2017. Weighting for External Validity. NBER Working Paper 23826.

Ariely, D., Wertenbroch, K., 2002. Procrastination, deadlines, and performance: Self-control by precommitment. Psychol. Sci. 13 (3), 219–224.

Ashraf, N., Bandeira, O., Jack, K., 2014. No margin, no mission? A field experiment on incentives for public service delivery. J. Public Econ. 120, 1–17.

Ashraf, N., Bandiera, O., Lee, S.S., 2015. Do-Gooders and Go-Getters: Career Incentives, Selection, and Performance in Public Service Delivery. HBS and NBER, Department of Economics and STICERD, Harvard Medical School and Harvard Business School.

Azulai, M., Bandiera, O., Blum, F., Kleven, H., La Ferrara, E., Padró, G., Tejada, C.M., 2014. State Effectiveness, Growth, and Development. Technical report, International Growth Centre.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., Walton, M., 2017a. From proof of concept to scalable solutions, with an application. J. Econ. Perspect. 31 (4), 73–102.

Banerjee, A., Chassang, S., Snowberg, E., 2017b. Decision theoretic approaches to experiment design and external validity. In: Duflo, E., Banerjee, A. (Eds.), Handbook of Field Experiments, Vol. I. Elsevier B.V., pp. 142–174.

Banerjee, A., Chattopadhyay, R., Duflo, E., Keniston, D., Singh, N., 2021. Improving police performance in rajasthan, India: experimental evidence on incentives, managerial autonomy, and training. American Economic Journal: Economic Policy 13 (1), 36–66.

Banuri, S., Dercon, S., Gauri, V., 2017. Biased Policy Professionals. World Bank Policy Research Working Paper 8113.

Barlevy, G., Neal, D., 2012. Pay for percentile. Amer. Econ. Rev. 102 (5), 1805–1831.

