Strengthening Fragile States: Evidence from Mobile Salary Payments in Afghanistan[*]

Joshua E. Blumenstock[†]
U.C. Berkeley

Michael Callen[‡]
London School of Economics

Anastasiia Faikina
Analysis Group

Stefano Fiorin[§]
Bocconi

Tarek Ghani[¶]
Washington University

June 8, 2023

Abstract

Building state capacity is uniquely challenging in fragile states. We report results from a randomized evaluation of a major Afghan government initiative to increase capacity by modernizing its payroll. The reform, which required teachers to biometrically register and receive salary payments via mobile money, did little to reduce payments to non-existent "ghost" workers, but significantly reduced delays. The reform also improved educational outcomes and increased formal financial inclusion. The impacts were not immediate – highlighting the importance of long time-horizons – and were largest in urban areas. The results have implications for state-building and are potentially actionable for policymakers.

[*] *Authors' Note*: We are grateful to the many devoted public servants in the Ministry of Education, the Policy Coordination Unit of the Office of the President, the Asan Khedmat unit of the Ministry of Information Communications Technology, the Treasury Department of the Ministry of Finance, as well as the dedicated staff at Roshan and Afghan Wireless. We thank Eddy Chebelyon, Miguel Fajardo-Steinhäuser, Shahim Kabuli, Jiyoung Kim, Aarya Nijat, Ofir Reich, Kelsey Reiff, and Sami Safiullah for excellent research assistance. Oriana Bandiera, Eli Berman, Gharad Bryan, Ernesto Dal Bó, Craig McIntosh, Karthik Muralidharan, Paul Niehaus, Rohini Pande, Imran Rasul, Jacob Shapiro, Noam Yuchtman and many others provided insightful feedback. We thank audiences at Bocconi, CEGA, CERGE-EI, CEPR, UCSD, Y-RISE, FCDO, ESOC, QMUL, SIOE, SEEDEC, Stanford, USC, and Warwick. We acknowledge funding from the International Growth Center, Innovations for Poverty Action, and the Jameel Poverty Action Lab. This RCT was registered in the American Economic Association Registry (AEARCTR-0001641). IRB approval was obtained from Harvard University, UCSD, UC Berkeley, and Washington University of St. Louis. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†] University of California at Berkeley, School of Information, Berkeley, CA 94720. jblumenstock@berkeley.edu

[‡] London School of Economics, Department of Economics, Houghton Street, London, WC2A 2AE. m.j.callen@lse.ac.uk

[§] Bocconi University, Department of Economics, Via Roentgen 1, 20136 Milano, Italy. stefano.fiorin@unibocconi.it

[¶] Washington University in St. Louis, Olin Business School, 1 Brookings Drive, St. Louis, MO 63130. tghani@wustl.edu

1 Introduction

By 2030, one third of the world's poor are expected to live in 15 extremely fragile countries and 80% of the world's poor are expected to live in 60 fragile countries (OECD, 2022).[1] There is a consensus among development economists that increasing state capacity is critical to reducing poverty, especially in these settings (Dal Bó, Finan and Rossi, 2013; Finan, Olken and Pande, 2017; Besley et al., 2022). However, state-building is uniquely challenging when basic administrative structures do not function properly. Effective governance relies on public employees to implement essential functions — from the maintenance of order and tax collection to the provision of services like health and education (Weber, 1919; Tilly, 1985; Bates, 2001; Besley and Persson, 2009). But what happens when states lack the ability to effectively identify and pay civil servants?

With its rudimentary administrative systems, violent competition for control of territory, and limited financial and economic development, Afghanistan exhibits key features of an extremely fragile state. This paper reports on a randomized evaluation of a multi-year Afghan government initiative – called the Mobile Salary Payment (MSP) reform – designed to strengthen the basic architecture needed to pay civil servants at large scale. The reform involved two main components. The first required employees to provide fingerprint-based biometric identification in order to register for a mobile money wallet. The second transitioned employee salary payments to a mobile money platform, from the status quo system that relied on banks and a cash-based network of "trusted agents."

These reforms, which were initially focused on the Ministry of Education (hereon, MoE), had five goals: (1) to remove "ghost" workers (i.e., fake employees added to the payroll so that others can capture their salary); (2) to reduce "leakage" between what was disbursed by the government and what was received by employees; (3) to reduce delays and improve the salary experience; (4) to improve educational outcomes of students; and (5) ultimately to promote broader adoption of mobile money and thereby strengthen the digital financial system. Reforms such as these were also enthusiastically supported by international donors, who viewed salary payments – especially for the police and the army – as essential to keeping the Afghan state intact against the ongoing insurgency, and so covered about two-thirds of the state's total wage bill.[2]

[1]The 15 extremely fragile countries are Afghanistan, Burundi, Central African Republic, Chad, Congo, Democratic Republic of the Congo, Eritrea, Equatorial Guinea, Haiti, Iraq, Somalia, South Sudan, Sudan, Syrian Arab Republic, and Yemen.

[2]The MoE was the third largest ministry (behind defense and interior) in terms of expenditure with an annual budget of 35 billion AFN (450 million USD). In 2018, the Afghan government spent 71% of its total

Our evaluation centers on the salary reforms implemented between 2017-2020 by the Afghan MoE, with support from the Ministry of Finance and the Ministry of Communications and Information Technology. The evaluation involved approximately 34,000 public employees (primarily teachers) working in roughly 1,500 schools across three conflict-affected provinces that account for about 15% of Afghanistan's population. The government prioritized the MoE because it employs roughly 70% of the country's civilian public servants, and because the status quo payroll system was plagued by inefficiencies. At baseline, for example, 54% of teachers reported receiving their salaries with (often substantial) delays.

Working with the Afghan government, we designed a field experiment to separately study the registration and mobile money payment components of the reform. We divided the 1,500 schools into 401 registration zones of about 80 employees each, and randomly assigned zones into three treatment groups that varied the planned timing of implementation: (i) an early payment group who were both registered and paid using mobile money as soon as practicable ("Early registration, Early mobile money payments" or "EE" for short); (ii) a delayed payment group, who were registered at the same time as the first, but intended to begin mobile payments with a six month delay ("Early registration, Delayed mobile money payments" or "ED" for short); and (iii) a control group that were meant to be both registered and begin mobile payments after the experiment ("Delayed registration, Delayed mobile money payments" or "control" for short).[3]

Our study provides five sets of results, one for each of the five goals listed above. First, we provide estimates of the number of ghost workers who were being paid but not actually working. Superficially, we find that 2.8% of employees never register biometrically, despite official warnings that failing to do so would prevent them from being paid. We expect this is an underestimate of the true number of ghost workers given reports of non-working "stand-ins" providing biometric information in order to continue receiving salary payments. We therefore used a forensic "litmus test" survey to quiz registered employees on basic facts that any real employee should know (e.g., the name of the school's principal). By comparing the quiz performance of certified employees (who were verified to be present during a separate round of unannounced audit visits) with the performance of other registered employees, we estimate that the true share of employees who are ghost workers is likely between 8.4% and

annual 3.4 billion USD recurrent budget on salaries, with roughly two-thirds funded by international donors (World Bank, 2019).

[3]We describe these groups in detail in Section 2.

20.4% of paid employees. In practice, officials only removed half of non-registered employees from the payroll, limiting savings to about 1.2% of the wage bill, which is statistically indistinguishable from turnover in the control group.

Second, the leakage of salaries during disbursement – an object of major policy concern because it necessitated the use of trusted agents or state employees who must physically transport salaries as cash – appears to have been a minor issue. At baseline, employees report paying about 24 AFN, or less than 0.5% of salary, to receive their monthly payment. While treatment reduces this further after two years to a level near zero, salary leakage of this form does not appear to be a major source of corruption in Afghanistan. Even though transporting physical cash is the only mechanism for many governments to pay rural employees, our evidence suggests this does not necessarily imply leakage.

Third, the reform improved the salary experience of employees – but the effects were not immediate.[4] Over the course of the reform, payment delays – which were pervasive – were substantially reduced, from affecting 55% to 30% of all employees. This improvement was driven mainly by reductions in "first mile" delays (i.e., those occurring at the Education and Finance Ministries in Kabul): in treatment zones, first mile delays dropped from 45.7% to 13.7% after one year; in control zones, they dropped from 41.7% to 27.6% (s.e. of endline difference = 0.3%; p-value=0.000).[5] By contrast, treatment induced an *increase* in "last mile" delays (i.e., those associated with local bank branches and trusted agents), driven by delays relating to the mobile network operator lacking the staff necessary to cash out employees. By the second year, however, we estimate that the reform reduced total delays by 26.5 percentage points in cities, though they remained elevated in rural areas. Correspondingly, while treatment only marginally increased employees' stated support for the reform in the first year, after two years a clear majority of treated employees supported expanding the reform nationally.

Fourth, we find evidence that the reform had impacts on student learning. Since no student-level administrative data were available, our team conducted in-home learning assessments with a representative sample of 1,101 students using methods validated in prior research (Burde, Middleton and Samii, 2017). While pandemic-related fieldwork restrictions limit our assessments to one year into the reform, these tests indicate that the reforms led to a

[4]The statistics reported in this paragraph correspond to those reported in Figures 4 and 5, which are based on comparisons of the EE treatment and the control group.

[5]The standard error of the difference at baseline is 3.8% and the corresponding p-value is 0.298.

significant improvement in test scores in urban areas where the reforms were most successful. The effect in rural areas was negative but statistically insignificant.

Fifth and finally, the reform increased mobile money adoption and use. For example, ToT estimates indicate that self-reported mobile money use for peer-to-peer transfers increased by 30 percentage points relative to 6% usage in the control group (s.e. of estimate = 4.8pp; p=0.000). The effects observed in survey data are also apparent in administrative data on mobile money use, where we observe treatment effects on mobile money deposits, mobile money purchases, and mobile airtime transfers. Over time, mobile money account balances increase by roughly 0.6% of salary for every month that the user remained on the platform.

Collectively, these results yield three main insights about strengthening extremely fragile states. First, we find consistent evidence that the government's salary reforms were most effective in secure, urban areas. Any country considering whether to expand mobile payments faces the basic question of whether to build the ecosystem in cities or rural areas. That this program would succeed most rapidly in cities was not obvious *ex ante*: many involved in this program believed that salary payment modernization should begin in rural areas where state capacity was more limited. In particular, the Finance Ministry argued that state employees in the rich and relatively secure provincial capitals were already served by banks and that salary issues and corruption were more severe in rural areas. In practice, however, the salary reforms were most immediately successful in urban areas: that is where individual support rose quickest and where payment delays were most substantially reduced.

Second, our reform highlights the potential for a targeted reform to produce downstream benefits across different sectors, and that sometimes the most pronounced effects are not where policymakers (or researchers) anticipate. In our case, the reform focused on cleaning up the government payroll. Biometric registration was expected to reduce ghost workers and mobile payments to reduce salary leakage, but our data suggest that these measures were not especially successful. Instead, the ease with which stand-in employees could provide biometric data to keep salary lines open remained a significant issue that undermined the goal of eliminating ghost workers. However, our results on student learning and mobile money usage suggest that, when successfully implemented, salary payment modernization can catalyze improvements in service delivery and in financial inclusion.

Finally, while political volatility can limit the window for policy reform in extremely fragile countries, our work in Afghanistan underscores the importance of sustained engagement across relatively long time horizons. In particular, the reforms implemented in the MoE were only possible because they built on a decade of related efforts to introduce mobile salary payments, first in the private sector (Blumenstock et al., 2015, 2022) and then in a small ministry in Kabul, before expanding across a much broader geography. And while the reforms eventually achieved many of the stated objectives, had the reform been called off after only one year, it would have appeared a failure by several measures. The immense pressure to achieve quick impacts is often at odds with the requirements for well-designed and sustainable policy change.

Our study is closely linked to the extensive literature on strengthening state capabilities in low-income countries, as summarized in Finan, Olken and Pande (2017), Bandiera et al. (2019), and Besley et al. (2022). The experiments in this literature provide valuable insights into enhancing states' ability to perform essential functions such as tax collection, healthcare provision, and education, through measures such as improving selection (cf. Ashraf et al., 2020; Finan, Bó and Rossi, 2013) and enhancing incentives (cf. Mbiti et al., 2019; Joppe, Muralidharan and Menno Pradhan, 2017; Muralidharan and Sundararaman, 2011; Khwaja, Khan and Olken, 2016, 2019). Our study complements this literature by examining a reform that sought to develop the basic administrative systems – often rudimentary in highly fragile settings – required to enhance recruitment and improve incentives. We also contribute to the literature on measuring the degree of waste in service provision, such as absence and leakage (Bandiera, Prat and Valletti, 2009; Chaudhury et al., 2006; Reinnika and Svensson, 2004; Olken, 2007; Olken and Pande, 2012).

Other related studies have focused on leveraging technology to enhance state effectiveness. These efforts include using biometric technology to develop citizen identification systems (Muralidharan, Niehaus and Sukhtankar, 2016, Forthcoming; Giné, Goldberg and Yang, 2012), which are essential for functions such as providing social protection, as well as tracking payments (Banerjee et al., 2020), and increasing state accountability (Duflo, Hanna and Ryan, 2012; Callen and Long, 2015; Callen et al., 2016; Dhaliwal and Hanna, 2017; Bossuroy, Delavallade and Pons, 2019). The mobile money results also connect to recent literature on how mobile money affects financial inclusion (Aker et al., 2016; Blumenstock, Eagle and Fafchamps, 2016; Blumenstock, Callen and Ghani, 2018; Jack and Suri, 2014; Suri and Jack, 2016).

Finally, our paper contributes to the academic literature focused on the early stages of state formation (cf. Besley and Persson, 2011; Acemoglu and Robinson, 2012; Commission on State Fragility, Growth, and Development, 2018), and on rebuilding fragile states, particularly in the context of international intervention (cf. Fearon and Laitin, 2004; Weinstein, 2005; U.S. Army, 2006). Within this literature, there is a small but growing experimental literature on how states begin to develop basic capabilities (Weigel, 2020; Balán et al., 2022; Sanchez de la Sierra, 2020). However, the evidence base for such states remains limited, given the challenges of collecting high-quality data and conducting rigorous research in these contexts. For example, only 0.6% of the randomized control trials registered with the AEA are in the 15 countries identified as extremely fragile by the OECD, and only 0.2% of articles in the top five economics journals since 2000 use data from these states.[6] An even smaller share of the literature is at large scale (Muralidharan and Niehaus, 2017) and in collaboration with fragile governments.

The rest of the paper proceeds as follows. Section 2 describes the MSP reform, our research design, and administrative and survey data sources. Section 3 presents results on biometric registration and ghost employees. Section 4 presents results on the effects of MSPs on the salary payment experience, financial inclusion, and education quality. Section 5 concludes.

2 Institutional Details, Research Design, and Data

The research design aims to measure the extent to which the payroll reforms implemented by the Afghan government achieved the five policy objectives enumerated above. This section describes the experimental sample; provides details on the reforms; discusses treatment assignment and treatment compliance; and describes the administrative and survey data used to evaluate the reform.

2.1 Experimental Sample

Our experimental sample consists of the 34,422 MoE employees who appear at least once on official government payroll records between March 2017 and February 2020, ostensibly because they worked in one of 1,530 schools in Kandahar, Nangarhar, and Parwan provinces (spanning a total of 42 districts). These schools were divided into 401 experimental registration zones using information on schools' location, number of employees, and security. In consultation with the implementing partners, about 300 other institutions were instead deemed to be inappropriate for the experiment because they belonged to other zones that

[6] The statistics reported in this sentence are from calculations by the authors. The data on top five economics journal articles come from the Web of Science and cover the period January 1, 2000 - January 30, 2023. Calculations are available on request.

were less secure and consisted in large part of non-typical schools (e.g., madrassas and vocational training institutes) or schools with less than ten employees. Nevertheless, our sample contains some insecure regions, some non-traditional schools, and some small schools. Approximately 12% of the teachers in our sample are female and over half of the schools taught girls.

2.2 Details of the MSP Reform

The MSP reform involves three stages:

Stage 1 - Registration: A Mobile Network Operator (MNO) contracted by the government attempted to provide a mobile wallet to every teacher on the payroll during a pre-scheduled and pre-announced visit to one school (registration center) in each registration zone. During the visit, field agents collected biometric measurements of ten fingers, and the teachers' national ID number. The purpose was to provide mobile wallets (in the form of mobile-money-enabled SIM cards) and to identify whether the teachers currently on the payroll could actually be found in the schools. Importantly, before registration teams visited schools, teachers were informed by the MoE that their ability to continue receiving a salary depended on appearing for registration.

Stage 2 - Adjudication: The government reform team compared the registration lists against the payroll record and transmitted both lists and their discrepancies to a committee at the MoE. The committee was meant to determine who should continue getting paid based on these data. While a comparison of registration and payroll data could identify potential ghost workers, it was ultimately up to the MoE to remove them.

Stage 3 - Payment: The list of verified employees created during the adjudication step was transmitted to the MNO. The MNO activated the wallet-enabled SIM cards for teachers and required biometric authentication when employees exchanged mobile money for cash at mobile money agents.

2.3 Assignment to Treatment

The top panel of Figure 2 shows the timeline of the reform. The basic design for the evaluation randomizes the timing of the registration process and the implementation of mobile money payments at the registration zone level. Here, we describe the design as conceptualized at the beginning of the reform. The 401 experimental zones were randomly assigned to one of the three treatment arms, with treatment assignment stratified at the district level:

1. Early registration, Early mobile money payments (EE): 137 zones where payroll verification was scheduled to begin in May 2018 and MSP payments were scheduled to begin in October 2018;

2. Early registration, Delayed mobile money payments (ED): 129 zones where payroll verification was also scheduled in May 2018, but MSP payments were scheduled to begin six months after those in the EE group (April 2019);

3. Delayed registration, Delayed mobile money payment (DD or control): 135 zones where payroll verification and MSP payments were delayed respectively by four (September 2018) and six months (April 2019) with respect to the EE group.

