Modernizing the State During War: Experimental Evidence from Afghanistan*

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

Michael Callen[‡] London School of Economics and CEPR

Anastasiia Faikina[§] U.C. San Diego

Stefano Fiorin[¶] Bocconi, LEAP and CEPR

Tarek Ghani[‖] Washington University

September 7, 2022

This is a preliminary draft. Please do not distribute.

Abstract

This paper studies whether it is possible to build basic capacity while state control remains contested. We report evidence from a randomized evaluation of a major reform intended to improve the Afghan government's ability to pay its employees, which involved over 30,000 employees of the Ministry of Education between 2018 and 2020. The first element of the program, designed to eliminate 'ghost' workers, required employees to register for a mobile money wallet with biometric identification. This helped eliminate ghosts from the payroll (1.3% of employees), and allows estimation of the true share of ghosts (8.4% - 20.4%). The second element transitioned employees from receiving their salary in cash to receiving it via direct mobile money transfers. This led to a 26 percentage point increase in support for the reform to be scaled nationally, reduced salary delays, and caused employees to dramatically increase activity on the mobile money network, demonstrating a potential pathway toward expanding Afghanistan's formal financial system. Because the experiment spanned both secure cities and contested rural regions, we can examine whether state control complemented the reform. While the reform brought benefits everywhere, improvements materialized faster in cities. Our results highlight the importance of long-term horizons in state-building efforts and provide evidence that progress toward a modern bureaucracy is possible, even in the shadow of war and while the broader state is under threat.

Authors' Note: We are grateful to the many devoted public servants in 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, and the dedicated staff at Afghan Wireless and at Roshan for their partnership. We thank Eddy Chebelyon, Shahim Kabuli, Jiyoung Kim, Ofir Reich, and Sami Safiullah for excellent research assistance and Aarya Nijat and Kelsey Reiff for outstanding project management. Oriana Bandiera, Eli Berman, Gharad Bryan, Craig McIntosh, Karthik Muralidharan, Paul Niehaus, Rohini Pande, Jacob Shapiro, Noam Yuchtman and many others provided insightful feedback. We thank audiences at Bocconi, CERGE-EI, CEPR, UCSD, Y-RISE, FCDO, ESOC, PacDev, Queen Mary University London, MWIEDC, 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 for randomized control trials under trial number AEARCTR-0001641. 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

[§]University of California at San Diego, Department of Economics, La Jolla, CA 92093. afaikina@ucsd.edu

[¶]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

In fragile countries, where the World Bank forecasts two-thirds of the world's extreme poor may reside by 2030 (Corral et al., 2020), building state capacity is challenging but fundamental for political stability, economic development, and advancing human welfare. Correspondingly, rich countries have spent trillions of dollars in the 21st century aiming to catalyze institutional development in fragile states.[1] A core element of these efforts is modernizing the civil service. State employees implement most essential functions of the state – 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).

Current trends make it clear that ending global poverty means ending it in the world's most fragile states (Pande and Page, 2018; Commission on State Fragility, Growth, and Development, 2018). This raises the question of whether it is possible to build institutions while control of the state remains contested. It is thus critical to understand whether establishing a monopoly on violence must precede the development of a bureaucracy capable of delivering services. This paper reports on an effort to use technology to register and pay civil servants in Afghanistan – the highest-profile recent example of an international effort to build state institutions – conducted across secure and contested regions. Afghanistan's varied landscape set the stage for a near-constant policy debate about whether state-building efforts should focus on easier gains in areas under control or target more contested areas. Secure areas tended to be urban, easier to access and monitor, with more developed market institutions. Contested areas were exclusively rural, with much more limited infrastructure and institutions, and which required state actors to negotiate with local leaders to implement policy.

Our analysis centers around a randomized evaluation of a reform implemented by the Afghan government that was designed to improve state capacity to serve two essential bureaucratic functions: to identify government employees and to pay them. The reform we evaluate was implemented by Afghanistan's Ministry of Education in 2017-2020, and included an experimental sample of just over 30,000 public employees (primarily teachers) working in roughly 1,500 schools across three large provinces. The Afghan government selected these three provinces precisely because they exhibited substantial heterogeneity in the degree of state control and thus permitted a test of whether state control complements efforts to modern-

[1] With its focus on formal institutions, state-building is distinct from 'nation-building', which emphasizes the development of national identity. While the highest-profile cases of U.S.-supported state-building efforts include Germany and Japan after World War II and Afghanistan and Iraq in the post-9/11 era, Lake (2010) identifies more than a dozen other relevant cases since 1900.

ize the civil service.[2] In addition, the Afghan government opted to focus on the Ministry of Education because it employs roughly 80% of the country's civilian public servants, is important to the country's development, and because the status quo payroll system was plagued by inefficiencies.[3] International donors believed paying government salaries to be a fundamental part of keeping the Afghan state intact against the ongoing insurgency, and so covered about two-thirds of the state's total wage bill.[4] Afghanistan, therefore, is an ideal setting to study whether state control complements efforts to modernize the civil service.

The reform itself involved two main components. First, employees were required to register for a mobile money wallet and to provide their fingerprints for biometric identification. Second, employees began to receive their salaries via direct mobile money transfers, instead of the status quo system that relied partly on a network of 'trusted agents' to deliver cash payments. We refer to these two steps jointly as the Mobile Salary Payment (MSP) reform. The reform had the following objectives. First, eliminate ghost workers (fake employees added to the payroll so that others can capture their salary) by requiring employees to provide their biometric data and register for a mobile wallet in order to continue receiving their salary. Second, eliminate leakage by creating an end-to-end electronic record of distributed funds. Third, reduce delays, travel costs, and generally improve the quality of salary payments. And last, promote financial inclusion for workers and broader adoption of mobile money – which exhibits substantial network effects (Katz and Shapiro, 1994) by creating a new and substantial number of users. Prior to the reform, employees had very limited formal savings, with only 5% saving in bank accounts. Afghanistan's rough geography and persistent insecurity make it an especially challenging environment to create a financial system; the government hoped that this reform could kickstart financial development.

Our design involves two treatment arms, allowing us to study the two 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 in terms of the

[2]The MSP reform was led by Afghanistan's Office of the President, the Ministry of Communications and Information Technology, Ministry of Finance, and Ministry of Education, with financial and technical support from USAID and the World Bank.

[3]For instance, in a series of unannounced school visits we conducted prior to the MSP reform, almost half of the recorded teachers were not present despite having been paid in recent months. Issues with teacher salary payments were similarly severe: at baseline, only 28% of employees in our control group reported being very satisfied with their payment system and 54% experienced payment delays.