Barret, P., Treves, A., Shmis, T., Ambasz, D., Ustinova, M., 2019. The Impact of School Infraestructure on Learning. A Synthesis of the Evidence. The World Bank.
Benjamini, Y., Krieger, A.M., Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93 (3), 491–507.
Bertrand, M., Burgess, R., Chawla, A., Xu, G., 2020. The glittering prizes: Career incentives and bureaucrat performance. Rev. Econom. Stud. 87 (2), 626–655.
Best, M.C., Hjort, J., Szakonyi, D., 2017. Individuals and Organizations as Sources of State Effectiveness, and Consequences for Policy. NBER Working Paper No. 23350.
Bhargava, S., Manoli, D., 2015. Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. Amer. Econ. Rev. 105 (11), 3489–3529.
Bobonis, G., Camara, L., Schwabe, R., 2016. Monitoring corruptible politicians. Amer. Econ. Rev. 106 (8), 2371–2405.
Bruhn, M., McKenzie, D., 2009. In pursuit of balance: Randomization in practice in development field experiments. Amer. Econ. J.: Appl. Econ. 1 (4), 200–232.
Burgess, S., Propper, C., Ratto, M., Tominey, E., 2017. Incentives in the public sector: Evidence from a government agency. Econ. J. 127, 117–141.
Bursztyn, L., Gonzalez, A., Yanagizawa-Drott, D., 2018. Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia. NBER Working Paper 24736.
Callen, M., Gulzar, S., Hasanain, A., Khan, Y., 2014. The Political Economy of Public Employee Absence: Experimental Evidence from Pakistan. University of California, New York University, Lahore University of Management Sciences, International Growth Centre.
Callen, M., Gulzar, S., Hasanain, A., Khan, M., Rezaee, A., 2018. Data and Policy Decisions: Experimental Evidence from Pakistan. Working Paper, UC San Diego.
Calzolari, G., Nardotto, M., 2017. Effective reminders. Manage. Sci. 63 (9), 2915–2932.
Celhay, P.A., Gertler, P.J., Giovagnoli, P., Vermeersch, C., 2019. Long-run effects of temporary incentives on medical care productivity. American Economic Journal: Applied Economics 11 (3), 92–127.
Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., Rogers, F.H., 2006. Missing in action: Teacher and health worker absence in developing countries. J. Econ. Perspect. 20 (1), 91–116.
Chong, A., Karlan, D., Shapiro, J., Zinman, J., 2013. (Ineffective) messages to encourage recycling: Evidence from a randomized evaluation in Peru. World Bank Econ. Rev. 29 (1), 180–206.
Cialdini, R.B., Goldstein, N.J., 2004. Social influence: Compliance and conformity. Annu. Rev. Psychol. 55, 591–621.
Cilliers, J., Kasirye, I., Leaver, C., Serneels, P., Zeitlin, A., 2018. Pay for locally monitored performance? A welfare analysis for teacher attendance in Ugandan primary schools. J. Public Econ. 167, 69–90.
Dal Bo, E., Finan, F., 2016. At the Intersection: A Review of Institutions in Economic Development. University of California, Berkeley.
Dal Bo, E., Finan, F., Li, N., Schechter, L., 2018. Government Decentralization under Changing State Capacity: Experimental Evidence from Paraguay. NBER Working Paper 24879.
Dal Bo, E., Finan, F., Rossi, M.A., 2013. Strengthening state capabilities: The role of financial incentives in the call to public service. Q. J. Econ. 1169–1218.
Davis, J., Guryan, J., Hallberg, K., Ludwig, J., 2017. The Economics of Scale-Up. NBER Working Paper 23925.
De La O, A., Martel, F., 2015. Can Intrastate Accountability Reduce Local Capture? Results from a Field Experiment Mexico. Working Paper, Yale University.
Del Carpio, L., 2013. Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru. Princeton University Working Paper, Princeton, NJ.
DellaVigna, S., 2009. Psychology and economics: Evidence from the field. J. Econ. Lit. 47 (2), 315–372.
Delmas, M., Lessem, N., 2014. Saving power to conserve your reputation? The effectiviness of private versus public information. J. Environ. Econ. Manag. 67 (3), 353–370.
Dhaliwal, I., Hanna, R., 2017. The devil is in the details: The successes and limitations of bureaucratic reform in India. J. Dev. Econ. 124, 1–21.
Duflo, E., 2001. Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. Amer. Econ. Rev. 91 (4), 795–813.
Duflo, E., Greenstone, M., Pande, R., Ryan, N., 2013. Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. Q. J. Econ. 128 (4), 1499–1545.
Duflo, E., Hanna, R., Ryan, S.P., 2012. Incentives work: Getting teachers to come to school. Amer. Econ. Rev. 102 (4), 1241–1278.
Eyal, N., 2014. Nudging by shaming, shaming by nudging. Int. J. Health Policy Manag. 3 (2), 53.
Ferraro, P., Miranda, J., 2013. Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment. Resour. Energy Econ. 35, 356–379.
Ferraz, C., Finan, F., 2008. Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes. Q. J. Econ. 123 (2), 703–745.
Finan, F., Olken, B., Pande, R., 2017. The personnel economics of the developing state. In: Banerjee, A., Duflo, E. (Eds.), Handbook of Field Experiments, Vol. 2. Elsevier B.V., pp. 467–514, Chapter 6.
Foxcroft, D., Moreira, M., Almedida, N., Smith, L., 2015. Social Norms Information for Alcohol Misuse in University and College Students. Technical Report 2, Cochrane Database of Systematic Reviews.
Frey, B., Meier, S., 2004. Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. Amer. Econ. Rev. 94 (5), 1717–1722.
Gerber, A., Green, D., Larimer, C., 2010. An experiment testing the relative effectiveness of encouraging voter participation by inducing feelings of pride or Shame. Political Behav. 32 (3), 409–422.
Gerber, A., Rogers, T., 2009. Descriptive social norms and motivation to vote: Everybody's voting and so should you. J. Politics 71 (1), 178–191.
Gertler, P., Shah, M., Alzua, M.L., Cameron, L., Martinez, S., Patil, S., 2015. How does Health Promotion Work? Evidence from the Dirty Business of Eliminating Open Defecation. NBER Working Paper 20997.
Gilligan, D.O., Karachiwalla, N., Kasirye, I., Lucas, A.M., Neal, D., 2022. Educator incentives and educational triage in rural primary schools. J. Hum. Resour. 57 (1), 79–111.
Glewwe, P., Ilias, N., Kremer, M., 2010. Teacher incentives. Amer. Econ. J.: Appl. Econ. 2 (3), 205–227.
Grindle, M., 2012. Jobs for the Boys. Patronage and the State in Comparative Perspective. Harvard University Press.
Hallsworth, M., Egan, M., Rutter, J., Julian, M., 2018. Behavioural Government. Using Behavioural Science to Improve How Governments Make Decisions. The Behavioural Insights Team.
Hallsworth, M., List, J.A., Metcalfe, R.D., Vlaev, I., 2017. The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. J. Public Econ. 148, 14–31.
Hansen, W., Graham, J., 1991. Preventing alcohol, marijuana, and cigarette use among adolescents: Peer pressure resistance training versus establishing conservative norms. Prev. Med. 20 (3), 414–430.
Hanushek, E., 1995. Interpreting recent research on schooling in developing countries. World Bank Res. Obs. 10 (2), 227–246.
Hasnain, Z., Manning, N., Pierskalla, J.H., 2014. The promise of performance pay? Reasons for caution in policy prescriptions in the core civil service. World Bank Res. Obs. 29, 235–264.
Holmstrom, B., Milgrom, P., 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. JL Econ. Org. 7, 24.
Karachiwalla, N., Park, A., 2017. Promotion incentives in the public sector: Evidence from Chinese schools. J. Public Econ. 146, 109–128.
Karlan, D., McConnell, M., 2014. Hey look at me: The effect of giving circles on giving. J. Econ. Behav. Organ. 106, 402–412..
Karlan, D., McConnell, M., Mullainathan, S., Zinman, J., 2016. Getting to the top of mind: How reminders increase saving. Manage. Sci. 62 (12), 3393–3411.
Karlan, D., Morten, M., Zinman, J., 2012. A Personal Touch: Text Messaging for Loan Repayment. NBER Working Paper 17952.
Khan, A.Q., Khwaja, A.I., Olken, B.A., 2016. Tax farming redux: Experimental evidence on performance pay for tax collectors. Q. J. Econ. 131 (1), 219–271.
Khan, A.Q., Khwaja, A.I., Olken, B.A., 2019. Making moves matter: experimental evidence on incentivizing bureaucrats through performance-based postings. American Economic Review 109 (1), 237–70.
Kowalski, A.E., 2018. How to examine external validity within an experiment. Journal of Economics & Management Strategy.
Leaver, C., Ozier, O., Serneels, P., Zeitlin, A., 2021. Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools. Amer. Econ. Rev. 111 (7), 2213–2246.
Lee, D., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. Rev. Econom. Stud. 76 (3), 1071–1102.
Lewis-Faupel, S., Neggers, Y., Olken, B., Pande, R., 2016. Can electronic procurement improve infraestructure provision? Evidence from public works in India and Indonesia.. Amer. Econ. J.: Econ. Policy 8 (3), 258–283.
Lynch, J., Mowrey, R., Nesbitt, G., O'Neill, D., 2004. Risky business: Misperceived norms of sexual behavior among college students. NASPA J. 42 (1), 21–35.
Mackie, G., Moneti, F., Shakya, H., Denny, E., 2015. What are Social Norms? How are They Measured?. UNICEF/UC San Diego.
Meyer-Sahling, J., Schuster, C., Mikkelsen, K., 2018. Civil Service Management in Developing Countries: What Works? Evidence from a Survey with 23,000 Civil Servants in Africa, Asia, Eastern Europe and Latin America.
Miller, D., Prentice, D., 2016. Changing norms to change behavior. Annu. Rev. Psychol. 67, 339–361..
Molina, E., Carella, L., Pacheco, A., Cruces, G., Gasparini, L., 2017. Community monitoring interventions to curb corruption and increase access and quality in service delivery: A systematic review. J. Dev. Eff. 9, 462–499.
Mortensen, C., Neel, R., Ciladini, R., Jaeger, C., Jacobson, R., Ringel, M., 2019. Trending norms: A lever for encourging behaviors performed by the minority. Soc. Psychol. Personal. Sci. 10 (2), 201–210.
Muralidharan, K., Niehaus, P., 2017. Experimentation at scale. J. Econ. Perspect. 31 (4), 103–124.
Muralidharan, K., Niehaus, P., Sukhtankar, S., 2016. Building state capacity: Evidence from biometric smartcards in India. Amer. Econ. Rev. 106 (10), 2895–2929.