Figure 1 provides a map of the registration zones by status across our three provinces. Appendix Table A.1 checks for balance using the survey data described in Section 2.5 below.

2.4 Treatment Compliance and Estimation

The middle and bottom panels of Figure 2 display the evolution through time of the number of employees who registered for a mobile wallet and those who started being paid via MSP by treatment assignment. As is evident in the middle panel, the majority of employees in the EE and ED group were registered between May 2018 and July 2018. Most of the registration in the control group happened between November 2018 and December 2018. Due to capacity constraints of the MNO, registration activities in this group continued at a slow pace in 2019.

As can be seen in the bottom panel of Figure 2, employees in the EE group started to be paid by MSP in October 2018 in accordance with the schedule. However, because there were insufficient mobile money agents, the transition of employees into mobile payments was gradual, and many of the employees in the EE group received their first mobile payment either in November or December 2018. During these months, in disagreement with the initial plans, about half of the employees in the ED group were also transitioned to mobile payments because the government wished to show progress on increasing the number of employees paid via mobile money. Only a small minority of employees in the control group were paid via mobile payments during the study period, starting mostly in February and April 2019. After April 2019, the number of employees paid via mobile money did not increase in the ED and control group, reflecting the persisting capacity constraints of the MNO. In April 2019 (when our first endline survey occurred), 72% of employees in the EE group, 44% in the ED group, and 8% in the control group reported being paid by mobile money. In May 2020 (our second endline survey), these numbers increased to 81%, 49%, and 33%, respectively.

While compliance with the schedule predetermined by the treatment assignment was far from perfect, the randomization generated variation in the number of employees impacted by the reform across treatment groups, as displayed in Figure 2 and Appendix Table A.2. In order to account for imperfect compliance, in our main specification we estimate Treatment-on-the-Treated (ToT) effects by instrumenting an indicator for whether the respondent received their salary via mobile money with treatment dummies using the following 2SLS specification:

$$\begin{aligned} Y_{izdt} &= \gamma + \beta_{MSP} MSP_{izdt} + \mu_d + \varepsilon_{izdt} \\ MSP_{izdt} &= \theta + \varphi_{EE} EE_z + \varphi_{ED} ED_z + \eta_d + \epsilon_{izdt} \end{aligned} \tag{1}$$

where Y_{izd} is the outcome at time t (as captured in one of the surveys) for employee i from registration zone z located in district d; EE_z is an indicator for early registration, early mobile money payment treatment zones; and ED_z is an indicator for early registration, delayed mobile money payment treatment zones. We use EE_z and ED_z to instrument MSP_{izdt}, an indicator for whether employee i reported receiving their salary via mobile money at time t. We control for district (strata) fixed effects, μ_d and η_d. Standard errors are clustered at the registration zone (treatment unit) level.

2.5 Administrative and Survey Data

In our analysis, we use data from two main administrative sources. First, from the Afghan government, we obtained the payroll records for all employees in our sample. These are generated when schools submit monthly requests for salaries to the Ministry of Finance, and contain detailed information about schools, employees, and their salaries. Second, from the MNO, we obtained records of every individual who registered for a mobile wallet, as well as the complete and detailed mobile money transaction records for all employees impacted by the reform (which is the same as the evaluation sample described in Section 2.1).

We complemented the administrative data with several different surveys. The first set of 'audit' surveys were primarily intended to measure teacher attendance; as such, each surveyed school was visited without prior warning, and all present teachers were interviewed. In total, there were three such rounds of unannounced in-person surveys: May 2018 (baseline), November 2018 (midline), and April 2019 (endline 1). We attempted to conduct these surveys in one randomly-selected school in each of the 401 registration zones, but due to security issues, the survey company could visit only 375 schools at baseline, 369 at midline, and 362 at endline 1.

The second set of surveys were designed to measure the effect of the reform on teachers' payment experience and support for the reform. These were longer surveys conducted with a random subsample of up to three teachers among those present during the audit visit. In total, we conducted in-person interviews with 1,005 teachers at baseline and 974 teachers at endline 1. We also administered an additional round of surveys in May 2020 (endline 2) over the phone and using a shortened survey instrument, since in-person surveys were not possible due to the Covid-19 pandemic.[7]

Third, we conducted a litmus test survey in February–March 2019 that was designed to assess the extent to which individuals registered on the payroll were in fact legitimate employees. This was a phone survey conducted with a random sample of 2,663 employees. The survey asked for basic details about the employee's workplace and job that we would expect any bona fide employee to be able to answer; questions included the school's geographic location, the principal's identity, and the rank of their position. We use this survey to compare knowledge of legitimate employees to those of potential ghosts and stand-ins.

Fourth, in order to measure the effect of the reform on students' learning, in May 2019, we conducted a learning assessment of 1,001 children aged 6–10 from a random sample of households who lived in the proximity of the schools selected for the in-person teachers' surveys. One child in each household was randomly chosen to take the learning assessment, which tested both literacy and numeracy skills.

Finally, in 2019, we conducted semi-structured interviews with key stakeholders (including among others teachers, government officials, journalists and NGO workers) to gather their beliefs on the existence of ghost workers.

3 The Impact of Biometric Registration

Before the reforms we study, the Government of Afghanistan lacked a reliable list of MoE employees, and there were numerous reports suggesting high rates of teacher absenteeism

[7] We attempted to contact via mobile phone 945 employees whom we interviewed during the endline 1 survey and for whom we obtained a valid phone number. We successfully interviewed 739 of them.

and payroll leakage.[8] In practice, there were several issues affecting the payroll: some payroll entries are for people who do not exist (ghosts); some corresponded to individuals who appear to have never worked (stand-ins); and others are people legitimately hired by MoE who sometimes did not show up for work (absentees).

The registration reform was designed, in part, to eliminate ghosts and to make it more difficult for stand-ins to continue receiving salary on behalf of someone else. This is because the biometric registration required that the employee appear in person, with an ID card, in order to continue receiving salary – which would be difficult if the person were entirely fictitious. Post-reform, biometric authentication was required to receive the monthly salary payments, thus raising the cost of stand-ins by requiring a single individual to register and to cash out salaries monthly for each unique position. Biometric registration also provided a foundation for future reforms that could link teacher pay to biometrically-verified attendance at schools.

In practice, the impact of biometric registration was limited by both the incomplete removal of non-registered ghosts and also some enrollment of stand-ins. This section describes our first set of results, which (i) document the share of employees removed from the payroll following biometric registration and (ii) attempt to quantify the number of total ghosts, including stand-ins, on the payroll. Section 4 discusses the impacts of the reforms transitioning salary payments to mobile money.

3.1 Identifying and Removing Ghost Employees

We exploit the rollout of biometric registration to estimate the number of ghost MoE employees. As can be seen in Panel A of Appendix Table A.3, the majority of employees in early registration zones registered during the first registration wave (94%) or with some delays (3.2%).[9] Only 2.8% of these employees never registered, providing a conservative, lower bound estimate for the number of ghosts.[10] Their salaries amount to about AFN 3.4 million

[8] See, for example, Special Inspector General for Afghanistan Reconstruction (2016) and Special Inspector General for Afghanistan Reconstruction (2017).

[9] To generate conservative estimates, we exclude from this analysis 23 early registration zones in Nangarhar, in which less than 50% of employees registered due to ongoing violent conflict. We also exclude all employees who were paid only during some, but not all, of the six months before the registration, in order to conservatively estimate the share of employees who never registered among those who clearly appear as regular and stable employees according to the payroll records.

[10] The mobile network operator shared data about registered employees until August 5, 2020 (i.e., more than two years after the registration activities started), so employees are classified as “never registered” if they did not registered by this date. While a few employees might have registered after this date, it is unlikely that many might have done so: of the registered employees within our sample of interest, more than 99.5% did so by the end of 2018, and 99.9% by the end of 2019, meaning that only 0.01% registered in 2020.

(USD 44,000 using the 2018/19 exchange rate) or 2.7% of the total monthly wage bill in these schools.

However, not all of the people who failed to register were removed from the payroll. After the primary registration wave, the MoE formed an adjudication committee tasked with eliminating ghost employees. The committee compared payroll disbursement records to registration records and determined the identity of employees who did not register or whose records did not match. If an employee could not be confirmed, the committee was responsible for removing them from the payroll. Using payroll records, we can track whether MoE continued to request salaries for employees who never registered.[11] Panel B of Table A.3 (column 3) shows the share of never-registered employees who continued receiving their salaries after the registration process. 89.6% of them remain on the payroll one month after the start of the reform. This share decreases to 56.9% within 12 months of the registration. However, even 20 months after the registration a salary was still being requested for 53.6% of never-registered employees. As a result, the incomplete removal of employees who never registered limited savings to only 1.2% (AFN 1.5 million or USD 20,000) of the total monthly wage bill in the early registration zones.

The incomplete removal of non-registered employees limits the effect of biometric registration on the payroll. In Appendix Tables A.4, A.5, and A.6, we test for post-registration reductions in the total wage bill aggregating at either the individual, school or registration zone level respectively. While the results are consistently negative, they are not significant at conventional levels. Therefore, we cannot reject the null hypothesis that payroll reductions in the treatment group are equal to the standard turnover rates observed in the control group.

3.2 The Persistent Problem of Stand-in Employees

Biometric registration also did not fully address the issue of stand-in employees. In fact, the reform may even have created incentives for corrupt actors to send stand-ins to register, thereby maintaining salary lines that were previously associated with nonexistent ghosts. In this scenario, the reform would impose the burden of sending stand-ins every month to collect salaries on corrupt actors, but would not yield any financial benefits for the government.

To better quantify the extent of this problem, we conducted a short litmus test phone survey

[11]We cannot verify if these salaries are withdrawn and thus may underestimate payroll reductions.

with MoE employees between February and March 2019. The twin purposes of the survey were to confirm that employees who did not register were indeed ghosts, and to explore the extent to which registered employees might have actually been stand-ins.

We sampled a total of 2,663 employees who had a phone number listed in the payroll records. Importantly, these records are from a roster held by the government prior to the reform. From the Early Registration group, we included all employees who did not register, all who registered after the in-person registration drive, and a subset of those who registered on time for comparison (see Appendix B for further details on the sampling procedure). The litmus test then asked employees seven simple questions about their job, which any legitimate employee should be able to answer easily: (1) the name of the school in which they work; (2) the district in which their school is located; (3) the employee's rank and (4) position; (5) the principal's and (6) headmaster's names; and (7) the total number of employees working in the school.

The basic prediction is that real employees should know more facts about their school than stand-ins who have spent less time in the school. Therefore, we expect those who registered on time to score better than those who registered late, who, in turn, should score better than those who never registered.

As summarized in Figure 3, employees who never registered were less capable of answering these questions than employees who registered. For instance, the first two columns show the share of employees who did not answer the phone or were unavailable to respond to the survey. While only 21% of registered employees were not available to take the test, this share is 36% for employees who never registered ($p < 0.01$). In the remaining columns, we observe that employees who didn't register are 7.8 percentage points (s.e.= 2 pp) less likely to answer correctly all seven questions, 11.4 percentage points (s.e. = 2 pp) less likely to get a score of six, and 23.1 percentage points (s.e. = 3 pp) more likely to get a score of three or less.

How many stand-ins registered? An ideal litmus test would perfectly discriminate between true employees and stand-ins. Such a test requires a set of questions that: i) would be easy enough for all true employees to answer correctly, ii) would be non-trivial, so that stand-ins could not answer correctly, and iii) could be graded using official records. While our litmus test was not perfect (true employees can make mistakes and stand-ins can provide correct answers), it is nevertheless informative. Recognizing these limitations, Appendix Section B uses litmus scores to estimate bounds on the share of stand-ins. The idea is simple:

the lower the scores of registered employees are with respect to the scores of known true employees, the higher the number of stand-ins must be.[12]

We find that estimates of stand-ins range from 5.8% to 18.1% of registered employees (5.6% to 17.6% of all employees), depending on how strictly we mark the litmus test. Adding these potential stand-ins to the 2.8% of unregistered employees, we estimate that up to 20.4% of all employees might be ghosts. If they were removed from the payroll, the government would save AFN 26 million (USD 0.34 million) per month in the early registration zones. In Appendix D, we discuss the potential fiscal cost savings that a program like this might create if a ministry had the political will and capability to remove ghost workers.

Qualitative Evidence: We collaborated with Integrity Watch Afghanistan – one of Afghanistan's premier NGOs focused on governance – to conduct semi-structured interviews with 116 key informants with domain knowledge of public education in the three experiment provinces. This sample includes seven local elites, 16 teachers, 15 local education officials, 40 provincial and district officials, 11 senior officials in Kabul, 12 journalists, five ex-government officials, and 10 NGO workers.

We asked these respondents, "For every 100 employees on the MOE payroll, how many would you guess are ghost workers?" This is a challenging quantity to estimate, and 40 of our 116 respondents indicate that they had no clear estimate. The remaining 76 on average estimate that 11.82% (median = 9.75%) of teachers are ghosts (standard deviation = 15.02pp). This estimate is within our estimated range. Journalists, NGO workers, teachers, and provincial and district officials provide estimates very near to this average.[13] By contrast, senior officials provide estimates around 6%.[14] Stakeholders in the field appear to have more accurate beliefs about the share of ghost workers than senior officials in Kabul.

[12]Our exercise requires two assumptions. First, stand-ins would score zero in the litmus test. This implies that our estimates are lower bounds for the true number of stand-ins: we attribute low scores among registered employees to few stand-ins with zero scores, while they could be due to a higher number of stand-ins with low but positive scores. Second, the average score of all true employees is the same as the average score of true employees who could be verified through the registration process and unannounced visits to schools.

[13]The specific average estimates for each group are as follows: journalists 13.94%; NGO workers 13.75%; teachers 12.31%, and provincial and district officials estimate 13.36%.

[14]Specifically, MoE officials in Kabul estimate 6.81% on average and other senior officials estimate 5.5%.

4 The Impact of Mobile Salary Payments

The second part of the reform transitioned teacher salary payments from a hybrid bank and trusted agent system into an integrated system based on mobile money. We begin by assessing the extent to which the employees themselves supported these reforms, as that was critical to their long-term success (Section 4.1). We then show how the reform impacted leakage (i.e., the discrepancy between what was disbursed by the government and what was received by employees) and salary delays (Section 4.2). Given the debate regarding where to prioritize institutional reform, we pay special attention to differences in effects between rural and urban settings (Section 4.3) and attempt to disentangle the role of insurgent versus government control of territory (Section 4.3.1). Our final two sets of results show how the reform affected student learning (Section 4.4) and overall adoption and use of mobile money (Section 4.5).

4.1 Employee Satisfaction and Support for the Reform

For the reform to survive, it needed to gain the support of employees. As can be seen in the first panel of Figure 4, prior to implementation, support for the reform was very high in both the EE group (91%) and in the control group (94%).[15] Since our focus in this section is on the randomized timing of mobile salary payments, we simplify the exposition by omitting the ED group from Figure 4.[16] Correspondingly, for the rest of the section, we will refer to the EE group as the treatment group.

The second column of Figure 4 indicates that support fell precipitously after one year, both in the control (to 42%) and in the EE group (to 50%), with a statistically significant difference between the groups (8.1pp; s.e. = 4pp; $p = 0.04$). After two years, support rebounded to 64% and to 75%, respectively, with the difference between groups remaining significant (10.4pp; s.e. = 0.03; $p < 0.01$). As we discuss in the following sections, this

[15]Specifically, the figure indicates the share of surveyed employees agreeing with the statement, "If you were asked to vote on whether the entire ministry should switch to mobile salary payments, would you be in favor or against such a change? Please note this is a hypothetical question and you will not actually be asked to vote on this decision." The government intended this three-province pilot to be a test run for nationwide scaling, and considered responses to this question when considering the value of the reform. Employees were aware that their responses could influence whether the program continued, but it was made clear to employees that the question was hypothetical. Appendix Table A.7 indicates that there is not differential response rates by treatment status for any of our main outcomes.

[16]We show in Appendix Table A.8 that registration *per se* did not impact any outcomes related to the quality of salary payment. Appendix Table A.9 reports corresponding regression estimates for the full set of outcomes specified in our pre-analysis plan at both endline surveys and includes the EE and ED groups. Appendix Table A.10 shows that results change little if strata fixed effects are included.

sudden decrease in support, and then the subsequent and gradual rebound, likely reflects the substantial teething problems faced when the reforms were first implemented, but which were eventually resolved. In particular, the mobile network operator did not have enough capacity to quickly cash out digital payments, especially in rural areas.

We estimate the effect of the mobile salary reform more precisely in the first column of Table 1, which uses specification (1) to account for the imperfect adherence to treatment assignment (as discussed in Section 2.3). We observe that employees who were randomly induced to receive their salary via mobile money were roughly 18 percentage points (42 percent) more likely to support the reform after one year, and 26 percentage points (41 percent) more likely to support the reform after two years. In other words, first-hand experience of mobile salary payments increased support for the reform.

4.2 Impacts on Leakage and Delays: First vs. Last Mile

Why were employees who received mobile payments more supportive of the reform, and why did the support of both groups change so dramatically over time? Initially, we expected – as did policymakers – that the MSP reform would reduce last mile leakage of salary payments, i.e., the diversion of salaries by intermediaries between when the payment was released by the MoE and when it was received by the employee. Indeed, there was widespread concern that a large share of salaries, especially when distributed as cash by trusted agents, was being diverted (Special Inspector General for Afghanistan Reconstruction, 2022). In practice, however, we find very little evidence of this sort of leakage – both before and after the reform was enacted.