[4]The Ministry of Education 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 annual 3.4 billion USD recurrent budget on salaries, with roughly two-thirds funded by international donors (World Bank, 2019).

planned timing of implementation: (i) an early payment group who were both registered and paid digitally as soon as practicable (hereon "early"); (ii) a late payment group, who were registered at the same time as the first, but paid with a six month delay (hereon "late"); and (iii) a control group that were meant to be both registered and paid after the experiment (hereon, "control").[5] This design yields four sets of results corresponding to the four policy goals enumerated above.

First, 2.8% of employees never registered, even though all employees understood that failing to do so would likely mean the loss of future salary payments. We complement the registration exercise with a "litmus test" survey, which tests knowledge of facts that any real employee should know (e.g., the name of the school's principal). This exercise was meant to identify potential stand-ins who submitted biometric data to keep salaries flowing but likely never worked as teachers. A simple back-of-the envelope exercise indicates that the true share of ghosts is likely between 8.4% and 20.4% of paid employees.

Second, the leakage of salaries during disbursement – an object of major policy concern prior to the reforms – appears to have been a minor issue. The concern stemmed from the fact that trusted agents would carry cash to remote schools in order to distribute it and that this process was not easily audited. 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, despite widespread concern, did not appear to be a major source of corruption.

Third, the reform improved the salary experience of employees – but the effects were not immediate. Over the course of the reform, delays – which stand out as a fundamental issue plaguing salary payments – came down substantially (including in the controls) from affecting 55% to 30% of all employees. This happened primarily because a set of salary approvals which take place at the Ministries of Education and of Finance in Kabul – which necessarily impact all employees – were substantially reduced. In the control group, first mile delays go down from 41.7% to 27.6% after one year. In the treatment group, these first mile delays dropped further to only 13.7% after one year, reflecting a focus by bureaucrats in Kabul on making the reform work. By contrast, treatment induced a substantial increase in last mile

[5]The early group comprised 137 zones where payroll verification began in May 2018 and MSP payments began in October 2018; the late group comprised 129 zones where payroll verification also began in May 2018, but digital payments were scheduled to be delayed by six months with respect to the early group; and the control group comprised 135 zones where registration and payments were scheduled to be delayed respectively by four (September 2018) and six (April 2019) months.

delays in year one, primarily in rural areas where the mobile network operator lacked the staff necessary to convert digital mobile money transfers into cash. By year two, however, these issues were resolved. Correspondingly, while treatment increased employees' stated support to scale the reform across the ministry even in the first year, after two years a clear majority of employees in all treatment arms support rolling it out across the country.

Fourth, treatment causes a 30 percentage point increase in self-reported use of mobile money for peer-to-peer transfers, relative to about 6% of employees who have executed such a transfer in the control group. Moreover, using administrative data, we find substantial increases in several dimensions of mobile money use, including depositing funds into the account, using funds to purchase airtime, and using funds to make transfers to others using airtime (which does not require the receiver to have a mobile wallet). Correspondingly, mobile money balances increase by about 0.6% of salary for every month that the user is on the platform.

We also find clear heterogeneity in the effectiveness of the reform. Individual support to scale the reform goes up after one year in cities, but takes two years in rural areas. Payment delays increased after one year, especially in rural areas, but are reduced substantially after two years in cities. Mobile money use for applications other than receiving salary also go up most dramatically in urban areas. Many involved in this program – and comparable efforts over the last twenty years – articulated a view that salary payment modernization should begin in rural, contested areas. These arguments, made mainly by the Finance Ministry and leaders of Afghan banks, centered around the idea that state employees in the rich and relatively secure provincial capital were already served by banks and that salary issues and corruption were more severe in rural areas. A much smaller constituency – mainly those working for Afghan telecoms – argued the opposite, based on the idea that the logistics of executing such payments would be much simpler in urban settings. Cities have mobile money ecosystems and corresponding market density to support an agent network, and, perhaps most importantly, the security for mobile money agents to work as *de facto* agents of the state who help create a registry of employees and pay them.

Our data do not allow us to say exactly why the reform brings benefits more quickly to cities, or whether it is physical security *per se*, or the density, ease of logistics, mobile money networks, and other market-supporting institutions that state control supports. Nonetheless, a clear policy implication of this study is that state-building reforms like this one will meet more success earlier if they start in urban areas.

This study yields three key insights. First, the reform, which built on a decade of efforts to design mobile salary payments, first in the private sector (Blumenstock et al., 2015, Forthcoming), then in a small ministry in Kabul, and eventually across a much broader geography, achieved the important policy goals of cleaning the employee registry, reducing payment delays, and expanding digital financial inclusion. H.R. McMaster, whose involvement in the Afghan war culminated in serving as U.S. National Security Advisor, famously argued that Afghanistan was not a 20-year war but rather a one-year war fought 20 times over. His comment highlights the idea that the U.S. strategy shifted on about an annual basis, reflecting personnel rotations, American political cycles, and the need to find agendas that could show almost instant progress. More recently, former CIA Director David Petraeus argued that the U.S. forces aiming to build a state in Afghanistan attempted to do "too much too quickly" (Petraeus, 2022). These comments resonate with our results: had the reform been called off after only one year, it would have appeared a failure. Second, as mentioned, the program bore fruit sooner in secure, urban areas, but eventually also appeared to make improvements in rural areas. This is consistent with the idea that state control complements (but is not necessarily a prerequisite for) modernizing the bureaucracy. Third, much of the popular wisdom about where corruption concentrates is not obviously correct. We did not find large numbers of ghost workers, and even our upper bound estimates lie below the extreme levels many expected. Correspondingly, salary leakage is trivial. By contrast, the major issue affecting employees are delays in approvals by bureaucrats at central ministries in Kabul, who face limited accountability to front line employees.

A broad literature discusses the complementarities between state control and efforts to build a broader state capable of delivering services (cf. Besley and Persson, 2011; Acemoglu and Robinson, 2012; Commission on State Fragility, Growth, and Development, 2018; U.S. Army, 2006). However, there is less rigorous evidence on how fragile states can rebuild, particularly in the context of international intervention (cf. Fearon and Laitin, 2004; Weinstein, 2005). This paper partly addresses that gap while contributing to debates on the potential for technology to expand state capabilities (Muralidharan, Niehaus and Sukhtankar, 2016; Banerjee et al., 2020) and accountability in service delivery more generally (Duflo, Hanna and Ryan, 2012; Callen and Long, 2015; Callen et al., 2016; Dhaliwal and Hanna, 2017; Bossuroy, Delavallade and Pons, 2019; Muralidharan, Niehaus and Sukhtankar, 2020).