Olken, B.A., 2007. Monitoring corruption: Evidence from a field experiment in Indonesia. J. Polit. Econ. 115 (2), 200–249.
Pepinsky, T., Pierskalla, J., Sacks, A., 2017. Bureaucracy and service delivery. Annu. Rev. Political Sci. 20, 249–268.
Peters, J., Langbein, J., Roberts, G., 2018. Generalization in the tropics-development policy, randomized controlled trials, and external validity. World Bank Res. Obs. 33 (1), 34–64.
Rogers, T., Goldstein, N., Fox, C., 2018. Social mobilization. Annu. Rev. Psychol. 69, 357–381..
Rosenzweig, M.R., Udry, C., 2020. External validity in a stochastic world: Evidence from low-income countries. Rev. Econom. Stud. 87 (1), 343–381.
Savje, F., Aronow, P.M., Hudgens, M.G., 2017. Average treatment effects in the presence of unknown interference. arXiv preprint arXiv:1711.06399.
Schultz, P.W., 1999. Changing behavior with normative feedback interventions: A field experiment on curbside recycling. Basic Appl. Soc. Psychol. 21 (1), 25–36.
Sonntag, A., Zizzo, D., 2015. On reminder effects, drop-outs and dominance: Evidence from an online experiment on charitable giving. PLoS One.
Sparkman, G., 2021. Dynamic norm interventions: How to enable the spread of positive change. In: Walton, G., Crum, A. (Eds.), Handbook of Wise Interventions: How Social Psychology Can Help People Change. The Guilford Press.
Sparkman, G., Walton, G., 2017. Dynamic norms promote sustainable behavior, even if it is counternormative. Psychol. Sci. 28 (11), 1663–1674.
Vervloet, M., Linn, A., van Weert, J., de Bakker, D., Bouvy, M., van Dijk, L., 2012. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: A systematic review of the literature. J. Am. Med. Inform. Assoc. 19, 696–704.
Voigt, K., 2013. "If you smoke, you stink". Denormalisation strategies for the improvement of health-related behaviours: The case of Tobacco. In: Strech, D., Hirschberg, I., Marckmann, G. (Eds.), Ethics in Public Health and Health Policy. Concepts, Methods, Case Studies. Springer, pp. 47–62..
WHO, 2009. Changing Cultural and Social Norms that Support Violence. Technical report, World Health Organization.
Young, H., 2015. The evolution of social norms. Annu. Rev. Econ. 7, 359–387.