We measure leakage by employees' self-reported payments to someone to receive their salary. On average, control group employees report paying less than AFN 25 (USD 30 cents) to receive their salary in year one, which amounts to less than 1% of the average monthly salary. After two years, treatment reduced leakage essentially to zero (Table 1 column 5, panel B).

By contrast, delays in salary payments stand out as a major issue affecting employees – and one margin that the reform appears to have impacted. Salaries were subject to both first mile delays in the capital at the Education and Finance Ministries and Kabul Bank, as well as last mile delays in getting the final payment to the employee.[17] The general trends

[17]In addition to the slow government process of approving salaries, government salaries provide a key source of liquidity for Kabul Bank, providing incentives to move slowly. After salaries are approved and deposited in employees' bank accounts, they either need to be withdrawn (for employees paid by banks) or

are evident in the middle panel of Figure 4. Before the reform, 54.3% of employees in the control group and 62.6% of employees in the treatment group experienced delays, with a median delay of 12 days. In the subsequent two years of the reform, delays were substantially reduced for all employees – including in the control group. After one year, 40.6% of control employees and 66.7% of treated employees experienced delays; by the second year, delays in both groups came down to about 30%. Changes in the control group do not appear to be a direct spatial spillover from treatment zones.[18] Instead, the reason that delays went down and satisfaction went up in the controls appears to be due to system-wide improvements caused by the reform.

In particular, our data suggest that the reductions in delays – for both the treatment and control groups – were due to improvements in the first mile of salary delivery. In our surveys, we separately asked employees about their experience of first and last mile delays – these results are reported in Figure 5. In both the treatment and control groups, first mile delays were the major source of concern, with 41.7% (control) and 45.6% (treatment) of respondents reporting such delays at baseline. By contrast only 5.9% (control) and 10.1% (treatment) reported last mile delays.

An important consequence of the reform was that it immediately reduced first mile delays for both groups, and particularly for the treatment group. This can be seen by comparing pairs of columns in Figure 5. At baseline, 45.7% of the treatment group experienced first mile salary delays, but this number fell to 13.7% by April 2019 (a reduction of 32 percentage points, with $p < 0.01$); in the control group, the share experiencing first-mile delays fell from 41.7% at baseline to 27.6% in April 2019 (a reduction of 13.9 percentage points, $p < 0.01$).

At the same time, the reforms immediately *increased* last mile delays in the treated group, from a baseline of 10.1% to 34.9% in April 2019 (Figure 5). There was no such impact on the control group, presumably because their last mile delivery system was not directly impacted by the reform. Thus, we observe that treatment caused an increase in total delays of 40.1 percentage points in the first year (column 3 of Table 1). However, by the end of the second year, the gap in delays between treatment and control groups was no longer statistically significant.

withdrawn and physically transported as cash (for employees paid by trusted agents).

[18] Appendix Section C provides additional evidence that these effects represent system-wide changes rather than spatial treatment externalities.

As we discuss in more detail in the following section, the immediate increase in last mile delays in just the treatment group – and subsequent reduction in overall delays in the second year – appears to reflect 'teething problems' of the program, in which it took the mobile operator several months to establish an effective infrastructure to efficiently distribute mobile payments to employees. For instance, using administrative data from the mobile phone operator, we observe that there were fewer active mobile money agents when the program launched, and many of those agents had very little prior experience (Appendix Figure A.3).

The general patterns observed with salary delays are also reflected in employee responses about the time it took them to travel to receive their salary. As can be seen in the bottom panel of Figure 4, the reform increased travel time in the treatment group in the first year, but after two years travel times were similar in the two groups. This result is also evident in Table 1: treatment increased travel times by 43.2 minutes (panel A, column 3) in the first year, but after two years, the treatment effect was not statistically significant.

4.3 Heterogeneity Between Urban and Rural Areas

The results in the previous section indicate that the reforms reduced first-mile delays across the board (i.e., for both treated and control employees), but that the randomized treatment increased last mile delays and travel time during the first year of the program. Qualitative reports on the ground suggest that this pattern was driven largely by difficulties employees had – particularly in rural areas – accessing mobile money agents who could cash out their mobile payment. These teething problems relate closely to a central policy consideration in the design of the program.

There was debate among government stakeholders over whether the reform should begin in the cities or in the countryside, in part given the fragile security situation in rural areas.[19] In the end, the government selected Kandahar, Nangarhar, and Parwan as the three provinces for the experiment precisely because they contained major government-controlled cities (Kandahar City, Jalalabad, and peri-urban areas of Kabul) and rural areas actively contested by the Taliban.[20] Urban areas also had better mobile network coverage and more

[19]While counterinsurgency doctrine recommends consolidating state control in cities before branching into the countryside (Krepinevich, 2005; U.S. Army, 2006), some policymakers advocated prioritizing reforms in contested and insurgent-controlled areas. With mobile salary payments, the Ministry of Finance argued that state employees in the rich and relatively secure provincial capitals were already served by banks and that salary issues and corruption were more severe in rural areas. However, rural areas posed challenges for mobile salaries including limited economies of scale, mobile agent presence, and mobile network coverage.

[20]We define urban zones as those with a population density of greater than 300 inhabitants per squared kilometer.

mobile money agents.

The large scale of our experiment makes it possible to separately test the effectiveness of the reform in rural and urban areas. These results are reported in Table 2. We find that in the first year, the reform drove increases in support for the reform in urban areas, with no clear impact in rural areas (panel A, column 1). Delays and travel times increased everywhere (columns 3 and 4), but both increases were significantly larger in rural areas. However, by year 2, support had increased uniformly across the entire geography of the experiment (panel B, column 1); delays remained elevated in rural areas but decreased by 26.5 percentage points in urban areas, and travel time was no longer statistically different than in the control group. While salary leakage was not a major issue, the reduction in leakage caused by the reform was largest in rural areas (column 5).

The heterogeneity in impacts between urban and rural areas, and the shifting dynamics between the first and second year, indicate that while it took time to work out the mechanics of delivering salaries – especially in rural areas – the reform eventually reduced delays, particularly in urban areas where the logistics of delivery were considerably easier.

4.3.1 Heterogeneity Between Government- and Taliban-Controlled Areas

In the appendix, we examine the effectiveness of the reform in government-controlled and insurgent-controlled areas. To conduct this analysis, we adopt the methodology and data outlined in Wright (2023).[21] We gauge government control by assessing whether non-local survey enumerators are permitted to conduct surveys without seeking permission from local commanders. According to this criterion, all urban areas are classified as being under government control. However, we can differentiate rural areas based on whether they were under government control or contested/controlled by the Taliban. The results are presented in Appendix Table A.11. Overall, our findings indicate that the reform did not demonstrate superior efficacy in rural areas under government control compared to those under insurgent control. Conversely, when focusing exclusively on areas under government control, the reform exhibited a more positive impact in urban centers.

4.4 Impacts on Student Learning

Beyond the immediate goals of reducing leakage and improving the salary experience of teachers, an important long-term goal of the reform was to improve student learning. As

[21] We thank Austin Wright for sharing these data.

such, this experiment provides an opportunity to study whether improving basic ministerial functions can translate to better service delivery. This section assesses impacts on learning using an assessment conducted in May 2019, one year after registration and six months after mobile payments started.[22] We originally planned to conduct a second assessment in 2020, but this became impossible due to restrictions on fieldwork from the COVID-19 pandemic.

Results in Table 3 indicate that the reform had positive impacts on student learning outcomes in cities, as measured in our educational assessments. The table provides ToT estimates, where we define a student as treated if at least 50% of the teachers in the school that they attend received a payment via mobile money by April 2019 (the month prior to the learning assessment). We instrument for treatment with the school's randomly assigned treatment status (i.e., a dummy variable indicating whether the school was in the EE or ED arm of the experiment). ToT estimates from the urban sample (columns 2, 6, and 8) indicate a math score increase of about 0.228σ ($p = 0.07$), and a combined score increase of 0.192σ ($p = 0.09$). Estimates in the full sample (columns 1, 4, and 7) are positive and estimates in rural zones (columns 3, 6, and 9) are negative, though all are statistically insignificant. The corresponding ITT estimates, which directly regress test scores on the school's treatment assignment, are reported in Appendix Table A.12.

To help benchmark the magnitude of these impacts, Appendix Table A.13 reports estimates of the relationship between our learning measures and years of schooling. An additional year of schooling is associated with math score increases of 0.28σ (column 2), reading score increases of 0.24σ (column 4), and combined score increases of 0.26σ (column 6).[23] The null of no association between years of schooling and test outcomes is rejected at the 1% level in all tests, increasing our confidence in the learning assessments.[24] Using our ToT estimates from the urban sample, the additional learning created by the reform is equivalent to $0.23/0.28 = 0.82$ years of schooling in math and $0.19/0.26 = 0.73$ years of schooling for the combined score.

Our data do not allow us to speak precisely about the mechanism linking the reform to

[22]Since Afghanistan does not have any reliable and systematic data on educational outcomes, we collected these data ourselves. The educational assessment protocols we implemented are described in Section 2.5. We conducted assessments with a representative sample of 1,101 students aged between 6 and 9, but focus our analysis on the 939 students who had attended school at least once in the three months prior to our assessment. We thank Dana Burde and Joel Middleton for sharing with us the learning assessment methodologies they validated for use in Burde and Linden (2013), Burde, Middleton and Samii (2017), and Burde, Middleton and Wahl (2015).

[23]Our preferred specifications control for demographics and district fixed effects, though estimates are similar to those obtained from simple bivariate regressions.

[24]Estimates are also positively associated with parental income and socio-economic status.

improved learning outcomes. The pattern of results, however, mirrors those discussed earlier in this section. The reform created an immediate relative jump in enthusiasm for the future prospects for the reform in the urban sample, despite the early delays and increases in travel costs. The results reported in Appendix Tables A.14 and A.15 do not indicate an increase in teacher attendance, so it is unlikely this is the mechanism.[25] While we are thus left to speculate, conversations with teachers suggest that employees were optimistic that the system might lead to reduced delays and an improved salary experience; it is possible that this in turn might have led teachers to increase their effort in the classroom. Taken together, the education impacts we report point to the possibility that, at least in dysfunctional states, improving basic administrative processes can translate through to improved service delivery.

4.5 Impacts on Mobile Money Use

A final objective articulated by the architects of the reform was to encourage broader adoption of mobile money and digital financial services.[26] This was because, prior to the reform, employees did not actively make use of formal financial services: In the baseline survey, 67% of employees reported having a bank account at baseline, but only 5% reported saving money in their bank accounts, 12% reported having used their bank account to send money, and 3% reported that they had an outstanding bank loan.

The survey evidence in Table 1 column (6) – discussed above – indicates the reform increased peer-to-peer transfers by 29.3pp from a base of 6.3%. To further understand if and how the salary reforms impacted how employees used their mobile wallets, we draw on rich administrative data from the mobile phone company. We focus our analysis on estimating how the length of exposure to mobile salary payments increased the use of other mobile financial services, using the following dosage-response specification:[27]

$$Y_{izd} = \alpha + \beta\ Months_{izd} + \mu_d + \varepsilon_{izd} \tag{2}$$

[25] We observe teacher attendance through unannounced audits conducted at baseline (May 2018), midline (November 2018), and endline (April 2019). We restrict the sample to schools that are audited and to teachers that were on the payroll prior to the experiment. At baseline 40.8% of teachers in the control group are present overall and 35.6% of urban and 56.4% of rural teachers are present.

[26] The Afghan government hoped to catalyze a digital financial ecosystem by providing mobile money accounts to a large number of public servants. Indeed, the initial request for proposals to contract a mobile network operator specified that the winner should provide an inter-operable service. Related to this, the government also required bidders to recruit 'white label' agents, who could convert mobile money on any mobile money platform to cash.

[27] Less than 0.1% of registered employees used their mobile wallet before receiving their first mobile salary payment, clearly indicating that mobile money usage was not significantly impacted by the provision of a mobile-money-enabled SIM card at registration.

In the above, $Months_{izd}$ indicates the number of months that employee i from registration zone z in district d had received mobile payments at the time of the last recorded transaction. To facilitate interpretation, we normalize the outcomes Y_{izd} by the total monthly salary received in the month prior to the start of the reform. We estimate (2) using transaction-level data from the roughly 15,000 employees that were paid via mobile money from October 2018 to February 2020. Our results, in Table 4 show the effect (β) of an additional month of receiving mobile salaries, both using an OLS specification (Panel A) and an instrumental variables specification (Panel B) in which we instrument for $Months_{izd}$ using the randomly assigned treatment dummies.[28]

The first column of Table 4 shows the impact of the reform on mobile money deposits, which we consider to be a particularly important measure of adoption since it requires employees to add money to their wallet beyond what they receive for their salaries. In the first month of use, employees deposit a statistically insignificant 0.2 percentage points of their salary (shown in the '1-Month Mean' row). The ToT estimates (Panel B) indicate this increases by 0.4 percentage points for every month they are on the platform. The median time on the platform observed in our data is 7 months; this implies an increase in deposits of 2.8 percentage points more every month. Transfers to others via mobile money (column 2) go up by a similar amount (though this is not significant after instrumenting). Airtime purchases (column 3) and transfers of airtime to others (a basic version of a mobile money transfer that can be done to subscribers without a mobile money wallet) also increases considerably (column 4). Balances go up by 1.2 percentage points, or by 8.4 percentage points at the median (though this effect is smaller and not significant after instrumenting).[29]

5 Conclusion

Effectively selecting and incentivizing state personnel is essential for building state capacity (Finan, Olken and Pande, 2017). However, maintaining an efficient government payroll is particularly challenging in fragile states, as they often lack the prerequisite systems needed to reliably track and compensate civil servants. While a growing body of work has linked the capability of state institutions to economic development (Besley et al., 2022), evidence in fragile states is still scarce, with less than 1% of all registered RCTs and top five economics journal articles over the last 23 years focusing on the world's 15 most fragile states.

[28] The corresponding first stage is $\#Months_{izd} = \theta + \varphi_{EE}EE_z + \varphi_{ED}ED_z + \eta_d + \epsilon_{izd}$.

[29] Table A.16 reports results from a similar specification that interacts $Months_{izd}$ with whether the sample is urban or rural. The ToT estimates (Panel B) indicate that the financial inclusion benefits of the MSP reform are generally not statistically distinguishable between urban or rural areas.

This project, implemented at scale by a government in the midst of an active war, demonstrates the potential for technology-centered reforms to improve state capacity. Indeed, after starting with three provinces, the government decided to scale up the reform to eight additional provinces. Notably, the reform registered 24,101 unique employees and, by the end of the project, most employees supported the expansion of the reform. What broader lessons can be learned from this specific set of reforms in Afghanistan?

First, we found that – even with robust monitoring and a careful RCT-based evaluation – it took years to be able to discern the positive impacts of the reform. This approach contrasts with the demand for 'quick impact' projects contracted to work around and not through the state (Dercon, 2022; Pande and Page, 2018; Commission on State Fragility, Growth, and Development, 2018). U.S. National Security Advisor H.R. McMaster famously remarked that the war in Afghanistan was not a 20-year war but rather "a one-year war fought 20 times over."[30] Many international efforts to support fragile states shift strategy annually, reflecting personnel rotations and domestic political cycles. If the reform studied here – focused on building a basic but essential capability – had been called off after only one year, it would have appeared a failure. The reason this reform showed relative success was because it was initiated and sustained by Afghan leaders as a top, multi-year, priority. Plans to build basic elements of state capacity need builders who can tolerate periods of failure before seeing success.

Second, we observe that investment in basic state capabilities can yield broader benefits. In this project, for instance, we saw downstream impacts of the reforms on student learning outcomes. We also found that providing teachers with digital wallets created new opportunities to save and send money – potentially creating new avenues to improve risk sharing and reduce poverty (Jack and Suri, 2014; Suri and Jack, 2016). In addition, a new digital system based partly on lessons from this study now provides a channel for humanitarian assistance to vulnerable women (Callen et al., 2023). Such systems also open doors for behavioral interventions to boost savings (Blumenstock, Callen and Ghani, 2018), and creates a pathway to financial deepening.

Finally, our study highlights how the same reform can have different impacts in different places and at different timescales. In cities, we consistently found evidence that the reform improved the salary experience of employees, and had downstream impacts on education

[30] See, for example, (McMaster, 2021).

after one year. In contrast, in the countryside, the salary payment experience of teachers only improved after two years. We do not find evidence that these differences are driven by government control – safer rural districts do not fare better than rural districts contested by insurgents. Of course, cities have a broader base of potential users and agents, reinforcing the network effects inherent in a digital payments platform. Cities also benefit from better mobile network coverage and a potentially more tech-savvy population. Further research is needed to understand exactly how these institutional and market conditions interact to support the efficacy of state-building reforms and the deepening of financial inclusion.

This project challenges narratives that dismiss fragile states as hopelessly corrupt, inept, or mired in historical complexity. Instead, we observe that the efficacy of state-building reforms in fragile contexts depends on prioritization, patience, and context-sensitivity. We also find value in approaches that go through the state, complemented by rigorous evaluation, focused on meeting long-term goals rather than political demands for quick impacts.

References

Acemoglu, Daron, and James A Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty.* Crown.

Aker, Jenny C., Rachid Boumnijel, Amanda McClelland, and Niall Tierney. 2016. "Payment Mechanisms and Antipoverty Programs: Evidence from a Mobile Money Cash Transfer Experiment in Niger." *Economic Development and Cultural Change*, 65(1): 1–37.

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee. 2020. "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services." *American Economic Review*, 110(5): 1355 – 94.

Balán, Pablo, Augustin Bergeron, Gabriel Tourek, and Jonathan L. Weigel. 2022. "Local Elites as State Capacity: How Local Information Increases Tax Compliance in the D.R. Congo." *American Economic Review*, 112(3): 1 – 36.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti. 2009. "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment." *American Economic Review*, 99(4): 1278 – 1308.