Indeed, our paper resembles Muralidharan, Niehaus and Sukhtankar (2016) in that it explores the potential to use technology to build core state capacities (in the case of that paper, building rosters of and paying citizens, and in ours, building lists of and paying government

employees). Our paper is distinct in three important regards. First, we focus on an earlier stage in state modernization: organizing and incentivizing state employees, which is a critical first step in creating a modern state. Second, the Afghan government was at a more basic stage of development than the Indian state of Andhra Pradesh. Third, our experiment takes place during war in a setting where the monopoly on the use of violence was contested by a foe that would ultimately prevail. In addition, our paper relates 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). We also relate to the literature on experiments at scale (Muralidharan, Niehaus and Sukhtankar, 2016; Muralidharan and Niehaus, 2017) and 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). Last, our experiment provides evidence directly relevant to the question of whether externally-supported state-building is possible at all in the geopolitically important case of Afghanistan (Callen and Kabuli, 2022).

The rest of the paper proceeds as follows. Section 2 provides background and summarizes policy debates related to the reform. Section 3 describes the MSP reform, our research design, and administrative and survey data sources. Section 4 presents results on biometric registration and ghost employees. Section 5 presents results on the effects of MSPs on the salary payment experience, financial inclusion, and education quality. Section 6 concludes.

2 Background

Afghanistan consistently ranks among the five most corrupt countries in the Transparency International corruption perceptions index. Even after 20 years of substantial levels of international support, the Afghan government failed to develop many key competencies. Regarding salaries, delays in payment, garnishments, and payments to ghost employees – fake employees added to the payroll so that others can capture their salary – were widely thought to be common. Correspondingly, some reports estimate the share of ghost workers on government payrolls to be as high as 40 percent. Outside of this study, the best estimate of which we are aware is that 13.3% of the Afghan National Defense Security Forces, which includes the police and the army, were likely ghost workers in 2018.[6] These estimates, however, were contested by the Minister of Defense at the time. Despite a surplus of anecdotes illustrating

[6]This estimate is based on the change in ANSDF personnel from 314,242 to 272,465 after the introduction of the biometric Afghan Personnel Pay System (APPS). The approach to obtaining this estimate, therefore, is quite similar to the approach used in our two experiments.

these issues, there are very few systematic efforts to either measure the true magnitude of payroll leakage, or to correct the underlying issues.

2.1 Key Policy Debates

Both the Afghan government and the international donors who funded the majority of government salaries viewed developing effective systems for paying and for tracking employees as fundamental. Still, various stakeholders engaged in several intensive policy debates.

Work in controlled or contested territory? A general debate throughout the Afghan war was whether the Afghan state and its international partners should focus more effort and resources in cities or in the countryside. The idea that state control should first be consolidated in the city before branching into the countryside is reflected in the 'Clear, Hold, and Build' doctrine articulated in the US Army's manual on counterinsurgency (U.S. Army, 2006), and was also reiterated by leading counterinsurgency scholars in the 'ink spot' or 'oil spot' strategy (Krepinevich, 2005). Similarly, a fundamental insurgency insight, attributed to Mao Zedong, is to 'encircle the cities from the country-side' (Mao, 1939).

This logic is reflected in the earliest initiative to pay government salaries using mobile money, which started in 2009. Government officials, especially those working at the Ministry of Finance, argued that such reforms should begin in rural areas. The logic is that these are the areas plagued by corruption, under threat by the Taliban, and most in need of investments in state presence. In addition, these areas are not already served by banks, and so the reform might create the largest increases in financial inclusion and do the most to reduce corruption.

However, rural areas pose major challenges for Mobile Network Operators (MNOs). There are few existing mobile money agents, not enough mobile money business to sustain agents, gaps in phone coverage, and insecurity making agents vulnerable to theft. Correspondingly, for agents to be willing to gather and distribute enough cash to cover a corresponding set of digital payments, they demand a high premium.

Focus on 'first mile' processes in Kabul or 'last mile' processes in the field? There are two broad reasons why salaries are delayed. First, the can be delayed by slow-moving 'first mile' approvals and administrative steps that must be taken in the Ministry of Finance, the Ministry of Education and then by Kabul Bank, all in Kabul. As government salaries provides the main source of liquidity for Kabul Bank, there exist incentives to move slowly. Moreover, bureaucrats in Kabul face almost no accountability to front line employees. The

second reason for delays are 'last mile' issues. After salaries are approved and deposited in employees' bank accounts, they either need to be withdrawn (for employees paid by banks) or withdrawn and physically transported as cash (for employees paid by Motameds). Importantly, first mile delays are centralized in Kabul and so should affect employees similarly, while last mile delays occur in the field, and should vary by geography. Moreover, fixing 'first mile' issues requires reforming payment approval processes in the ministries in Kabul, while improving 'last mile' issues requires changing the payment technology. At baseline, we find 84.1% of delays due to first mile actors and 21.5% to last mile actors.

Can the government use a 'Big Push' approach to create a mobile financial ecosystem? Mobile money is subject to substantial direct network effects (Katz and Shapiro, 1994). Afghanistan was the second country in the world to get a mobile money service, after Kenya. Indeed, the dramatic adoption, and corresponding financial opportunities created in Kenya (Jack and Suri, 2014, 2016), motivated foreign donors to help establish a service in Afghanistan. However, a key advantage of Kenya's service is that the mobile network provider who launched the service (Safaricom) already had an 80% market share, allowing network effects to realize quickly. First generation mobile money services were not 'inter-operable', so pre-existing market share played a key role in determining success.

The Afghan government thus 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.

3 Institutional Details, Research Design, and Data

The primary goals of the reform are: (1) removing ghost workers from the payroll; (2) reducing the extent of leakage in salary payments; (3) improving the salary payment experience; and (4) promoting financial inclusion through the creation of a mobile money ecosystem.

The research design aims to measure the extent to which the reform achieved these goals. We begin by providing details of the reform. We then describe the sample of MoE employees affected by the reform, treatment assignment, and treatment compliance. Finally, we describe the survey and administrative data used to evaluate the reform.

3.1 Details of the MSP Reform

The MSP reform involves three stages: (i) registering employees in mobile wallets and thereby auditing the payroll (registration); (ii) having the MoE determine who should remain on the payroll after registration (adjudication), and (iii) the processing of payments using mobile money (payment). The MoE implemented the reform using a randomized and staggered rollout. Below are details regarding each of the 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 be found in the field. 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 Ministry of Education. 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 ghosts, it was up to the MoE to ultimately remove them.

Stage 3 - Payment: The list of verified employees created during the adjudication step was transmitted to the Mobile Network Operator (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.

3.2 Creating the Experimental Sample

The population for this study is all MoE employees working in Kandahar, Nangarhar, and Parwan provinces. We identify this population based on official government payroll records from March 2017 to January 2018, containing information about 34,666 employees working in 1,829 schools.[7] We focus on March 2017 to January 2018 as implementation began in

[7]The payroll records 1,921 schools, but 92 schools had no employees paid at all between March 2017 to January 2018. In our experimental results, we also exclude a sample of 29 schools in Nangarhar that were

March 2018.

The experimental sample is determined as follows. First, we divided 1,829 schools into 501 distinct registration zones using the information on schools' location, number of employees, and security. We removed 100 'buffer' registration zones from the experimental sample in the next step. These zones are, by design, less secure, do not have typical schools (e.g., schools are madrassas, vocational training institutes, or other non-traditional schools managed by MoE), or have ten or fewer employees.[8]. Note that these criteria were not rigidly and uniformly applied in creating the buffer. Our experimental sample still contains insecure regions, some non-traditional schools, and some small schools. We created buffer zones so that the implementing partner could iteratively work out and adapt both the process of biometrically registering employees and the logistics of managing agents to ensure that teachers could withdraw their salary on payday.

Figure A.1 provides an aerial view of a typical urban (Panel A) and rural (Panel B) registration zone. The urban zones are typically much smaller and comprised of a smaller number of schools, as there are more students per school in those areas.

Our final experimental sample consists of the 32,398 MoE employees paid at least once between March 2017 and June 2019 in one of the 1,516 schools belonging to the 401 experimental registration zones.[9] For each registration zone, we also randomly selected one school for full listing exercises of all the employees present at the school during an unannounced audit visit and extensive surveys of a sub-sample of present employees (up to 3 in each school).

3.3 Assignment to Treatment

The upper portion of Figure 2 depicts the timing of the reform. The basic design for the evaluation randomizes the timing of the payroll verification process and the implementation of MSPs payments at the registration zone level. Here, we describe the design as conceptualized at the beginning of the reform. After the buffer zones were removed, the remaining 401 zones were assigned to one of the three treatment arms:

1. Early registration, early payment (hereafter EE): 137 zones where payroll verification was scheduled to begin in May 2018 and MSP payments were scheduled to begin in

part of an initial pilot of registration activities.

[8]We include the final condition because of the high fixed cost of surveying a school with only ten teachers

[9]Our pre-analysis plan mentions 1,535 schools, but there were no paid employees in 19 of them.

October 2018;

2. Early registration, late payment (hereafter EL): 129 zones where payroll verification was also scheduled in May 2018, but MSP payments were delayed by six months with respect to the EE group (April 2019);

3. Late registration, late payment (hereafter LL): 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. Because our implementing partners were under immense pressure to show progress on increasing the number of government workers paid via mobile money, we could not negotiate the delay of the initiation of payment in the second and third treatment arms beyond April 2019.

Treatment assignment is stratified at the district level. The three studied provinces contain 49 districts, and our experimental sample includes 42 of them.[10]

3.4 Treatment Compliance

The central and lower portion 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 from the central portion of Figure 2, the majority of employees in the EE and EL group were registered between May 2018 and July 2018. Most of the registration in the LL 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 displayed in the lower portion of Figure 2, employees in the EE group started to be paid by MSP in October 2018 in accordance with the schedule. However, due to insufficient number of agents, the transition of employees into MSP payment was gradual, and many of the employees in the EE group received their first MSP payment either in November or December 2018. During these months, in disagreement with the initial plans, about half of the employees in the EL group were also transitioned to MSP payments to show progress

[10]The number of zones per treatment is not equal due to the stratification: not all districts have a number of registration zones that is a multiple of three. The different sizes are due to the random assignment of the one or two "remainder" zones in such districts.

on increasing the number of employees paid via mobile money. Only a small minority of employees in the LL group were paid via MSP during the study period, starting mostly in February and April 2019. Finally, even after April 2019 the number of employees paid via MSP did not increase in the EL and LL group, reflecting the persisting capacity constraints of the MNO. In April 2019, 72% of employees in the EE group, 44% in the EL group, and 8% in the LL group reported being paid by mobile money. In May 2020, these numbers increased to 81%, 49%, and 33%, respectively.

While the compliance with the schedule predetermined by the treatment assignment was far from perfect, the randomization generated variation in the number of employees involved in the reform across treatment groups, as displayed in Figure 2 and Appendix Table A.4.

3.5 Survey Instruments

We have four primary survey instruments:

Listing Survey: In each school, we administer a short survey of all available teachers. The exercise begins with an unannounced visit where an enumerator lists all employees present at the school. The teachers are then asked basic questions regarding their identity, job, and salary experience (with a special focus on the timing of salary payments and leakage). The enumerator arrives at the school with a pre-compiled list (built from the payroll records) of employees recently paid for work at that school. During the visit, the principal is asked for an explanation for absent employees. The enumerator also takes pictures of key administrative documents, including attendance sheets. The listing survey was conducted three times: at baseline in May 2018 (before the registration activities started in the EE and EL groups), at midline in November 2018, and at endline 1 in April 2019 (before the MSP payments were scheduled to start in the EL and LL groups).

We randomly selected 401 schools, one per registration zone, for the listing surveys. However, the survey company could not visit all the schools due to security issues. They visited only 375 schools at baseline, 369 at midline, and 362 at endline 1. The total number of employees listed on the payroll and present during the audit visits is 6,264 at baseline, 5,793 at midline, and 5,214 at endline 1, respectively.

Full Survey: After administering the listing exercise at baseline and endline 1, we also randomly selected up to three teachers in each school for full surveys. They were asked

about their salary payment experience, employment history, teaching, and mobile usage. In total, we interviewed 1,005 teachers at baseline and 974 teachers at endline 1.

Additionally, we administered an extra round of the endline survey (endline 2) in May–June 2020 to analyze the medium-term effects of MSPs. We attempted to contact via mobile phone 945 employees whom we interviewed in-person during the first endline survey and for whom we obtained a valid phone number. We successfully interviewed 739 of them with a non-response rate of 21.8% using a short version of the endline 1 survey instrument.

'Litmus Test' Survey: We anticipated that some employees who appear at registration might be 'stand-ins,' who are not legitimate teachers such that ghost salaries are not eliminated. In February–March 2019, we conducted a 'litmus test' phone survey to distinguish real employees from 'stand-ins.' We sampled 2,663 employees, including those who did not register during the first registration wave and, for comparison, who registered with delay. They were asked for basic details about their workplace and job, such as the school's geographic location, the principal's identity, and the rank of their position. Real employees should be able to answer these questions easily, while 'stand-ins' might find some of them difficult.