Bandiera, Oriana, Michael Callen, Katherine E. Casey, Eliana La Ferrara, Camille Landais, and Matthieu Teachout. 2019. "International Growth Centre Evidence Paper - State Effectiveness."

Banerjee, Abhijit, Esther Duflo, Clement Imbert, Santhosh Mathew, and Rohini Pande. 2020. "E-governance, Accountability, and leakage in Public Programs: Experimental Evidence from a Financial Management Reform in India." *American Economic Journal: Applied Economics*, 12(4): 39–72.

Bates, Robert H. 2001. *Prosperity and violence: the political economy of development.* Norton.

Besley, Timothy, and Torsten Persson. 2009. "The origins of state capacity: Property rights, taxation, and politics." *American economic review*, 99(4): 1218–44.

Besley, Timothy, and Torsten Persson. 2011. *Pillars of prosperity.* Princeton University Press.

Besley, Timothy, Robin Burgess, Adnan Khan, and Guo Xu. 2022. "Bureaucracy and development." *Annual Review of Economics*, 14: 397–424.

Blumenstock, J.E., N. Eagle, and M. Fafchamps. 2016. "Airtime Transfers and Mobile Communications: Evidence in the Aftermath of Natural Disasters." *Journal of Development Economics*, 120: 157–181.

Blumenstock, Joshua Evan, Michael Callen, Tarek Ghani, and Lucas Koepke. 2015. "Promises and Pitfalls of Mobile Money in Afghanistan: Evidence from a Randomized Control Trial." *ICTD '15*. Singapore:ACM.

Blumenstock, Joshua, Michael Callen, and Tarek Ghani. 2018. "Why do defaults affect behavior? Experimental evidence from Afghanistan." *American Economic Review*, 108(10): 2868–2901.

Blumenstock, Joshua, Michael Callen, Tarek Ghani, and Robert Gonzalez. 2022. "Violence and Financial Decisions: Evidence from Mobile Money in Afghanistan." *Review of Economics and Statistics*.

Bossuroy, Thomas, Clara Delavallade, and Vincent Pons. 2019. "Biometric Tracking, Healthcare Provision, and Data Quality: Experimental Evidence from Tuberculosis Control." *NBER Working Paper*.

Burde, Dana, and Leigh Linden. 2013. "Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools." *American Economic Journal: Applied Economics*, 5(3): 27 – 40.

Burde, Dana, Joel Middleton, and Cyrus Samii. 2017. "Assessment of Learning Outcomes and Social Effects of Community-Based Education: A Randomized Field Experiment in Afghanistan. ALSE Phase II Endline: Learning Assessment Survey."

Burde, Dana, Joel Middleton, and Rachel Wahl. 2015. "Islamic Studies as Early Childhood Education in Countries Affected by Conflict: The Role of Mosque Shools in Remote Afghan Villages." *International Journal of Educational Development*, 41: 70–79.

Callen, Michael, and James D Long. 2015. "Institutional Corruption and Election Fraud: Evidence From a Field Experiment in Afghanistan." *The American Economic Review*, 105(1): 354–381.

Callen, Michael, Clark Gibson, Danielle Jung, and James Long. 2016. "Improving Electoral Integrity with Information and Communications Technology." *Journal of Experimental Political Science*, 3(1): 4–17.

Callen, Michael, Miguel Fajardo-Steinhauser, Michael Findley, and Tarek Ghani. 2023. "Digital Aid Cost-Effectively Addresses the Humanitarian Needs of Vulnerable Women."

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perpsectives*, 20(1).

Commission on State Fragility, Growth, and Development. 2018. *Escaping the Fragility Trap.* International Growth Center.

Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service*." *The Quarterly Journal of Economics*, 128(3): 1169–1218.

Dercon, Stefan. 2022. *Gambling on Development: Why Some Countries Win and Others Lose.* C Hurst and Co.

Dhaliwal, Iqbal, and Rema Hanna. 2017. "Deal with the Devil: The Successes and Limitations of Bureaucratic Reform in India." *Journal of Development Economics*.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *The American Economic Review*, 102(4): 1241–1278.

Fearon, James D, and David D Laitin. 2004. "Neotrusteeship and the problem of weak states." *International security*, 28(4): 5–43.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. 2017. "Chapter 6 - The Personnel Economics of the Developing State." In *Handbook of Economic Field Experiments*. Vol. 2, , ed. Abhijit Vinayak Banerjee and Esther Duflo, 467–514. North-Holland.

Finan, Frederico, Ernest Dal Bó, and Martín Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *Quarterly Journal of Economics*, 128(3): 1169 – 1218.

Giné, Xavier, Jessica Goldberg, and Dean Yang. 2012. "Credit Market Consequences of Improved Personal Identification: Field Experimental Evidence from Malawi." *American Economic Review*, 102(6): 2923–2954.

Jack, William, and Tavneet Suri. 2014. "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution." *American Economic Review*, 104(1): 183–223.

Joppe, De Ree, Karthik Muralidharan, and Halsey Rogers Menno Pradhan. 2017. "Double for Nothing? Experimental Evidence on Unconditional Teacher Salary Increase in Indonesia." *Quarterly Journal of Economics*, 133(2): 993 – 1039.

Khwaja, Asim, Adnan Khan, and Benjamin A. Olken. 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors." *Quarterly Journal of Economics*, 131(1).

Khwaja, Asim, Adnan Khan, and Benjamin A. Olken. 2019. "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings." *American Economic Review*, 109(1).

Krepinevich, Andrew. 2005. "How to Win in Iraq." *Foreign Affairs*, 84(5): 87 – 104.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania." *Quarterly Journal of Economics*, 134(3): 1627 – 1673.

McMaster, H.R. 2021. "Afghanistan is America's Longest War – It's Time for the Delusion About It To End."

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in The Presence of Treatment Externalities." *Econometrica*, 72(1): 159–217.

Muralidharan, Karthik, and Paul Niehaus. 2017. "Experimentation at Scale." 31: 103–124.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experiment Evidence from India." *Journal of Political Economy*, 119(1): 39 – 77.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2016. "Building State Capacity: Evidence from Biometric Smartcards in India." *The American Economic Review*, 106(10): 2895–2929.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. Forthcoming. "Identity Verification Standards in Welfare Programs: Experimental Evidence from India." *Review of Economics and Statistics*.

OECD. 2022. *States of Fragility 2022.*

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

Olken, Benjamin A., and Rohini Pande. 2012. "Corruption in Developing Countries." *Annual Review of Economics*, 4: 479 – 509.

Pande, Rohini, and Lucy Page. 2018. "Ending Global Poverty: Why Money Isn't Enough." *Journal of Economic Perspectives*, 32(4): 173–200.

Reinnika, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence From A Central Government Transfer Program In Uganda." *Quarterly Journal of Economics*, 119(2): 679–705.

Sanchez de la Sierra, Raul. 2020. "On the Origins of the State: Stationary Bandits and Taxation in Eastern Congo." *Journal of Political Economy*, 128(1).

Special Inspector General for Afghanistan Reconstruction. 2016. "SIGAR 17-12-SP Schools in Herat Province: Observations from Site Visits at 25 Schools."

Special Inspector General for Afghanistan Reconstruction. 2017. "SIGAR 17-32-SP Schools in Balkh Province: Observations from Site Visits at 26 Schools."

Special Inspector General for Afghanistan Reconstruction. 2022. "SIGAR 22-34 Audit Report. Department of Defense's Salary Payments to the Afghan Ministry of Defense: DOD Did Not Use the Afghan Personnel and Pay System as Intended and Internal Control Weaknesses Raise Questions about the Accuracy of USD 232 Million in Salary Payments."

Suri, Tavneet, and William Jack. 2016. "The Long-Run Poverty and Gender Impacts of Mobile Money." *Science*, 354(6317): 1288–1292.

Tilly, Charles. 1985. "War Making and State Making as Organized Crime." In *Bringing the State Back In.* , ed. Dietrich Rueschemeyer Evans, Peter B. and Theda Skocpol, 169 – 191. Cambridge University Press.

U.S. Army. 2006. *Field Manual 3-24: Counterinsurgency.* United States Army.

Weber, Max. 1919. "1946. Politics as a Vocation." *From Max Weber: Essays in Sociology*, 77–128.

Weigel, Jonathan L. 2020. "The Participation Dividend of Taxation: How Citizens in Congo Engage More with the State when It Tries to Tax Them." *Quarterly Journal of Economics*, 135(4): 1849 – 1903.

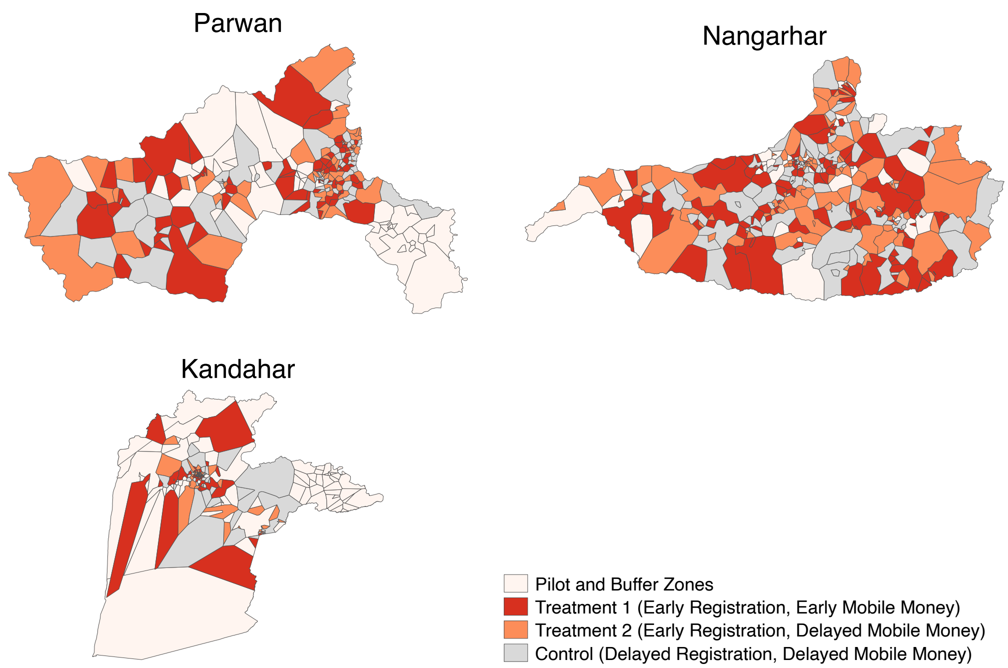
Weinstein, Jeremy M. 2005. "Autonomous recovery and international intervention in comparative perspective." *Available at SSRN 1114117.*

World Bank. 2019. "Afghanistan: Public Expenditure Update."

Wright, Austin. 2023. "The Economic Origins of Territorial Control."

Figures and Tables

Panel A: Studied Provinces of Afghanistan



Panel B: Registration Zones and Treatment Status

Figure 1: Map of Provinces and Registration Zones

Notes: Panel A plots the map of Afghanistan, with three studied provinces highlighted in red. Panel B plots registration zones colored according to the treatment status in studied provinces. Registration zones are designed to encompass approximately 50 teachers and are created by a spatial grouping algorithm. Areas that either had no cellular coverage or that were known to be completely under Taliban control were not considered for the reform.

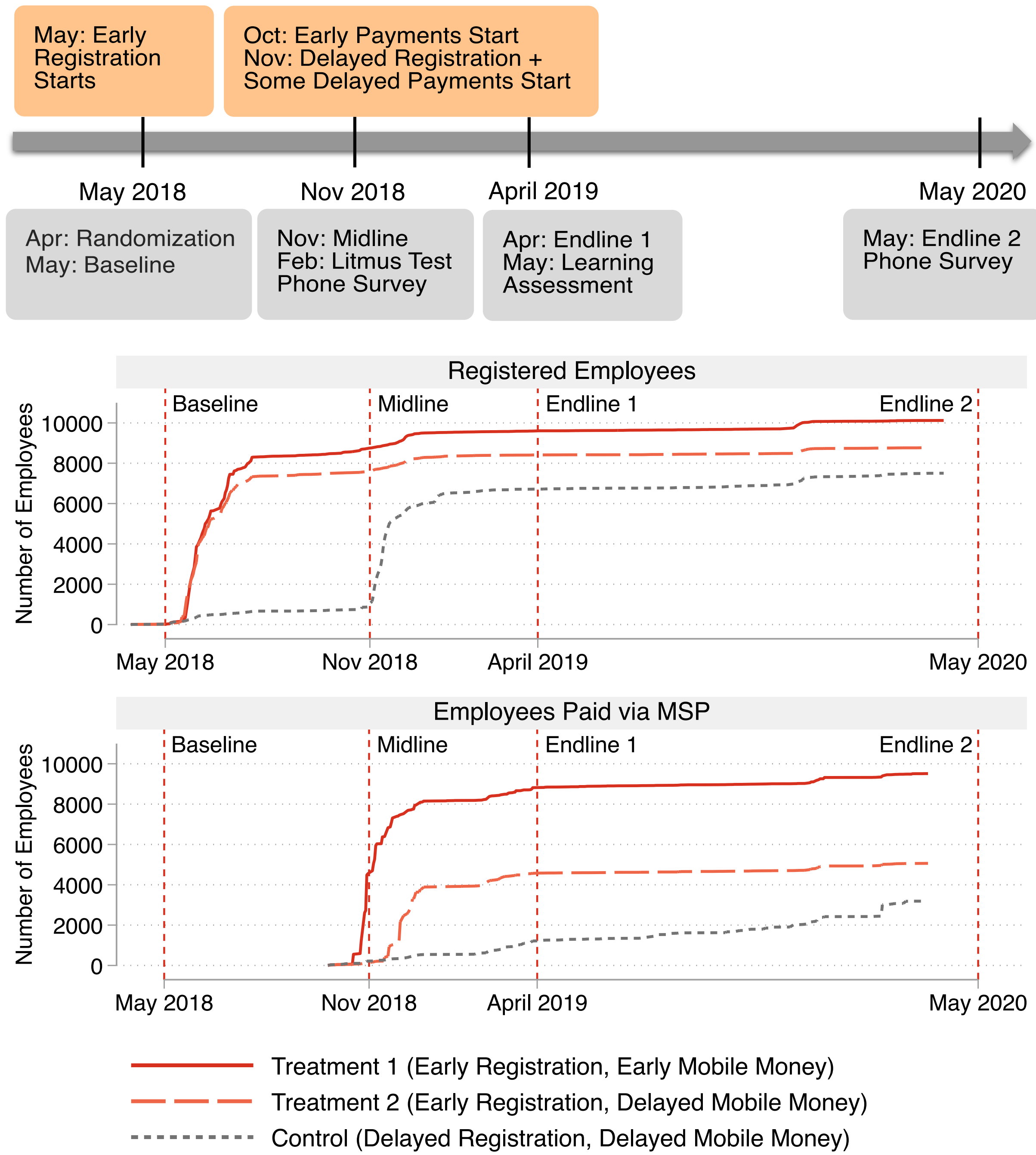


Figure 2: Project Implementation Timeline and Treatment Compliance

Notes: This figure plots the timeline of project implementation. The top panel presents the dates of the main milestones: the start of registration and payments in different treatment groups and administered surveys. The center panel presents the cumulative number of employees registered using administrative registration data. The bottom panel presents the cumulative number of employees paid via mobile salary payments using the administrative transaction data.

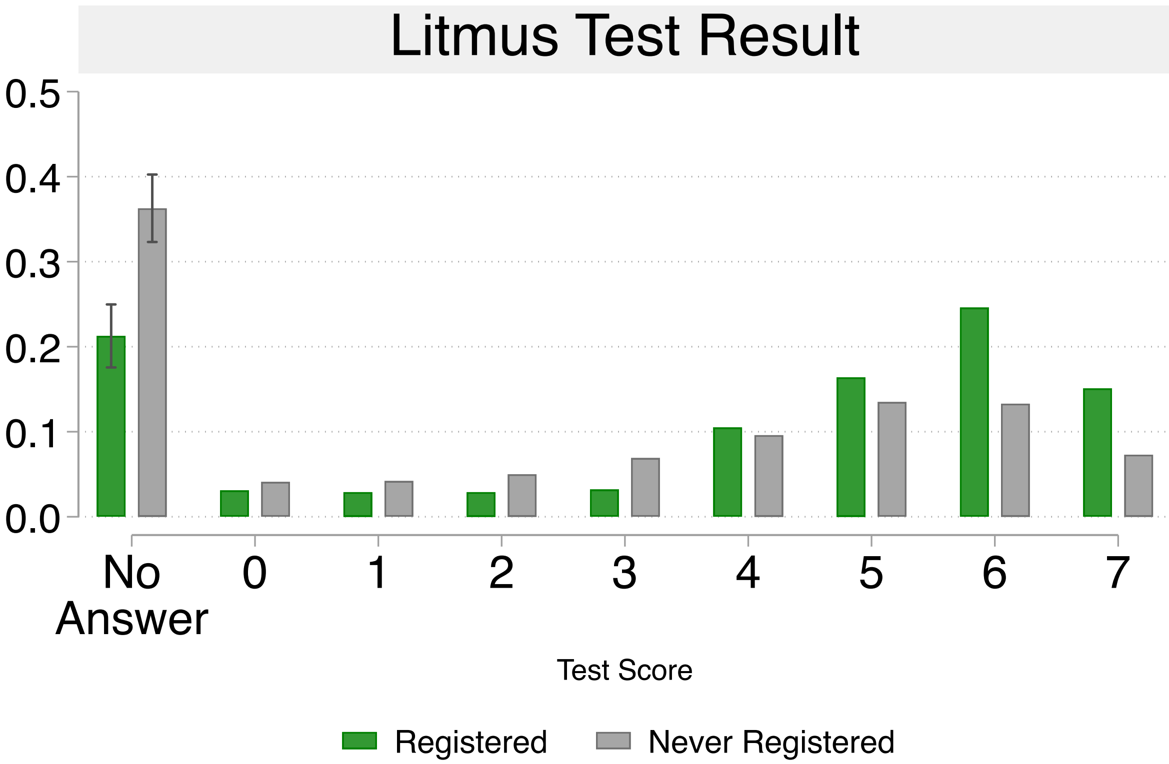


Figure 3: Litmus Test Results

Notes: This figure presents the results designed to confirm whether employees knew basic facts about the schools where they were assigned. The sample consists of 2,545 employees from Early Registration zones, selected for the Litmus Test in February of 2019. The litmus test score is defined as the total number of correct answers to the following seven questions: school district and name, employee's rank and position, principal's and headmaster's names, and the total number of employees. Employees were interviewed on the phone based on a phone number registry that pre-dated the experiment. No Answer corresponds to employees who did not answer the phone. 2.78% of the 14,184 employees in the 243 Early Registration zones who were paid for six months before the reform never registered.