Learning Assessment: In May 2019, we conducted a learning assessment of 1,001 children aged 6–10 who live in the proximity area of 367 accessible schools selected for the listing exercise. We administered a short household survey with an adult to collect basic information about the children within each randomly selected household. Afterward, one child in each household was randomly chosen to take the learning assessment. The instruments were adopted with the permission of the New York University's Assessment of Learning Outcomes and Social Effects of Community-based Education (ALSE) in Afghanistan. The learning assessment, modeled off the Early Grade Reading Assessment (EGRA) and Early Grade Math Assessment (EGMA), tested literacy and numeracy skills such as counting, number and letter recognition, addition, subtraction, reading, and reading comprehension.

3.6 Administrative Data

We also collected administrative data from the following two sources:

Payroll data (M-41): These data comprise the primary payroll records for the universe of employees generated when the Ministry of Education submits monthly requests for salaries

to the Ministry of Finance. These records contain employee-level information, including names, basic demographics, and monthly salary payment transactions.

Mobile Money Transactions data: These data record each payment issued via mobile money as a separate entry. These also contain information about salary availability, disbursement, and mobile money usage including deposits, remaining balances, peer-to-peer transfers, merchant payments, and airtime top-ups.

4 Biometric Registration and Ghost Employees

4.1 Identifying Ghost Employees

Before the reform, the government lacked a reliable list of employees in the Minisry of Education (henceforth, MoE).[11] Moreover, numerous reports indicated a high share of ghost workers, an impression that is supported by the widespread absenteeism of teachers in schools.

Several features of the registration process help to verify employees. First, the MoE provided clear directions to employees that they must appear for registration to continue receiving their salaries. This gave an incentive for legitimate employees to register while also potentially incentivizing corrupt actors to send 'stand-ins' to submit their biometric data in order to hold positions open. In addition, biometric identification required employees to provide all ten fingerprints. Cashing out salary payments after registration required biometric authentication. Therefore, adding a stand-in to the payroll required that this person appeared every month to cash out their salary.

Our data allow us to directly identify both employees who did not physically register and those who were eventually removed from the payroll. This provides a direct measure of the number of employees removed from the payroll as a result of registration and the associated cost savings.

Panel A of Appendix Table A.2 summarizes registration outcomes. Among employees in the early registration zones (i.e., in the early and late treatment group) who were paid in each of the six months before registration, 94% registered during the primary registration

[11] Specifically, the MoE maintained an human resources roster and a payroll, which differed substantially from each another.

wave, 3.2% registered with some delays, and 2.8% never registered.[12] Hence, a conservative, lower bound estimate for the number of ghosts is 2.8% of all employees. Their salaries amount to about AFN 3.4 million (USD 44,000 using the 2018/19 exchange rate) or 2.7% of the total monthly wage bill in these schools.

4.2 Removing Ghost Employees from the Payroll

After the primary registration wave, the MoE formed an adjudication committee tasked with removing ghost employees from the payroll. 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 the payroll records, we can track whether MoE continued to request salaries for employees who never registered. Panel B of Table A.2 (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 twelve months of the registration. However, even twenty 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.

4.3 Validating Ghost Employees

Between February and March 2019, we conducted a short "litmus test" phone survey with MoE employees. The purpose of the survey was both to validate whether never-registered employees were indeed ghosts and to explore the extent to which registered employees might have actually been 'stand-ins' registering, so that ghost salaries are not eliminated. We sampled a total of 2,663 employees who had a phone number listed in the payroll records. 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 Section B for further details on the sampling procedure).

[12]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 only among those who clearly appear as regular and stable employees according to the payroll records.

Litmus Test Design: We asked employees seven simple questions about their job: (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. We designed and piloted our litmus test in collaboration with the MoE in order to identify a set of questions that legitimate employees should be able to easily answer. The primary purpose was to identify uninformed stand-ins who may have more difficulty answering these questions.

Figure 3 displays the results. 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 ($p < 0.01$). This is the first indicator that some never-registered employees might indeed not be real employees (or even real people).

The remaining columns report the distribution of scores in the litmus test (conditional on participating in the survey). Employees who never registered are less knowledgeable than employees who registered: for example, they are 7.8 percentage points (s.e.= 2 pp) less likely to answer correctly all seven questions and 11.4 percentage points (s.e. = 2 pp) less likely to get a score of six, and are instead 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 (even though some of them might make mistakes), 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 is not perfect (true employees might make mistakes and stand-ins might answer correctly the simplest questions), 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.[13]

[13]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

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, 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 only in early registration zones in three provinces.

5 Mobile Salary Payments Impacts

The reform aimed to improve matters for employees in two key regards: improving the timeliness and quality of salary payments and expanding digital financial access. This section reports on whether the reform achieved these goals and checks for impacts on attendance and on student learning.

5.1 Impacts on Payment Quality

For the reform to survive – especially given the intense opposition of Afghanistan's banks described in Section 2 above – it needed to gain the support of employees by materially improving the payment process. This section reports estimates of impacts on delays, transaction costs, and general support for the reform.

Figure 4 depicts means of three main outcomes measuring salary payment quality separately for the early group and the control group at baseline, after one year in April 2019, and after two years in May 2020. Table 1 reports corresponding Treatment-on-the-Treated estimates.[14]

The first panel of Figure 4 reports 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?" The Government of Afghanistan intended the three-province pilot described in this paper to be a test run for nationwide scaling, and employees were informed about this when being recruited to registration. Em-

true employees who could be verified through the registration process and unannounced visits to schools.

[14] As we report in Appendix Table A.9, registration *per se* did not impact any outcomes related to the quality of salary payment (which, indeed, should only change when electronic payments comments), so we omit the late group from Figure 4. We also relegate closely related pre-specified outcomes to the appendix. Appendix Table A.5 reports regression estimates of impact for both the early and late groups comprehensively across all pre-specified outcomes and shows baseline balance. Appendix Table A.6 reports estimates from similar specifications, adding strata fixed effects.

ployees could therefore have perceived their responses to this question as consequential for whether the program continued. The government gave it substantial weight when considering the value of the reform.[15]

In May 2018, before the implementation of the reform, 94% of employees in the control group supported it. After one year, support fell precipitously to much lower levels, both in the control (42%) and in the early group (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). Corresponding ToT estimates in Table 1 indicate that being paid through mobile money caused a 17.8 pp increase in support for the reform after one year and a 26.3pp increase after two years.[16]

Delays stand out as a key issue affecting salary payments. Before the reform, 54% of employees in the control group experienced delays, with a median delay of 10 days. After one year, 41% of control employees, and 67% of early group employees experienced delays (Figure 4, middle panel). By the second year, delays in both groups came down to about 30%. In Section 5.3 below we provide more details about the effect of the reform on delays.

Travel time (Figure 4, bottom panel) shows a comparable pattern; the reform increased travel time in the first year, but not after two years. The corresponding ToT results in Table 1 report a similar pattern. Treatment increased the share experiencing some delay by 40.1pp (panel A, column 2) and travel times by 43.2 minutes (panel A, column 3) in the first year. In two years, by contrast, the ToT impacts on these measure are statistically indistinguishable from zero.

We also find very little evidence of leakage in last mile of salary payments. We measure leakage by employees' self-reported payments to someone to receive their salary. On average, employees report having paid less than AFN 25 (USD 30 cents, less than 1% of the average monthly salary) to receive their salary even prior to the reform.

[15]Nevertheless, it was made clear to employees that the question was hypothetical. At baseline, the question was formulated as "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."

[16]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. Appendix Table A.4 reports first stage results.

It is surprising that treatment increased support for reform during the first year, while also increasing delays and travel times to retrieve salaries. Speculatively, this may be because of the usefulness of mobile money itself, the use of which treatment increased rather substantially in the first year. Treatment increased the share of employees sending mobile money transfers by 29.3% relative to 6.3% in the controls (Panel A, column 4). In subsection 5.4, we describe more evidence that the reform created a clear increase in the use of mobile money.

5.2 Did the reform reduce salary delays?

As mentioned above, over the two years of the reform, delays were substantially reduced for all employees, including in the control group. The share of employees experiencing delays in the control group decreased from 54% (s.e. 0.04) to 41% (s.e. 0.04) and 32% (s.e. 0.03), respectively. Correspondingly, the share in the control group reporting satisfaction with salary payments went up from only 30% (s.e. 0.03) at baseline to 63% (s.e. 0.04) by year 1 and 73% (s.e. 0.03) by year 2 (Appendix Table A.5). Changes in the control group do not appear to be a direct spatial spillover from treatment zones.[17] 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, which we now describe.

As we discuss in section 2, salaries are subject to both 'first mile' and 'last mile' delays. Our data allow us to precisely separate first mile and last mile issues. Figure 5 reports these results. At baseline, 41.7% of the control group experienced first mile salary delays, while only 5.9% experienced last mile delays. By April 2019, first mile delays went down to 27.6% in the controls and fell even further to 13.7% in the early group. This corresponds to a treatment effect of 13.9 percentage points (s.e. = 3.1pp ; $p < 0.01$). By contrast, while last mile delays are flat in the control group, reflecting the fact that the payment modality did not change, they increase dramatically to 34.9% in the early group, reflecting an estimated treatment effect of 30.1 percentage points (s.e. = 2.9pp; $p < 0.01$). Moreover, these increases are entirely in rural areas, reflecting the fact that the mobile network operator running the reform faced challenges recruiting agents to disburse salaries. Appendix Figure A.3 provides evidence that the initial increase in delays partly stemmed from inexperienced mobile money agents, and that agent tenure subsequently increased during the period when payments improved.

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

The reform immediately reduced a set of first mile delays, especially in the early group, while raising last mile delays during the first year. This appears to reflect 'teething problems' in rural areas, where the mobile network operator took more time to develop an effective agent network. Reasonably, the subsequent system-wide reduction in delays in year 2, and the corresponding increase in satisfaction, reflects a new equilibrium whereby first mile delays are reduced, and last mile problems are also evened out across treatment arms. Therefore, it appears that the system-wide reduction in delays is attributable to the reform.

Importantly, the reductions in total delays is the result of the Ministry of Finance streamlining and improving it approvals process. Doing so does not require mobile money or biometric registries, but does require awareness of the degree of the issue and the political will to improve matters.

5.3 Does government control matter?

Section 2 describes the debate over whether the reform should begin in the city or in the countryside and about the importance of government control. Indeed, the government selected Kandahar, Nanagarhar, and Parwan as the three locations for the experiment precisely because they contained major government-controlled cities (Kandahar City, Jalalabad, and peri-urban areas of Kabul), peaceful rural areas, and rural areas actively contested by the Taliban.[18] One of the most important aspects of this experiment is that it operated at a time and place and at a scale that allows discussion of the importance of government control for efforts to modernize the state.

Insurgent conflicts are characterized by ambiguous areas of control. We use two measures of the territorial extent of government control during our experiment. Following the discussion in Wright (2022), we use a measure based on whether survey non-local enumerators can conduct surveys without the permission of local commanders as our primary measure, and we check the robustness to an additional measure created by the Long War Journal (LWJ).[19]

Table 2 reports ToT effects by urban areas, rural controlled areas, and rural contested areas. In year 1 the increase in delays caused by the reform were largest in rural contested areas, followed by rural controlled areas (panel A, column 2) with travel time also increasing more in rural areas (panel A, column 3). By contract, employees in all three areas become substantially more likely to report using mobile money (panel A, column 4). The reform

[18] Urban zones are those that have a population density of more than 300 inhabitants per squared kilometer.
[19] See Appendix Tables A.10 and A.11 for the analysis using LWJ data.

drove increases in support for the reform in urban areas, with no clear impact in rural areas (panel A, column 1).

By contrast, in year 2, support increased uniformly across the entire geography of the experiment (panel B, column 1), with delays going down in urban areas while possibly remaining elevated in rural areas (column 2) and travel time not being appreciably affected relative to the controls (column 3).

The difference in estimated impacts on delays and travel times across year 1 and year 2 indicate that working out the mechanics of delivering salaries, especially to rural areas, takes time, but that the reform can eventually reduce delays, especially in urban areas where the logistics of delivery are considerably easier.

5.4 Effects on Financial Inclusion

The reform required that employees use mobile money. Despite being the second in the world, after Kenya, to have a mobile money platform, adoption languished in Afghanistan. Mobile money exhibits substantial network effects (Katz and Shapiro, 1994). A key goal of the reform, therefore, was to move a critical mass of users onto a mobile money platform to kickstart the development of mobile financial infrastructure.

Before the reform, few employees used formal financial institutions. Even though 67% of employees report having a bank account at baseline, most used it solely to withdraw their salaries. Only 12% of employees send money via banks, 5% save in their bank accounts, and 3% have an outstanding bank loan. Hence, mobile money could serve as the first formal account enabling financial inclusion.

Survey data, captured in both the treatment and control, indicate that mobile money caused many employees to start making onward financial transfers via mobile money (Table 4 column 4). We complement this analysis using much more granular transaction records covering the 16,000 employees paid via mobile money from October 2018 to February 2020. These data record transactions for employees with a mobile wallet and so these employees come primarily from the early and late treatment groups.

Because we observe these data primarily in the treatment groups, we take a dosage response approach to estimate the causal effect of time spent receiving MSPs on other types

of mobile transactions. We estimate:

$$Y_{izd} = \alpha + \beta \; Months_{izd} + \mu_d + \varepsilon_{izd} \tag{1}$$

where $Months_{izd}$ is the number of months an employee i from the registration zone z in district d received MoE payments to his mobile money account at the time of the last transaction. β measures a change in outcome Y with an additional month of receiving salaries via mobile money.