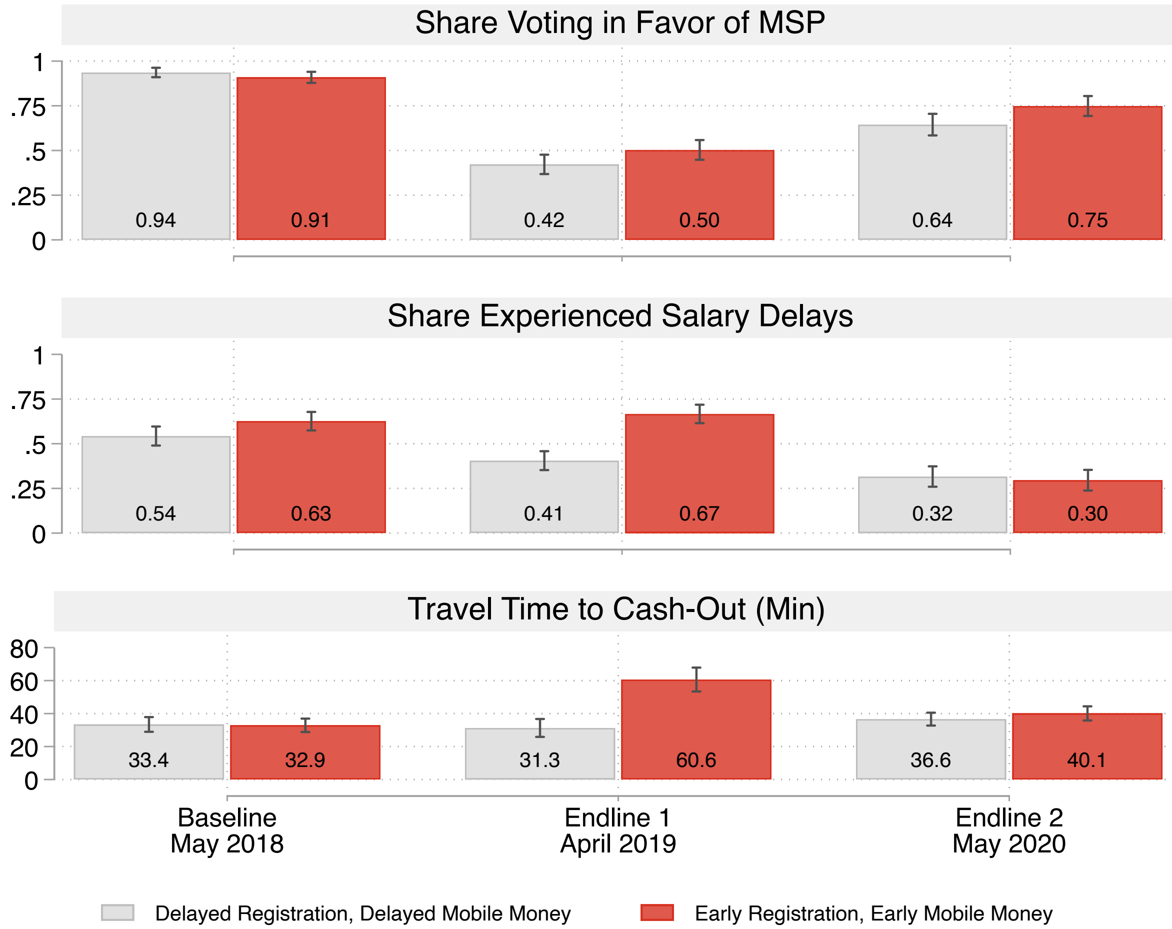


Figure 4: Effects on Payment Quality

Notes: This figure presents means of outcomes measuring payment quality at baseline (bars 1–2), endline 1 (bars 3–4), and endline 2 (bars 5–6). The sample consists of MoE employees who participated in the full baseline, endline 1, and endline 2 surveys, respectively, and responded to all questions. Appendix Table A.9 reports the corresponding treatment effects for both the EE and ED treatment arms estimated without stratum fixed effects for all outcomes specified in our pre-analysis plan, and Appendix Table A.10 shows corresponding estimates with stratum fixed effects.

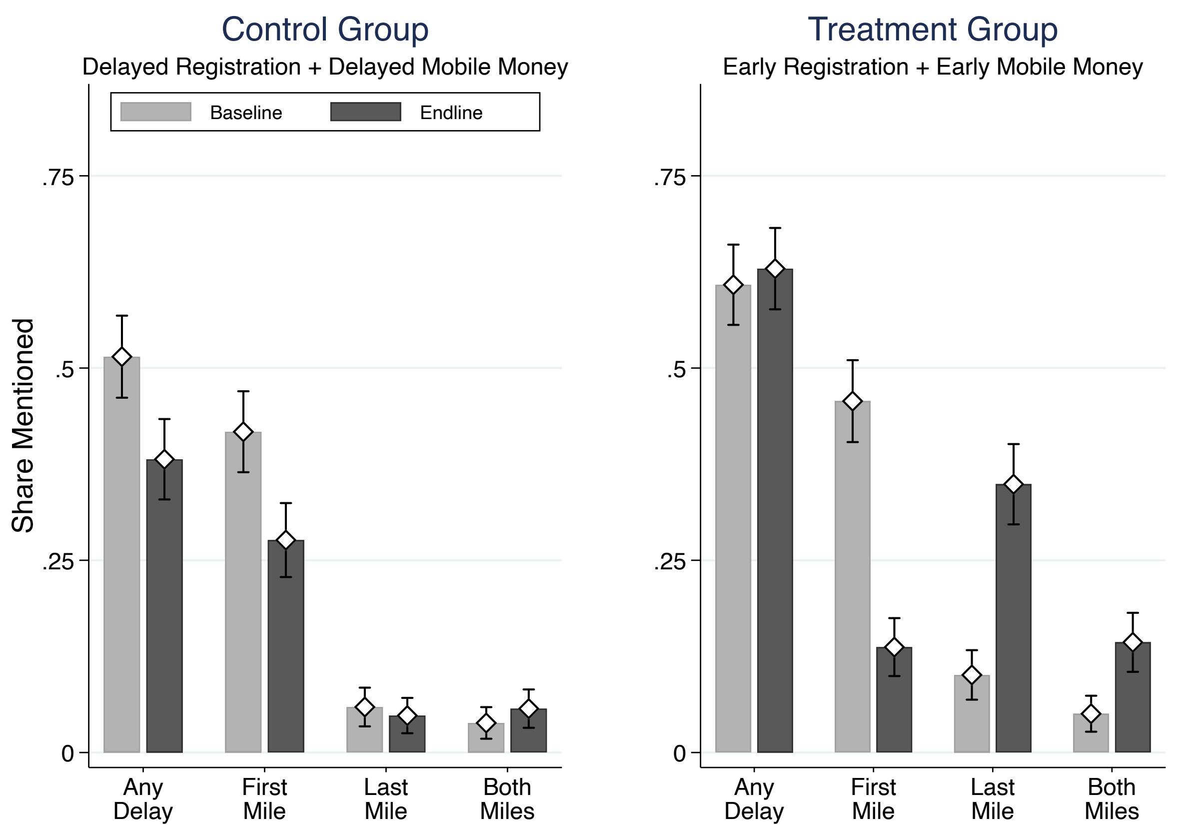


Figure 5: Separating First Mile and Last Mile Delays

Notes: This figure depicts the cause of delays. Respondents who reported receiving their salary late were asked who was responsible for the delay. First mile delays are those due to the MoE, the Ministry of Finance, or the Central Bank (Da Afghanistan Bank). Last mile delays are those due to New Kabul Bank, Motameds, the mobile network operator, or the mobile money agents. The causes of delay are comprehensive, such that the share reporting First Mile, Last Mile, and First and Last Mile sum to the total share reporting delays. The sample comprises 1,005 respondents interviewed at baseline in May 2018 and at endline 1 in April 2019. We did not include questions about the source of delays in our endline 2 survey conducted by phone in May 2020 due to Covid-19.

Table 1: Estimates of Impact on Employees' Payment Experience

	Vote in Favor of MSP (1)	Very Satisfied with Pay System (2)	Experienced Delay (3)	Travel Time to Cash-Out Salary (Min) (4)	Payment to Receive Transfer (5)	Conducted Mobile Money Transaction (6)
Panel A. Year 1 Outcomes						
Mobile Salary Payment (=1)	0.178***	–0.397***	0.401***	43.208***	–4.966	0.293***
	(0.067)	(0.076)	(0.073)	(8.496)	(5.002)	(0.048)
Control Mean	0.422	0.628	0.405	31.276	22.311	0.063
Observations	950	966	969	922	959	965
R squared	0.334	0.205	0.269	0.273	0.270	0.149
# Reg. Zones	350	352	352	344	352	352
Panel B. Year 2 Outcomes						
Mobile Salary Payment (=1)	0.263***	–0.001	–0.064	3.677	–17.563***	
	(0.080)	(0.079)	(0.081)	(5.943)	(3.849)	
Control Mean	0.645	0.727	0.316	36.628	12.479	
Observations	712	736	735	725	738	
R squared	0.269	0.132	0.117	0.090	0.249	
# Reg. Zones	322	328	328	324	327	

Notes: This table reports ToT estimates of impacts of mobile salary payments on payment quality and mobile money use. Corresponding ITT estimates are reported in Appendix Tables A.9 and A.10. Favor MSP is a dummy variable equal to one if the respondent indicates they would support scaling the reform across the Ministry of Education. Satisfied with Pay System is a dummy equal to one if the respondent indicates a very high level of satisfaction. Experienced Delay is a dummy variable equal to one if the respondent reports their salary being delayed. Travel Time is the time to convert the mobile money payment to cash in minutes. Payment to Receive measures what respondents report paying to receive their salary in Afghanis. For comparison, the average (net) salary in the sample is 8560 Afg. Money Transfer is a dummy variable equal to one if the respondent indicates making a mobile money transfer to someone else in the previous month. This outcome is not recorded in year 2 because the survey was abbreviated due to the pandemic. Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 2: Estimates of Impact on Employees' Payment Experience - Urban vs. Rural

	Vote in Favor of MSP (1)	Very Satisfied with Pay System (2)	Experienced Delay (3)	Travel Time to Cash-Out Salary (Min) (4)	Payment to Receive Transfer (5)	Conducted Mobile Money Transaction (6)
Panel A. Year 1 Outcomes						
β_1: MSP × Urban	0.306***	–0.274***	0.239**	33.788***	–6.855	0.299***
	(0.089)	(0.102)	(0.102)	(8.187)	(7.817)	(0.063)
β_2: MSP × Rural	–0.015	–0.580***	0.641***	55.686***	–2.262	0.285***
	(0.104)	(0.116)	(0.107)	(17.098)	(5.373)	(0.078)
Observations	950	966	969	922	959	965
R squared	0.013	0.032	0.071	0.080	0.005	0.066
Control Mean	0.422	0.628	0.405	31.276	22.311	0.063
p-value $\beta_1 = \beta_2$	0.020	0.049	0.007	0.249	0.629	0.890
# Reg. Zones	350	352	352	344	352	352
Panel B. Year 2 Outcomes						
β_1: MSP × Urban	0.236**	0.238**	–0.265**	–0.275	–9.878**	
	(0.115)	(0.105)	(0.116)	(7.140)	(4.286)	
β_2: MSP × Rural	0.296**	–0.321**	0.203*	9.090	–27.846***	
	(0.117)	(0.125)	(0.114)	(10.732)	(7.125)	
Observations	712	736	735	725	738	
R squared	0.087	–0.044	–0.031	0.005	0.059	
Control Mean	0.645	0.727	0.316	36.628	12.479	
p-value $\beta_1 = \beta_2$	0.716	0.001	0.004	0.468	0.031	
# Reg. Zones	322	328	328	324	327	

Notes: This table reports ToT estimates of impacts of mobile salary payments on payment quality and mobile money use. Favor MSP is a dummy variable equal to one if the respondent indicates they would support scaling the reform across the Ministry of Education. Satisfied with Pay System is a dummy equal to one of the respondent indicates a very high level of satisfaction. Experienced Delay is a dummy variable equal to one if the respondent reports their salary being delayed. Travel Time is the time to convert the mobile money payment to cash in minutes. Payment to Receive measures what respondents report paying to receive their salary in Afghanis. For comparison, the average (net) salary in the sample is 8560 Afg. Money Transfer is a dummy variable equal to one if the respondent indicates making a mobile money transfer to someone else in the previous month. This outcome is not recorded in year 2 because the survey was abbreviated due to the pandemic. Urban areas are those with a population density greater than 300 inhabitants per squared kilometer. Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table 3: Estimates of Impact on Student Learning

	Math Score			Reading Score			Combined Score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treated School (=1)	0.084	0.228*	-0.097	0.072	0.155	-0.019	0.078	0.192*	-0.058
	(0.096)	(0.124)	(0.143)	(0.095)	(0.122)	(0.151)	(0.086)	(0.114)	(0.126)
Untreated School Mean	0.07	0.04	0.10	0.05	0.09	0.00	0.06	0.07	0.05
Sample	Full	Urban	Rural	Full	Urban	Rural	Full	Urban	Rural
District FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓	✓	✓	✓	✓
R-squared	0.30	0.26	0.36	0.25	0.24	0.27	0.31	0.28	0.35
# Reg. Zones	362	201	161	362	201	161	362	201	161
# Students	939	516	423	939	516	423	939	516	423

Notes: This table reports the impacts of the Mobile Salary Payment Reform on student learning outcomes measured in the assessment described in Section 2.5. Scores are standardized using the control group mean and standard deviation. The sample comprises 939 students aged between 6 and 10 years who attended a public school eligible for the Mobile Salary Payments reform within the last three months. Treated School is a dummy variable equal to one for students who attend a school where at least 50% of teachers were paid using mobile money by April 2019, the month before the learning assessment. Covariates include years of education, years of age, a dummy variable equal to one if a student lives in a high socio-economic status household, and gender. High SES households are those that are on a maintained road with access to water and to electricity. 458 of the 939 students are high SES. 343 of the students are female. Standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

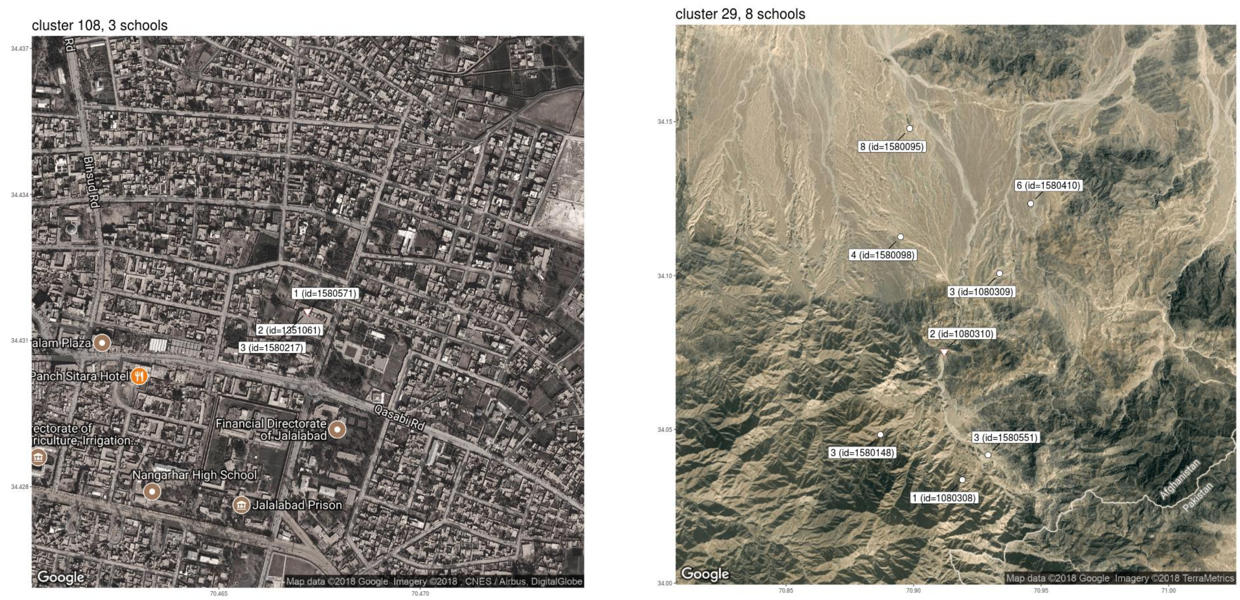
Table 4: Impacts on Mobile Money Use from Administrative Data

	Deposit (1)	Transfer to Cust. Wallet (2)	Own Airtime Top-up (3)	Other Airtime Top-up (4)	Balance (5)
Panel A. OLS					
MSP Months	0.004***	0.003**	0.001***	0.002***	0.012***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.002)
1-Month Mean	0.002	0.000	0.000	0.000	0.139
Observations	14,780	14,780	14,780	14,780	14,780
R^2	0.005	0.015	0.022	0.042	0.125
# Reg. Zones	301	301	301	301	301
Panel B. 2SLS					
MSP Months	0.004***	0.004	0.001***	0.002***	0.005
	(0.001)	(0.003)	(0.000)	(0.000)	(0.003)
Observations	14,780	14,780	14,780	14,780	14,780
R^2	0.002	0.001	0.011	0.017	0.005
# Reg. Zones	301	301	301	301	301

Notes: This table reports estimates of the impact of receiving salary via mobile money on other dimensions of mobile money use. The MSP months variable counts the number of months that an employee has been paid by mobile money. 2SLS outcomes instrument the number of months with the treatment assignment. Outcome data reflect transactions conducted between October 2018 and December 2020. Deposits are money added to the mobile money wallet via agents or bank transfers. Transfers to customer wallet are peer-to-peer transfers to another mobile money user. Airtime top-ups are money added to the pre-paid mobile phone call plan. Pre-pay balance is the remaining balance the day before salary payment for each employee. All variables are normalized by monthly salary. Robust standard errors clustered at the registration zone (treatment unit) level are reported in parentheses.
Levels of significance: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix - For Online Publication

A Supplementary Figures and Tables



Panel A: Urban Registration Zone

Panel B: Rural Registration Zone

Figure A.1: Example of Two Registration Zones in Nangarhar

Notes: This figure shows examples of two registration zones located in urban and rural areas of Nangarhar Province. White windows display the location and identifiers of the schools located in these zones.