To identify the causal impact of time spent receiving MSPs, we estimate the following 2SLS specification:

$$\begin{aligned} Y_{izd} &= \gamma + \beta \; Months_{izd} + \mu_d + \varepsilon_{izd} \\ \#Months_{izd} &= \theta + \varphi_{EE} EE_z + \varphi_{EL} EL_z + \eta_d + \epsilon_{izd} \end{aligned} \tag{2}$$

where we instrument the number of months, $Months_{izd}$, with treatment zone indicators, EE_z and EL_z. By design, employees in EL and LL zones started receiving payments via mobile money in later months compared to employees in EE zones. All outcomes are normalized by the total monthly received in the month prior to the start of the reform.

Table 3 reports impacts on mobile money use. Deposits comprise a particularly important measure of adoption, as this requires employees to add money to their wallet above what they receive for their salaries. In the first month of use, employees deposit a statistically insignificant 0.2 percentage points of their salary. The ToT estimates indicate this increase 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; by this point employees are depositing 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). Purchasing airtime (column 3) and purchasing airtime to transfer it to others (the crudest version of a mobile money transfer, allowing value to be transferred to others without a mobile money wallet) also increase considerably (column 4). Balances go up by 0.6 percentage points, or by 4.2 percentage points at the median.

Financial inclusion impacts by government control: Table **??** reports results estimates a similar specification that interacts $Months_{izd}$ with whether the sample is urban, rural and controlled, or rural and contested. The ToT estimates (Panel B) indicate that the financial inclusion benefits of the MSP reform are localized primarily in urban areas. More suggestively, their appear to be benefits in rural contested zones, possibly reflecting

the usefulness of mobile money as a secure means of storing and transferring value.

Taken together, these results indicate that mobile salary payments facilitate the expansion of the mobile money ecosystem and financial inclusion. Employees increase their mobile money usage by holding larger balances and sending more transfers, which are critical for further growth.

5.5 Effects on teacher attendance and on learning outcomes

The reform generated several substantial changes within the MoE. Conceivably, this could affect the quality of education that schools provide.

We measure the quality of education using two metrics. First, we follow the literature on teacher absenteeism in developing countries (Chaudhury et al., 2006) and look at the effects on teacher attendance. In particular, we measure the share of employees who were present during unannounced visits to schools. The unannounced visits conducted at Midline allow us to separately estimate the effects of registration since mobile payments were only starting in the treatment zones at that time. Panel A of Table A.1 shows that 52% of employees paid at Baseline were present during the Midline unannounced visits to control schools located in the Late Registration (LL) zones.[20] The attendance rates were slightly lower but not significantly different in treatment schools located in the treatment zones. Hence, biometric registration did not have clear effects on subsequent employee attendance.

Similarly, the unannounced visits to schools at Endline 1 allow us to measure the total effects of mobile salary payments on teacher attendance one year after the start of the reform. 46% of employees paid at Baseline were present during the Endline 1 unannounced visits to control schools, a comparable level to the treatment schools.

Second, even though teachers were equally likely to show up to schools, they could have changed their efforts or teaching methods due to changes in their salary payment experience. We measure the effects on students' knowledge by conducting an independent learning assessment of 1,001 children aged 6–10 who live nearby 367 experimental schools. Following Burde, Middleton and Samii (2017) who performed a similar survey in Afghanistan, we

[20]Note that these shares can underestimate the actual teacher presence rates in the visited schools. The listing exercise sample included both single-shift and multiple-shift schools, and unannounced visits sometimes spanned only part of the business day. There is also a natural turnover of employees across survey waves.

construct math, reading, and total test scores and standardize them relative to the scores of children living nearby control schools in the Late Registration, Late Payments (LL) zones.

Panel B of Appendix Table A.1 presents the results for students' learning outcomes. The effects of mobile salary payments on students' test scores are positive, but not statistically significant. Children living in the catchment area of treated schools in early (late) zones have a 0.05 (0.06) standard deviations higher test score than the control zones. This difference reflects slightly positive but insignificant effects on math and reading test scores.

Collectively, we do not find evidence that biometric registration or mobile salary payments impacted the quality of education during the first year of the reform. Importantly, the teething problems of the reform in the first year do not impact teachers' attendance or learning outcomes. It is plausible that improved payment experience during the second year could have positively affected the quality of education in the longer term. It was not feasible to conduct unannounced visits to schools or learning assessments at the time of Endline 2 due to COVID-19.

6 Conclusion

Over the last two decades, the war in Afghanistan carried tremendous financial and human costs. The U.S. spent 2.3 trillion dollars, and around 176,000 people, mostly Afghans, were killed (Brown Watson Institute, 2022). In addition, at the time of writing, more than half of the country's population – an estimated 22.8 million people – face life-threatening food insecurity as the economy crumbles (UN WFP, FAO, 2021), and many of the advances in human rights achieved during the previous 20 years have been quickly reversed. An unprecedented international effort to modernize Afghan institutions seems to have ended very near to where it started: with Afghanistan under the control of a brutal Taliban theocracy.

It is hard to imagine a scenario that could more starkly call ideas about whether and how to build states into question. Indeed, it makes a case that any such effort, no matter how it is executed, may not be worth the phenomenal financial and human costs. The record of international efforts to transform Afghanistan, by Americans, Soviets, the British, and even Alexander the Great, further suggests that the failure of external intervention is inevitable.

The experiment reported in this paper reveals some small glimmers of hope regarding the feasibility of building fragile states, and a corresponding set of lessons related to the broad

questions of whether and how to proceed. The first lesson regards the broad question of whether state control must necessarily precede the development of a bureaucracy capable of delivering services. Our results indicate that stability and state control certainly make professionalizing the bureaucracy much easier, but that it remains possible (although challenging) to develop service delivery functions even in a contested space.

Other lessons are more practical. First, many popular notions of where corruption would concentrate, especially high rates of ghost workers and substantial leakage of salaries, are not supported in our data. Structured qualitative interviews with just over 100 administrators, teachers, journalists, and community leaders reinforce this: rural elites did not benefit from, and indeed were frustrated by poorly functioning schools. By contrast, the salary delays that plagued employees had to do with processes in Kabul, not the last mile of delivery. It is widely argued that the substantial sums spent to support governments in conflict-affected countries backfire and blunt incentives to create self-sustaining states (Dercon, 2022; Easterly, 2007). Afghanistan provides a key case in point. Yet, the mechanics of this unintended consequence are not easy to understand. A lesson in our data is that rigorous measurement of these distortions is not easy, and often not in line with perceptions based on anecdote.