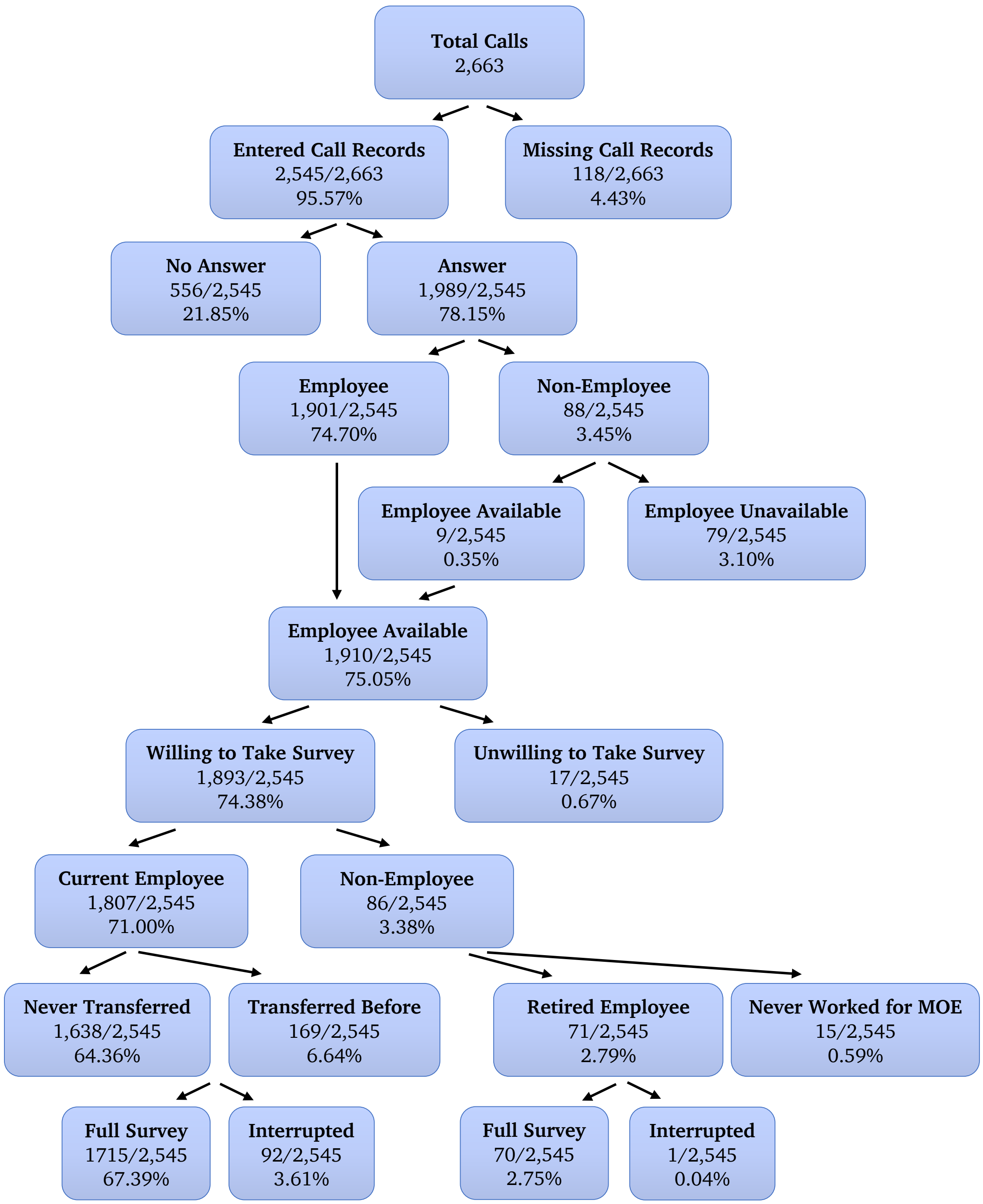


Figure A.2: Litmus Test Survey Diagram

Notes: This figure plots the diagram detailing responses to the 'Litmus Test' survey.

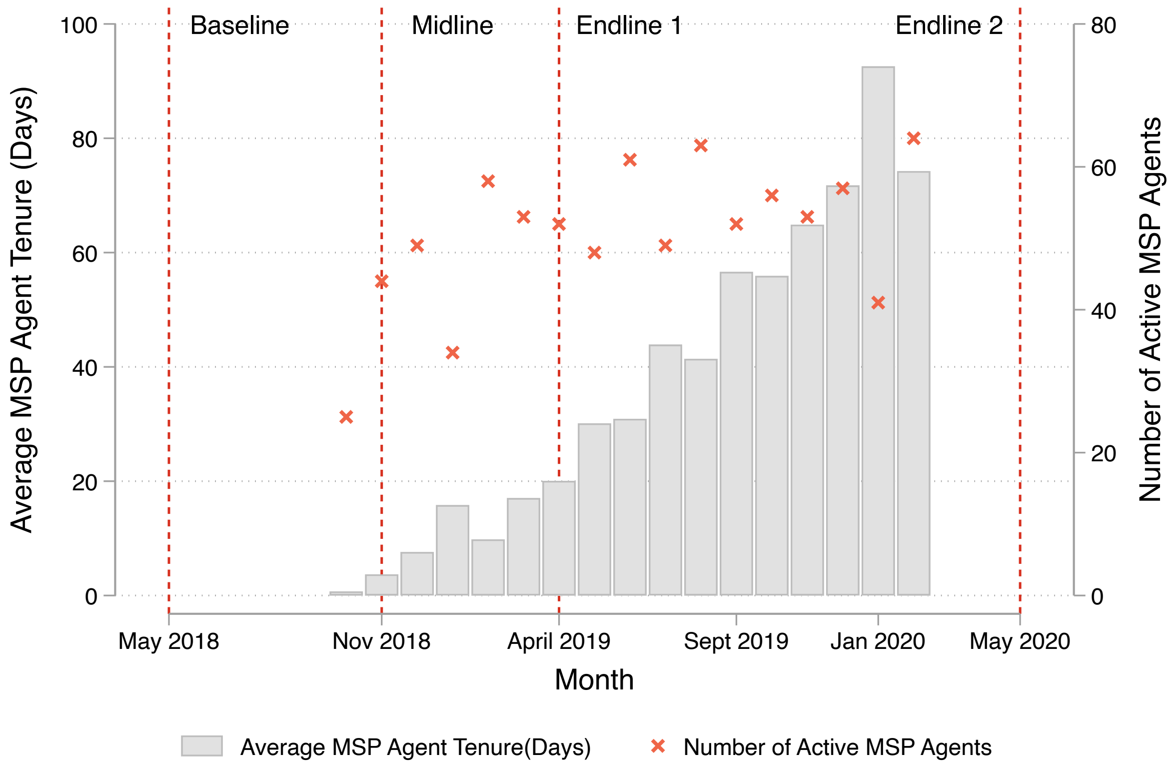


Figure A.3: Tenure of MSP Agents

Notes: This figure plots the average MSP tenure and the total number of active mobile money agents across months of program implementation. The tenure is measured as the total number of days an agent was active in prior months.

Table A.1: Balance Checks for Survey Outcomes

	DD Mean (1)	EE Effect (2)	ED Effect (3)	N Obs (4)
Panel A. Without Stratum FEs				
Vote in Favor of MSP	0.936	–0.027	–0.026	985
	(0.019)	(0.029)	(0.028)	
		[0.623]	[0.623]	
Very Satisfied with Pay System	0.297	–0.020	–0.041	999
	(0.034)	(0.046)	(0.047)	
		[0.949]	[0.623]	
Experienced Delay	0.543	0.083	–0.017	1,001
	(0.039)	(0.053)	(0.056)	
		[0.074]	[0.976]	
Travel Time to Cash-Out (Min)	33.412	–0.541	–0.983	928
	(3.192)	(4.370)	(4.297)	
		[0.976]	[0.976]	
Payment to Receive Salary (Afg)	23.533	0.806	–3.109	990
	(4.167)	(5.371)	(5.160)	
		[0.976]	[0.880]	
Panel B. With Stratum FEs				
Vote in Favor of MSP	0.936	–0.027	–0.024	984
	(0.019)	(0.027)	(0.024)	
		[0.649]	[0.649]	
Very Satisfied with Pay System	0.297	–0.022	–0.041	998
	(0.034)	(0.042)	(0.044)	
		[0.862]	[0.649]	
Experienced Delay	0.543	0.085**	0.005	1,000
	(0.039)	(0.042)	(0.043)	
		[0.024]	[0.885]	
Travel Time to Cash-Out (Min)	33.412	–2.378	–1.838	927
	(3.192)	(3.647)	(3.348)	
		[0.821]	[0.862]	
Payment to Receive Salary (Afg)	23.533	1.089	–3.956	989
	(4.167)	(3.735)	(4.216)	
		[0.885]	[0.649]	

Notes: This table checks balance on the main outcomes reflecting employees' salary experience. Robust standard errors are clustered at the registration zone (treatment unit) level. FWER-adjusted p-values within each panel are reported in squared brackets (following Romano & Wolf, 2005, using 1000 repetitions).
Levels of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table A.2: First Stage Estimates with Strata Fixed Effects

	DD Mean (1)	EE Effect (2)	ED Effect (3)	EE - ED Difference (4)	F-stat (5)	N Obs. (6)
Panel A. Endline 1 – April 2019						
Share Registered	0.65	0.26***	0.25***	0.01	25.03	970
	[0.04]	[0.04]	[0.04]	[0.02]		
Share Self-Reported MSP	0.08	0.63***	0.37***	0.27***	117.37	970
	[0.02]	[0.04]	[0.04]	[0.05]		
Panel B. Endline 2 – May 2020						
Share Registered	0.66	0.29***	0.25***	0.03	25.93	739
	[0.05]	[0.04]	[0.04]	[0.02]		
Share Self-Reported MSP	0.33	0.49***	0.15***	0.32***	68.09	739
	[0.04]	[0.04]	[0.04]	[0.04]		

Notes: This table reports the first stage estimates, controlling for the district (strata) fixed effects. The Endline 1 (2) sample consists of 970 (739) MoE employees who participated in the full survey and self-reported the payment system. Share registered is the share of employees who registered for a mobile money wallet before Endline 1 and 2, respectively. Share self-reported MSP is the share of employees that self-reported receiving their last salary payment via Mobile Salary Payments at Endline 1 and 2, respectively. Robust standard errors clustered at the registration zone (treatment unit) level in squared brackets. *Levels of significance*: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.3: Registration Outcomes

	Registered On Time (1)	Registered With Delay (2)	Never Registered (3)
Panel A: Share Registered			
% Employees	94.00%	3.22%	2.78%
	13333	457	394
Panel B: Share Paid After Registration			
1 Month After Registration	99.74%	99.34%	89.59%
	13299	454	353
12 Months After Registration	93.89%	88.84%	56.85%
	12518	406	224
20 Months After Registration	89.63%	80.31%	53.55%
	11950	367	211

Notes: This table presents the share and number of employees by their registration outcome (Panel A) and by their propensity of getting paid after the start of registration (Panel B). The sample consists of employees in Early Registration Zones who were paid in each of the 6 months before the start of registration. 23 Registration Zones in the Province of Nangarhar were omitted due to incomplete registration with less than 50% registration rate. The employee is "Registered On Time" if he or she registered for a mobile money wallet during the main registration wave (before July 18, 2018). Alternatively, the employee is "Registered With Delay" if he or she registered after the main registration wave (after July 18, 2018).

Table A.4: Wage Bill Changes Post-Registration At Individual Level

	May Treatment		Nov. Treatment	
	All	Restricted	All	Restricted
	(1)	(2)	(3)	(4)
Post × Treatment	–4.604	0.605	–10.040	–3.347
	(20.435)	(16.946)	(21.342)	(18.628)
Observations	999,827	863,174	999,827	863,174
R squared	0.679	0.664	0.679	0.664
Control Mean (Pre)	7294.450	7294.450	7260.410	7260.410
# Reg. Zones	401	400	401	400

Notes: This table estimates the impact of the reform on the wage bill. Outcome is monthly salary. Treatment is a dummy equal to one for individuals in the EE and ED treatment groups. Post is a dummy equal to one starting from either May 2018 (columns 1 and 2) or November 2018 (column 3 and 4). Whole sample or restricting to those paid 6 months before the intervention. Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.5: Wage Bill Changes Post-Registration At School Level

	May Treatment				November Treatment			
	Employees	Net Total	Net	Salary	Employees	Net Total	Net	Salary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Whole Sample								
Post × Treatment	–0.175	120.411	–2.691	–609.429	–0.230	–676.938	–154.574	–664.631
	(0.134)	(1499.929)	(1332.909)	(950.226)	(0.139)	(1688.863)	(1446.216)	(1005.278)
Observations	55,080	55,080	55,080	55,080	55,080	55,080	55,080	55,080
R squared	0.982	0.940	0.974	0.990	0.982	0.940	0.974	0.990
Control Mean (Pre)	17.240	1.5e+05	1.4e+05	1.3e+05	17.100	1.5e+05	1.4e+05	1.3e+05
# Reg. Zones	401	401	401	401	401	401	401	401
Panel B. Restricted Sample								
Post × Treatment	–0.082	810.573	–71.754	–320.592	–0.085	722.996	–108.319	–369.144
	(0.099)	(1035.729)	(869.229)	(686.152)	(0.121)	(1124.233)	(993.175)	(823.699)
Observations	51,876	51,876	51,876	51,876	51,876	51,876	51,876	51,876
R squared	0.994	0.960	0.978	0.992	0.994	0.960	0.978	0.992
Control Mean (Pre)	16.990	1.5e+05	1.4e+05	1.3e+05	16.970	1.5e+05	1.4e+05	1.3e+05
# Reg. Zones	400	400	400	400	400	400	400	400

Notes: This table estimates the impact of the reform on the wage bill. Treatment is a dummy equal to one for schools in the EE and ED treatment groups. Post is a dummy equal to one starting from either May 2018 (columns 1 to 4) or November 2018 (column 5 to 8). Whole sample or restricting to those paid 6 months before the intervention. Aggregated at the school level. Outcome variables are number of employed teachers (columns 1 and 5), net total monthly wage bill (columns 2 and 6, includes for example overtime payments), net monthly wage bill (columns 3 and 7), and gross monthly salary (columns 4 and 8). Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A.6: Wage Bill Changes Post-Registration At Registration Zone Level

	May Treatment				November Treatment			
	Employees	Net Total	Net	Salary	Employees	Net Total	Net	Salary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Whole Sample								
Post × Treatment	–0.662	–1220.517	–1450.051	–2840.663	–0.924*	–5306.418	–2683.089	–3379.369
	(0.521)	(5868.961)	(5140.781)	(3810.658)	(0.548)	(6379.207)	(5564.581)	(4174.463)
Observations	14,436	14,436	14,436	14,436	14,436	14,436	14,436	14,436
R squared	0.982	0.925	0.971	0.991	0.982	0.925	0.971	0.991
Control Mean (Pre)	67.950	585365.625	550405.812	504775.656	67.390	578794.750	547659.812	502922.125
# Reg. Zones	401	401	401	401	401	401	401	401
Panel B. Restricted Sample								
Post × Treatment	–0.183	2926.333	–359.887	–782.632	–0.169	2474.371	–541.319	–863.931
	(0.392)	(3759.610)	(3143.914)	(2435.445)	(0.450)	(4074.044)	(3595.988)	(2831.133)
Observations	14,400	14,400	14,400	14,400	14,400	14,400	14,400	14,400
R squared	0.995	0.960	0.976	0.993	0.995	0.960	0.976	0.993
Control Mean (Pre)	62.490	536239.562	517383.594	477013.281	62.420	533839.938	517029.375	476881.625
# Reg. Zones	400	400	400	400	400	400	400	400

Notes: This table estimates the impact of the reform on the wage bill. Treatment is a dummy equal to one for registration zones in the EE and ED treatment groups. Post is a dummy equal to one starting from either May 2018 (columns 1 to 4) or November 2018 (column 5 to 8). Whole sample or restricting to those paid 6 months before the intervention. Aggregated at the registration zone level. Outcome variables are number of employeed teachers (columns 1 and 5), net total monthly wage bill (columns 2 and 6, includes for example overtime payments), net monthly wage bill (columns 3 and 7), and gross monthly salary (columns 4 and 8). Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.7: Balance of Nonresponse Rates for Main Outcomes

	Sample Mean (1)	DD Mean (2)	EE Mean (3)	ED Mean (4)	p-value DD=EE=ED (5)	N Obs. (6)
Panel A. Baseline – May 2018						
Vote in Favor of MSP	0.02	0.02	0.02	0.02	0.82	1005
	(0.14)	[0.01]	[0.01]	[0.01]		
Share Very Satisfied with Pay System	0.01	0.00	0.01	0.00	0.5	1005
	(0.08)	[0.00]	[0.01]	[0.00]		
Experienced Delay	0.00	0.00	0.00	0.01	0.25	1005
	(0.06)	[0.00]	[0.00]	[0.01]		
Payment to Receive Salary (Afg)	0.01	0.02	0.01	0.02	0.88	1005
	(0.12)	[0.01]	[0.01]	[0.01]		
Travel Time to Cash-Out (Min)	0.08	0.08	0.08	0.07	0.86	1005
	(0.27)	[0.02]	[0.02]	[0.02]		
Panel B. Endline 1 – April 2019						
Vote in Favor of MSP	0.02	0.03	0.01	0.02	0.31	970
	(0.14)	[0.01]	[0.01]	[0.01]		
Share Very Satisfied with Pay System	0.00	0.01	0.00	0.00	0.83	970
	(0.06)	[0.00]	[0.00]	[0.00]		
Experienced Delay	0.00	0.00	0.00	0.00	0.32	970
	(0.03)	[0.00]	[0.00]	[0.00]		
Payment to Receive Salary (Afg)	0.01	0.01	0.02	0.01	0.44	970
	(0.11)	[0.00]	[0.01]	[0.01]		
Travel Time to Cash-Out (Min)	0.05	0.05	0.04	0.06	0.64	970
	(0.22)	[0.02]	[0.01]	[0.02]		
Panel C. Endline 2 – May 2020						
Vote in Favor of MSP	0.04	0.06	0.02	0.03	0.14	739
	(0.19)	[0.02]	[0.01]	[0.01]		
Share Very Satisfied with Pay System	0.00	0.00	0.00	0.01	0.22	739
	(0.06)	[0.00]	[0.00]	[0.01]		
Experienced Delay	0.01	0.00	0.00	0.01	0.13	739
	(0.07)	[0.00]	[0.00]	[0.01]		
Payment to Receive Salary (Afg)	0.00	0.00	0.00	0.00	0.32	739
	(0.04)	[0.00]	[0.00]	[0.00]		
Travel Time to Cash-Out (Min)	0.02	0.02	0.01	0.03	0.48	739
	(0.14)	[0.01]	[0.01]	[0.01]		

Notes: This table reports balance of nonresponse rates for the main outcomes of the study. The sample consists of MoE employees who participated in the full survey at the baseline, endline 1 and endline 2, respectively. Standard errors in parentheses and robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.

Table A.8: Registration Effects on the Main Outcomes

Outcome	Unregistered Mean (1)	Registration Effect (2)	N Obs. (3)
Panel A: Endline 1 – April 2019			
Vote in Favor of MSP	0.35	0.13	296
	[0.06]	[0.09]	
Very Satisfied with Pay System	0.90	0.02	305
	[0.03]	[0.08]	
Experienced Delay	0.36	0.05	307
	[0.06]	[0.08]	
Travel Time to Cash-Out (Min)	31.69	-1.44	290
	[4.26]	[6.46]	
Payment to Receive Salary (Afg)	32.41	-1.30	305
	[4.69]	[9.29]	
Panel B: Endline 2 – May 2020			
Vote in Favor of MSP	0.46	0.01	158
	[0.04]	[0.14]	
Very Satisfied with Pay System	1.00	-0.01	172
	[0.00]	[0.01]	
Experienced Delay	0.46	-0.07	172
	[0.06]	[0.14]	
Travel Time to Cash-Out (Min)	37.58	1.42	171
	[3.89]	[7.99]	
Payment to Receive Salary (Afg)	21.95	-3.06	172
	[3.30]	[6.32]	

Notes: This table reports estimates of impacts of registration on payment quality by comparing respondents who were not yet registered (for which we report the mean of each dependent variable) to those who were already registered at the time of the survey. In order to isolate the effect of registration, we exclude from the sample respondents who reported receiving their salary via mobile money at time of the survey. Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A.9: Effects on Salary Payment Experience – Without Stratum FEs

	DD Mean (1)	EE Effect (2)	ED Effect (3)	ToT Effect (4)	N Obs (5)
Panel A. Endline 1					
Vote in Favor of MSP	0.422	0.081	0.052	0.128	950
	(0.040)	(0.057)	(0.057)	(0.088)	
		[0.069]	[0.307]		
Very Satisfied with Pay System	0.628	–0.257***	–0.146***	–0.401***	966
	(0.037)	(0.053)	(0.055)	(0.085)	
		[0.001]	[0.001]		
Experienced Delay	0.405	0.261***	0.118**	0.404***	969
	(0.037)	(0.052)	(0.057)	(0.080)	
		[0.001]	[0.002]		
Travel Time to Cash-Out (Min)	31.276	29.355***	20.677***	46.084***	922
	(3.241)	(6.298)	(7.224)	(9.423)	
		[0.001]	[0.001]		
Payment to Receive Salary (Afg)	22.311	–3.735	–0.452	–5.589	959
	(2.626)	(4.110)	(3.986)	(6.240)	
		[0.307]	[0.830]		
Conducted Mobile Money Transfer	0.063	0.170***	0.163***	0.276***	965
	(0.015)	(0.033)	(0.037)	(0.050)	
		[0.001]	[0.001]		
Panel B. Endline 2					
Vote in Favor of MSP	0.645	0.104**	0.011	0.233***	712
	(0.035)	(0.047)	(0.051)	(0.090)	
		[0.014]	[0.983]		
Very Satisfied with Pay System	0.727	–0.022	–0.022	–0.041	736
	(0.031)	(0.045)	(0.043)	(0.092)	
		[0.972]	[0.972]		
Experienced Delay	0.316	–0.021	–0.015	–0.039	735
	(0.030)	(0.044)	(0.043)	(0.089)	
		[0.972]	[0.983]		
Travel Time to Cash-Out (Min)	36.628	3.469	0.163	7.535	725
	(2.185)	(3.264)	(2.976)	(6.632)	
		[0.603]	[0.983]		
Payment to Receive Salary (Afg)	12.479	–8.062***	0.986	–18.396***	738
	(1.853)	(2.266)	(3.281)	(3.874)	
		[0.001]	[0.983]		

Notes: This table reports treatment effects on salary payment experience. The Endline 1 (2) sample consists of 970 (739) MoE employees who participated in the full survey and self-reported their payment system. ToT is a treatment-on-the-treated effect obtained by instrumenting self-reported MSP payments with the treatment group assignment. Robust standard errors clustered at the registration zone (treatment unit) level. FWER-adjusted p-values within each panel in squared brackets (following Romano & Wolf, 2005, using 1000 repetitions).
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.10: Effects on Salary Payment Experience – With Stratum FEs

	DD Mean (1)	EE Effect (2)	ED Effect (3)	ToT Effect (4)	N Obs (5)
Panel A. Endline 1					
Vote in Favor of MSP	0.422	0.110**	0.076*	0.178***	950
	(0.040)	(0.044)	(0.045)	(0.068)	
		[0.001]	[0.039]		
Very Satisfied with Pay System	0.628	–0.252***	–0.145***	–0.397***	966
	(0.037)	(0.049)	(0.048)	(0.078)	
		[0.001]	[0.001]		
Experienced Delay	0.405	0.254***	0.146***	0.401***	969
	(0.037)	(0.048)	(0.049)	(0.075)	
		[0.001]	[0.001]		
Travel Time to Cash-Out (Min)	31.276	27.193***	21.824***	43.208***	922
	(3.241)	(5.880)	(6.052)	(8.684)	
		[0.001]	[0.001]		
Payment to Receive Salary (Afg)	22.311	–3.208	–1.644	–4.966	959
	(2.626)	(3.402)	(2.996)	(5.109)	
		[0.294]	[0.395]		
Conducted Mobile Money Transfer	0.063	0.178***	0.168***	0.293***	965
	(0.015)	(0.033)	(0.034)	(0.049)	
		[0.001]	[0.001]		
Panel B. Endline 2					
Vote in Favor of MSP	0.645	0.122***	0.013	0.263***	712
	(0.035)	(0.043)	(0.040)	(0.082)	
		[0.003]	[0.981]		
Very Satisfied with Pay System	0.727	–0.003	–0.015	–0.001	736
	(0.031)	(0.041)	(0.040)	(0.081)	
		[0.991]	[0.981]		
Experienced Delay	0.316	–0.033	–0.018	–0.064	735
	(0.030)	(0.042)	(0.040)	(0.083)	
		[0.878]	[0.981]		
Travel Time to Cash-Out (Min)	36.628	1.661	–0.244	3.677	725
	(2.185)	(3.185)	(2.999)	(6.110)	
		[0.981]	[0.991]		
Payment to Receive Salary (Afg)	12.479	–7.787***	1.252	–17.563***	738
	(1.853)	(2.033)	(2.549)	(3.955)	
		[0.001]	[0.981]		

Notes: This table reports treatment effects on salary payment experience, controlling for the district (strata) fixed effects. The Endline 1 (2) sample consists of 970 (739) MoE employees who participated in the full survey and self-reported their payment system. ToT is a treatment-on-the-treated effect obtained by instrumenting self-reported MSP payments with the treatment group assignment. Robust standard errors clustered at the registration zone (treatment unit) level. FWER-adjusted p-values within each panel in squared brackets (following Romano & Wolf, 2005, using 1000 repetitions).
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.11: Heterogeneity Treatment Effects – Security

	Vote in Favor of MSP (1)	Very Satisfied with Pay System (2)	Experienced Delay (3)	Travel Time to Cash-Out Salary (Min) (4)	Payment to Receive Transfer (5)	Conducted Mobile Money Transaction (6)
Panel A. Year 1 Outcomes						
β_1: MSP × Urban	0.306***	–0.274***	0.239**	33.788***	–6.855	0.299***
	(0.089)	(0.102)	(0.102)	(8.192)	(7.821)	(0.063)
β_2: MSP × Rural Controlled	–0.070	–0.586***	0.670***	62.507***	–7.784	0.269***
	(0.119)	(0.129)	(0.110)	(21.734)	(6.332)	(0.084)
β_3: MSP × Rural Contested	0.023	–0.543**	0.523**	33.527*	14.735	0.365**
	(0.215)	(0.250)	(0.253)	(18.506)	(12.375)	(0.178)
p-value $\beta_2 = \beta_3$	0.706	0.880	0.594	0.311	0.106	0.625
p-value $\beta_1 = \beta_2$	0.012	0.059	0.004	0.217	0.926	0.773
p-value $\beta_1 = (\beta_2 + \beta_3)/2$	0.031	0.095	0.038	0.388	0.324	0.879
Control Mean	0.422	0.628	0.405	31.276	22.311	0.063
Observations	950	966	969	922	959	965
R squared	0.012	0.034	0.076	0.088	0.002	0.067
# Reg. Zones	350	352	352	344	352	352
Panel B. Year 2 Outcomes						
β_1: MSP × Urban	0.236**	0.238**	–0.265**	–0.275	–9.878**	
	(0.115)	(0.105)	(0.116)	(7.145)	(4.289)	
β_2: MSP × Rural Controlled	0.253**	–0.314**	0.142	8.179	–30.732***	
	(0.127)	(0.126)	(0.107)	(9.935)	(7.762)	
β_3: MSP × Rural Contested	0.414	–0.320	0.336	15.431	–26.572	
	(0.265)	(0.297)	(0.302)	(27.369)	(16.595)	
p-value $\beta_2 = \beta_3$	0.583	0.984	0.545	0.803	0.821	
p-value $\beta_1 = \beta_2$	0.923	0.001	0.010	0.490	0.019	
p-value $\beta_1 = (\beta_2 + \beta_3)/2$	0.602	0.004	0.011	0.457	0.064	
Control Mean	0.645	0.727	0.316	36.628	12.479	
Observations	712	736	735	725	738	
R squared	0.085	–0.043	–0.034	0.010	0.056	
# Reg. Zones	322	328	328	324	327	

Notes: This table reports ToT estimates of impacts of mobile salary payments on payment quality and mobile money use by security and population density. Favor MSP is a dummy variable equal to one if the respondent indicates they would support scaling the reform across the Ministry of Education. Satisfied with Pay System is a dummy variable equal to one of the respondent indicates a very high level of satisfaction. Experienced Delay is a dummy variable equal to one if the respondent reports their salary being delayed. Travel Time is the time to convert the mobile money payment to cash in minutes. Payment to Recieve measures what respondents report paying to receive their salary in Afghanis. For comparison, the average (net) salary in the sample is 8560 Afg. Money Transfer is a dummy variable equal to one if the respondent indicates making a mobile money transfer to someone else in the previous month. This outcome is not recorded in year 2 because the survey was abbreviated due to the pandemic. Urban areas are those with a population density greater than 300 inhabitants per squared kilometer. Contested areas are those that ACSOR categorizes as restricted access for survey enumerators (categories 3 to 5). Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.12: Estimates of Impact on Education - ITT

	Math Score			Reading Score			Combined Score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
EE Effect	0.071	0.180	-0.096	0.065	0.142	-0.033	0.068	0.161	-0.064
	(0.083)	(0.117)	(0.115)	(0.082)	(0.114)	(0.119)	(0.075)	(0.107)	(0.100)
ED Effect	0.048	0.236**	-0.199	-0.017	0.081	-0.145	0.015	0.158	-0.172
	(0.085)	(0.111)	(0.125)	(0.079)	(0.107)	(0.115)	(0.075)	(0.101)	(0.107)
$H_0: T1 + T2 = 0$ p-value	0.41	0.03	0.16	0.73	0.25	0.38	0.52	0.08	0.19
Control Group Mean	0.01	-0.07	0.12	0.01	0.00	0.01	0.01	-0.03	0.07
Sample	Full	Urban	Rural	Full	Urban	Rural	Full	Urban	Rural
District FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓	✓	✓	✓	✓
R-squared	0.30	0.28	0.36	0.25	0.24	0.27	0.31	0.29	0.35
# Reg. Zones	362	201	161	362	201	161	362	201	161
# Students	939	516	423	939	516	423	939	516	423

Notes: This table reports the impacts of the Mobile Salary Payment Reform on student learning outcomes measured in the assessment described in Section 2.5. Scores are standardized using the control group mean and standard deviation. The sample comprises 939 students aged between 6 and 10 years who attended a public school eligible for the Mobile Salary Payments reform within the last three months. EE Effect is a dummy variable equal to one for students who attend a school in the EE treatment group. ED Effect is a dummy variable equal to one for students who attend a school in the ED treatment group. Covariates include years of education, years of age, a dummy variable equal to one if a student lives in a high socio-economic status household, and gender. High SES households are those that are on a maintained road with access to water and to electricity. 458 of the 939 students are high SES. 343 of the students are female. Standard errors clustered at the registration zone level are reported in parentheses. *Levels of significance*: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A.13: Years of Education and Learning Assessment Scores

	Math Score		Reading Score		Combined Score	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.284***	0.279***	0.233***	0.242***	0.259***	0.261***
	(0.041)	(0.041)	(0.042)	(0.043)	(0.039)	(0.039)
Female (=1)		-0.148		-0.079		-0.114
		(0.108)		(0.115)		(0.103)
High SES (=1)		0.406***		0.311**		0.358***
		(0.123)		(0.128)		(0.116)
Fixed Effects	None	District	None	District	None	District
Mean Outcome (0 Yrs Ed.)	-0.53	-0.53	-0.39	-0.39	-0.46	-0.46
Mean Outcome (5 Yrs Ed.)	0.59	0.59	0.52	0.52	0.56	0.56
R-squared	0.14	0.36	0.10	0.29	0.14	0.34
# Children	313	313	313	313	313	313

Notes: This table reports on the relationship between students' years of education and their standardized scores on the learning assessment conducted by our team. The learning assessment is described in Section 2.5. Scores are standardized using the control group mean and standard deviation. The sample comprises students aged between 6 and 10 years who attended a public school eligible for the Mobile Salary Payments reform within the last three months for whom we have demographic data in control registration zones. High SES is a variable for students living in High Socio-Economic Status households, which are on a maintained road with access to water and to electricity. 152 of 313 students are characterized as high SES. 114 of the students are female. Robust standard errors are reported in parentheses. *Levels of significance*: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A.14: Treatment Effects on Attendance

	DD Mean (1)	EE Effect (2)	ED Effect (3)	ToT Effect (4)	N Obs (5)
Share Present at Baseline	0.408	0.007	0.006		15,323
	[0.026]	[0.029]	[0.027]		
Share Present at Midline	0.402	–0.019	–0.020		15,253
	[0.022]	[0.022]	[0.023]		
Share Present at Endline 1	0.368	–0.033	–0.009	–0.046	15,006
	[0.021]	[0.023]	[0.022]	[0.033]	

Notes: This table reports treatment effects for teachers' attendance during the audit exercise at baseline, midline and endline 1, controlling for the district (strata) fixed effects. ToT is the treatment-on-the-treated effect obtained by instrumenting receiving MSP before endline 2 with the treatment group assignment. Robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A.15: Treatment Effects on Attendance – By Urban & Rural

	DD Mean (1)	EE Effect (2)	ED Effect (3)	ToT Effect (4)	N Obs (5)
Panel A. Urban					
Share Present at Baseline	0.356	0.018	0.017		11,448
	[0.029]	[0.037]	[0.033]		
Share Present at Midline	0.358	–0.018	–0.011		11,448
	[0.023]	[0.027]	[0.028]		
Share Present at Endline 1	0.331	–0.033	–0.010	–0.046	11,340
	[0.024]	[0.027]	[0.027]	[0.039]	
Panel B. Rural					
Share Present at Baseline	0.564	–0.032	–0.033		3,875
	[0.034]	[0.035]	[0.031]		
Share Present at Midline	0.533	–0.022	–0.051		3,805
	[0.036]	[0.036]	[0.037]		
Share Present at Endline 1	0.483	–0.029	–0.002	–0.046	3,666
	[0.029]	[0.041]	[0.034]	[0.062]	

Notes: This table reports treatment effects for teachers' attendance during the audit exercise at baseline, midline and endline 1, controlling for the district (strata) fixed effects. Panel A restricts the sample to those schools located in urban areas. Panel B restricts the sample to those schools located in rural areas. ToT is the treatment-on-the-treated effect obtained by instrumenting receiving MSP before endline 2 with the treatment group assignment. Robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A.16: Heterogeneity Impacts on Mobile Money Use from Administrative Data – Urban & Rural

	Deposit (1)	Transfer to Cust. Wallet (2)	Own Airtime Top-up (3)	Other Airtime Top-up (4)	Balance (5)
Panel A. ITT					
β_1: MSP Months × Urban	0.003***	0.002***	0.000***	0.001***	0.010***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
β_2: MSP Months × Rural	0.002***	0.002*	0.000***	0.000***	0.025***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)
Observations	23,234	23,234	23,234	23,234	23,234
R^2	0.009	0.011	0.019	0.032	0.133
1-Month Mean	0.002	0	0.000	0.000	0.139
p-value $\beta_1 = \beta_2$	0.043	0.612	0.191	0.013	0.000
# Reg. Zones	301	301	301	301	301
Panel B. 2SLS					
β_1: MSP Months × Urban	0.003***	0.002*	0.000***	0.001***	0.008***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)
β_2: MSP Months × Rural	0.002***	0.001	0.000***	0.000**	0.024***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)
Observations	23,234	23,234	23,234	23,234	23,234
R^2	0.006	0.002	0.006	0.008	0.061
p-value $\beta_1 = \beta_2$	0.079	0.493	0.351	0.016	0.000
# Reg. Zones	301	301	301	301	301

Notes: This table reports estimates of the impact of receiving salary via mobile money on other dimensions of mobile money use. The MSP months variable counts the number of months that an employee has been paid by mobile money. 2SLS outcomes instrument the number of months with the treatment assignment. Outcome data reflect transactions conducted between October 2018 and December 2020. Deposits are money added to the mobile money wallet via agents or bank transfers. Transfers to customer wallet are peer-to-peer transfers to another mobile money user. Airtime top-ups are money added to the pre-paid mobile phone call plan. Pre-pay balance is the remaining balance the day before salary payment for each employee. All variables are normalized by monthly salary. Robust standard errors clustered at the registration zone (treatment unit) level are reported in parentheses. Urban zones are those that have a population density of more than 300 inhabitants per squared kilometer. Impacts on Mobile Money Use from Administrative Data

Table A.17: Spatial Treatment Externalities on Salary Payment Experience (within 2 km)

	DD Mean (1)	EE Effect (2)	ED Effect (3)	# EE Neigh Schools 2 km (4)	# ED Neigh Schools 2 km (5)	# Total Neigh Schools 2 km (6)	N N Obs. (7)
Panel A. Endline 1 – April 2019							
Vote in Favor of MSP	0.42	0.11**	0.09**	0.00	–0.02	0.00	950
	[0.04]	[0.05]	[0.04]	[0.02]	[0.02]	[0.01]	
Experienced Delay	0.63	–0.25***	–0.14***	0.00	0.00	0.00	966
	[0.04]	[0.05]	[0.05]	[0.02]	[0.02]	[0.01]	
Payment to Receive Salary (Afg)	0.41	0.25***	0.15***	0.00	–0.01	0.01	969
	[0.04]	[0.05]	[0.05]	[0.02]	[0.02]	[0.01]	
Travel Time to Cash-Out (Min)	22.31	–4.05	–1.04	0.64	–0.73	–0.09	959
	[2.63]	[3.02]	[3]	[1.2]	[0.94]	[0.55]	
Conducted Mobile Money Transfer	31.28	28.1***	22.17***	–0.32	0.19	–0.77	922
	[3.24]	[6.31]	[6.52]	[1.56]	[1.69]	[0.79]	
Panel B. Endline 2 – May 2020							
Vote in Favor of MSP	0.64	0.11***	0.01	0.01	0.00	0.00	712
	[0.03]	[0.04]	[0.04]	[0.01]	[0.02]	[0.01]	
Experienced Delay	0.73	–0.01	–0.02	0.01	0.00	0.00	736
	[0.03]	[0.04]	[0.04]	[0.01]	[0.01]	[0.01]	
Payment to Receive Salary (Afg)	0.32	–0.03	–0.03	0.00	0.01	–0.01	735
	[0.03]	[0.04]	[0.04]	[0.01]	[0.01]	[0.01]	
Travel Time to Cash-Out (Min)	12.48	–8.72***	1.54	0.83	–0.49	–0.11	738
	[1.85]	[2.25]	[2.59]	[0.85]	[0.9]	[0.39]	
Conducted Mobile Money Transfer	36.63	3.36	–0.24	–1.35	0.42	–0.03	725
	[2.18]	[3.27]	[3.05]	[1.1]	[1]	[0.61]	

Notes: This table reports spatial treatment externalities for the main outcomes of the study within a two-kilometer radius, controlling for the district (strata) fixed effects. The Endline 1 (2) sample consists of 970 (739) MoE employees who participated in the full survey and self-reported the payment system. EE (ED) Neigh Schools 2 km is the number of neighboring schools within a two-kilometer radius located in EE (ED) treatment zone, Total Neigh Schools 2 km is the total number of neighboring schools within the same radius. Robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

B Validating Ghost Employees: Details

B.1 Survey Sampling

We sampled a total of 2,663 employees who had a phone number listed in the payroll records, with a breakdown as follows:[31]

1. 753 employees who did not register by the time of the phone survey. We included in the sample all of the employees who had not registered even if they should have. This is the group that we expect to have the highest proportion of ghost workers.

2. 987 employees who did not register by July 19, 2018, but did register before the phone survey took place. These employees are potentially suspicious because they failed to appear during the registration drive, but did eventually get registered. These could be either 'stand-ins', who are not employees in any real sense, or simply employees who genuinely could not make registration. We sampled all employees fitting this description.

3. Three additional categories of employees, all of whom registered by July 19, 2018 and belong to the same schools as the samples 1 and 2. These provide useful comparisons to the two suspicious cases above, since they are similar employees, working at the same school, who were given the same opportunity to register, but *did* appear for registration. These three categories are:

   a. 175 employees who registered in time, are working in the 182 schools which were visited for the baseline unannounced visit, and who were present during this audit visit. We sampled one employee in each school, but 7 schools had no employee satisfying these criteria. This is the sample of employees who are most likely to be genuine employees: they were both present at school at baseline and registered on time. For this reason, we consider them our main comparison group.

   b. 153 employees who registered in time and are working in the 182 schools which were visited for the baseline audit, but who were absent during the audit. We sampled one employee in each school, but 29 schools had no employee satisfying these criteria. These might be either real employees who happened to be absent at the time of the audit, or 'stand-ins' who registered during the first wave of registration.

   c. 356 employees who registered on time and are working in the 356 schools which were not selected for the baseline audit. We sample one employee in this category from each school.

[31]Almost all employees have a (possibly outdated) phone number listed in their payroll records. For employees who appear for registration, we also observe the new phone number that was given to them to open the mobile wallet account. Moreover, for employees who participated in our baseline survey, we also have the phone numbers that they reported as currently using. Using these phone numbers, when available, could have improved our chance of reaching those employees. However, to ensure that all employees had an *ex ante* equal chance of being contacted, we only called numbers listed in the payroll records before the experiment started. Appendix Figure A.2 summarizes the litmus survey response outcomes.

4. 239 employees who registered on time and who worked in the 239 schools that had no ghost workers at all (we sample one employee per school). These are potentially of interest because they either work in extremely well-run schools, or the perfect registration record reflects a successful attempt to provide stand-ins for each slot on the payroll.

B.2 Stand-ins Estimates

We introduce some notation for clarifying how we use the results of the litmus test to bound the number of stand-ins who registered. Let T be a dummy variable equal to 1 if a respondent is a true employee, Z be a dummy variable which is instead equal to 1 if the respondent is a stand-in, and R be a dummy for employees who registered, with $R = T + Z$.

Ideally, we would have liked to design a litmus test which could discriminate perfectly between true employees and stand-ins. Let L be a dummy variable equal to 1 if the respondent passed the litmus test (for example, because he answered all 7 questions correctly, but the threshold could be even a different one): then we would like the probability of success μ for true employees to be $\mu_T = P(L = 1|T = 1) = 1$ and $\mu_Z = P(L = 1|Z = 1) = 0$, that is all true employee will pass the test and all stand-ins will fail it. The success rate among registered employees μ_R can be decomposed into $\mu_R = \mu_T \times P(T = 1) + \mu_Z \times P(Z = 1)$, and with $\mu_T = 1$ and $\mu_Z = 0$, $1 - \mu_R = P(Z = 1)$ identifies the proportion of stand-ins among registered employees.

Designing such a litmus test, however, proved challenging: indeed, we need to ask a set of question which: i) would be easy enough for all true employees to answer correctly (even though some of them might have low literacy rates or other characteristics which would make answering the test hard), ii) would be non-trivial so that stand-ins could answer correctly, and iii) could be graded by us using information available (for example in the existing payroll records). While we ultimately failed to design a test which could discriminate perfectly true employees from false one, we were able to design a test which was nevertheless informative: employees who were more likely to be true employees (having registered in time and being present during our baseline audit visit) had on average higher scores than those less likely to be current employees (having not registered and being absent during our baseline audit).

This implies that, while we don't have $\mu_T = 1$ and $\mu_Z = 0$, it is possible to use the phone survey responses to design litmus test L such that $\mu_T > \mu_Z$. Moreover, below we show that adding some assumption to the exercise, it is still possible to recover information about the proportion $P(Z = 1)$ of stand-ins in the population of registered employees starting from the proportion of μ_R of employees who pass the test among those who register.

ASSUMPTION 1 First, we note that $\mu_R = \mu_T \times P(T = 1) + \mu_Z \times P(Z = 1)$ can be rewritten as $P(Z = 1) = \frac{\mu_R - \mu_T}{\mu_Z - \mu_T}$, so that assuming $\mu_Z = 0$ allows us to calculate a lower bound for $P(Z = 1)$.

Intuitively, the score among registered employees can be low either i) because a lot of stand-ins get registered, or ii) because stand-ins have extremely low scores: so, assuming that stand-ins have a score of zero bounds from below the possible size of the stand-in population. In this sense, assumption 1 is rather unproblematic.

However, while μ_R can be estimated through the proportion of respondents who pass the litmus test in the data, μ_T is unobserved, so that even the lower bound $P(Z=1)_{\mu_Z=0} = 1-\frac{\mu_R}{\mu_T}$ cannot be computed. Nevertheless, while we don't observe the success rate of all true employees because we cannot in general know who is a true employee and who is not, we observe it for a subsample of them: indeed, we can reliably consider employees who registered early and who were present during the baseline audit visit as true employees.

ASSUMPTION 2 We assume that there exist a test L such that the probability of success of all true employees μ_T is equal to the probability of success of the subset of true employees who registered early and who were present during the baseline audit visit μ_V (V for 'verified employees').

We consider this second assumption as more demanding: indeed, it could be the case that verified employees know more about their school with respect to true employees who are often absent from work. Using the threshold of a score of 7 in the litmus test would then be problematic, because verified employees would outperform absent true employees. The problem here is that in calculating $P(Z=1)_{\mu_Z=0} = 1-\frac{\mu_R}{\mu_V}$ we would attribute differences between the performance of all registered employees μ_R and the performance of verified employees μ_V entirely to the presence of stand-ins scoring zero, while it is actually in part attributable to the lower performance of true but absent employees ($\mu_V > \mu_T$). This problem can be alleviated by lowering the threshold for the litmus test: for example, using a threshold of 4 rather than 7, should make it more likely that the proportion of true employees scoring at least 4 would be well approximated by the proportion of verified employees scoring at least 4 (the fact that the true employees score relatively more 4 and 5, and the verified employees score more 6 and 7 would not matter). The issue with lowering the threshold too much is that the first assumption that $\mu_Z = 0$, which is likely justifiable for a threshold of 7, might become less reasonable for a threshold of 4, if relatively many stand-ins could be getting such a score, leading to a lower bound too far from the true $P(Z=1)$.

Using these assumptions and the results of the phone survey, we find that using a score of 7 results in a rate of stand-ins equal to at least 18.1% of registered employees ($1-\frac{15.14}{18.56}$, where the numerator is the percentage of registered employees who scored 7 on the litmus test, and the denominator is the percentage of employees who registered early, were present at the school baseline audit visit, and scored 7), a score of 4 in a lower bound of 8.8% ($1-\frac{66.79}{73.21}$), and simply answering the phone in a bound of 5.8% ($1-\frac{81.08}{86.11}$). Adjusting these numbers for the share of registered employees (97.2%), the bounds range from 5.6% to 17.6% of all employees.

C Spatial Treatment Externalities

The MSP reform could have created spatial externalities given the scale of its implementation and the amount of attention it received from policymakers. First, with the start of the reform, policymakers and ministries involved in its implementation could have paid more attention to improving the payment experience in all schools, regardless of their treatment status. This is partly evident from the changes in payment experience outcomes in the control group. Second, schools located near those in the treatment group could have felt pressure to improve the payment experience of their employees. Conversely, they could have also experienced deterioration in the payment experience if the reform caused local disruptions in the payment process.

While it is not feasible to formally test for the program-wide externalities due to data limitations, we test for spatial externalities across studied schools. Following Miguel and Kremer (2004), we estimate it using the following specification:

$$Y_{iszd} = \alpha + \beta_{EE}EE_z + \beta_{EL}EL_z + \eta_{EE}\,\#\,EE\ Neighbors^{2km}_{szd} + \eta_{EL}\,\#\,EL\ Neighbors^{2km}_{szd} + \\ + \gamma\,\#\,Total\ Neighbors^{2km}_{szd} + \mu_d + \varepsilon_{iszd} \quad (3)$$

where Y_{iszd} is the payment experience outcome for employee i from school s located in registration zone z and district d, $\#\,EE\ Neighbors^{2km}_{szd}$ and $\#\,EL\ Neighbors^{2km}_{szd}$ are the number of neighboring schools within a two-kilometer radius from school s located in EE and EL treatment zones, $\#\,Total\ Neighbors^{2km}_{szd}$ is the total number of neighboring schools within the same radius. The main identifying assumption is that, conditional on the total number of neighboring schools within a fixed radius, the number of treated neighboring schools is random.

Appendix Table A.17 presents the results. The estimates measuring treatment effects on payment experience in columns (2) and (3) remain similar to those reported in Appendix Table A.10. Moreover, the estimates measuring externalities in columns (4) and (5) are close to zero and not statistically significant. This implies that there is no evidence of spatial externalities across schools located within a two-kilometer radius.

D Cost Effectiveness

To provide a simple lower-bound estimate of the cost-effectiveness of the MSP reforms, we consider the fiscal returns to removing ghost workers from the payroll net of implementation costs.[32] We cost the MSP reform at the 195 AFN registration fee the government paid per employee successfully enrolled via biometric registration and the 100 AFN monthly fee per individual salary transfer for the mobile network operator to facilitate cash out by its mobile agents. While our experimental sample totals 34,422 employees, we focus on the 14,184 in the early registration group described in Table A.3, where we have the longest time series to observe payment outcomes. We observe that 2.8% (n=394) of the early registration group are never registered. The remaining 13,790 employees thus result in a one-time registration cost of approximately 2.7 million AFN and a monthly cost of 1.4 million AFN.

The fiscal benefits of the MSP reform entail the wages saved due to the removal of ghost workers. In the early registration group, 1.3% of employees (n=183) who never registered are eventually removed from the payroll, leading to a savings of 1.2% of the total monthly wage bill of 128 million AFN (e.g. approximately 1.5 million AFN). For ease of presentation, we

[32]This is likely a lower-bound as it does not account for efficiency gains related to time and travel costs to collect payment, or redistributive effects related to reduced payment delays and reduced leakage as in Muralidharan, Niehaus and Sukhtankar (2016). Each of these variables are statistically indistinguishable between the treatment and control groups by the time of the second endline survey in May 2020. This exercise also ignores the potential benefits to employees from increased financial inclusion discussed above.

assume costs and benefits are incurred simultaneously, which results in a breakeven period of 17.1 months and a modest return on investment of 3.0% after 24 months.[33] However, as mentioned above, we cannot reject the null hypothesis that the reduction in the wage bill in the treatment group is the same as in the control group, and so caution against a causal interpretation of the 1.2% number.

Counterfactually, if all of the 2.8% employees who never registered had been removed from the payroll, this would have produced a savings of 2.7% of the total monthly wage bill, or approximately 3.4 million AFN. Naturally, this scenario would result in a shorter break-even period of 1.3 months, and a much larger return on investment of 132% after 24 months. If the 5.6% of employees identified as *likely* ghosts (Appendix B) had been removed in addition to the 2.8% of employees who never registered, this would have produced a savings of 8.3% of the wage bill, or 10.6 million AFN. This alternative scenario would have resulted in a break-even period of 0.3 months, and a sizeable return of investment of 613% after 24 months – underscoring the large fiscal incentives to removing ghost workers from the payroll.

[33]In practice, there are delays both in the realization of costs – which the mobile network operator must invoice to the government – and in the realization of benefits – as the MoE adjudication committee must finalize its decisions to remove suspicious workers.