Second, early signs of success took years, not months. The 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). Moreover, Afghanistan is a country of rich variety, with stark differences between (at the time of the experiment) relatively safe and prosperous provincial capitals, unstable areas with active conflict, and quieter rural areas where the Taliban administered core state functions. Correspondingly, the possibilities for taking even the basic first steps toward creating a state, and the corresponding timelines, were always likely to vary.

Third, a reform that was an object of intense focus by senior Afghan policymakers did eventually improve salary payments, even though education quality did not clearly improve during the first year when it was possible to measure. In a closely related domain, the data reveal the green shoots of a digital financial economy. Employees began to use their mobile wallets to save and to make transfers. This suggests that such a reform, at scale, could help create a digital financial ecosystem in Afghanistan.

Even after the Taliban takeover, digital payments remain important in Afghanistan. The scarcity of hard currency and the unreliability of banks is prompting the adoption of broadly

interoperable smartphone-based financial apps. Donors are now channeling both urgently needed humanitarian assistance and government salaries using digital payments. Partly, this is of necessity, and partly it is meant to avoid capture by a hostile regime.

The experiment also points toward several areas for improvement in future iterations. First, registration needs to include additional layers of screening, beyond biometrics and along the lines of our litmus test, to filter out stand-ins. Second, relevant government authorities need to have the capacity and the political support to take more decisive action when ghosts and other forms of corruption are identified. Third, agent networks need to densify, perhaps with the government's support, to improve service quality and fully realize the financial inclusion aspects of the reform. Fourth, the early policy perspective that such reforms should start in rural and insecure areas needs to be revised. The traditional trusted agent payment system worked better than banks, and rural areas lacked the security, density, and other market-supporting institutions needed to catalyze mobile money agent networks. And last, such reforms require the government and mobile network operators to develop a set of 'back end' capacities to deliver on their promise. Indeed, most of the reductions in salary payment delays that occurred during this reform happened because of improvements to first mile procedures in Kabul, which necessitate no change in payment technologies. However, we find suggestive evidence that focusing on these capabilities can yield a better payment process for all government employees, including those paid by traditional means.

While certainly incomplete and constrained by the broader challenges in Afghanistan, the progress we document on reforming government salaries is noteworthy. Given our findings, pessimistic narratives that dismiss fragile states as hopelessly corrupt, inept, or mired in historical complexity should be revisited. Instead, the feasibility of state-building in fragile contexts will likely depend directly on the prioritization, patience, and context-sensitivity of institutional reforms. For external actors seeking to promote development and stability in such settings, the challenge is working through and not around the state, and avoiding the temptation to pursue quick impacts at the cost of long-term sustainability.

References

Acemoglu, Daron, and James A Robinson. 2012. *Why nations fail: The origins of power, prosperity, and poverty.* Currency.

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.

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.

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 New York.

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.

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. Forthcoming. "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*.

Brown Watson Institute. 2022.

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."

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, and Shahim Kabuli. 2022. "Three Sins: The Disconnect Between *de jure* Institutions and *de facto* Power in Afghanistan." *London School of Economics Public Policy Review*, 2(3).

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.

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.

Corral, Paul, Alexander Irwin, Nandini Krishnan, Daniel Gerszon Mahler, and Tara Vishwanath. 2020. *Fragility and Conflict: On the Front Lines of the Fight against Poverty.* The World Bank.

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.

Easterly, William. 2007. *White Man's Burden: Why the West's Efforst to Aid the Rest Have Done So Much Ill and So Little Good.* Oxford University Press.

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

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.

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

Katz, Michael L., and Carl Shapiro. 1994. "Systems Competition and Network Effects." *Journal of Economic Perspectives*, 93–115.

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

Lake, David. 2010. "The practice and theory of US statebuilding." *Journal of Intervention and Statebuilding*, 4(3): 257–284.

Mao, Ze Dong. 1939. "The Chinese Revolution and the Chinese Communist Party."

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, 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. 2020. "Identity Verification Standards in Welfare Programs: Experimental Evidence from India." *NBER Working Paper*.

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

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

Petraeus, David. 2022. "Afghanistan Did Not Have to Turn Out This Way." *Atlantic Monthly*.

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.

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. Peter B. Evans, Dietrich Rueschemeyer and Theda Skocpol, 169–191. Cambridge:Cambridge University Press.

UN WFP, FAO. 2021.

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.

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. 2022. "The Economic Origins of Territorial Control."

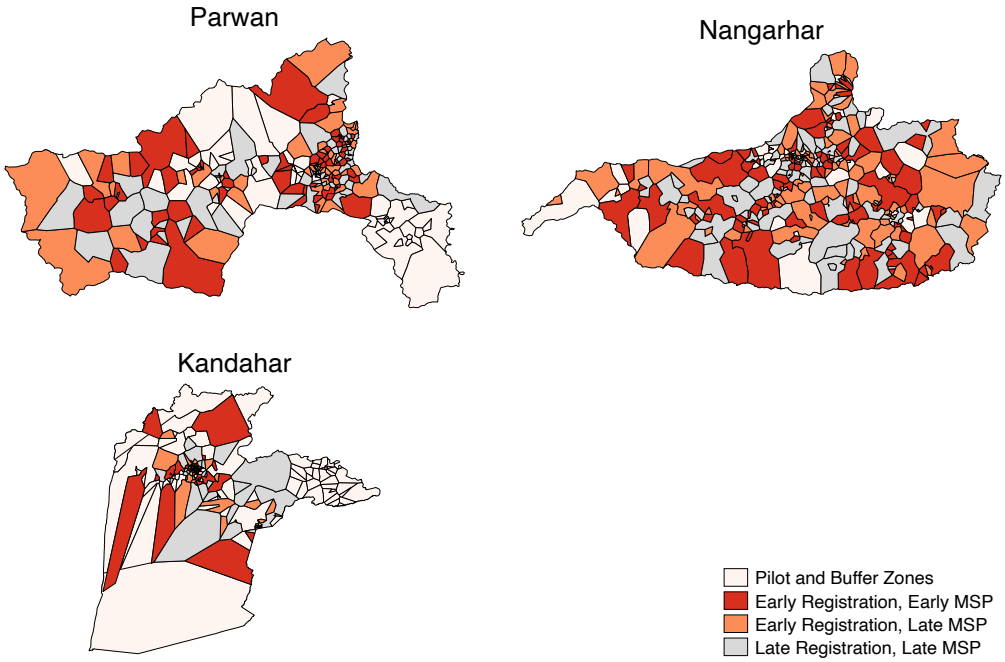
Figures and Tables

Figure 1: Map of Provinces and Registration Zones

Panel A: Studied Provinces of Afghanistan

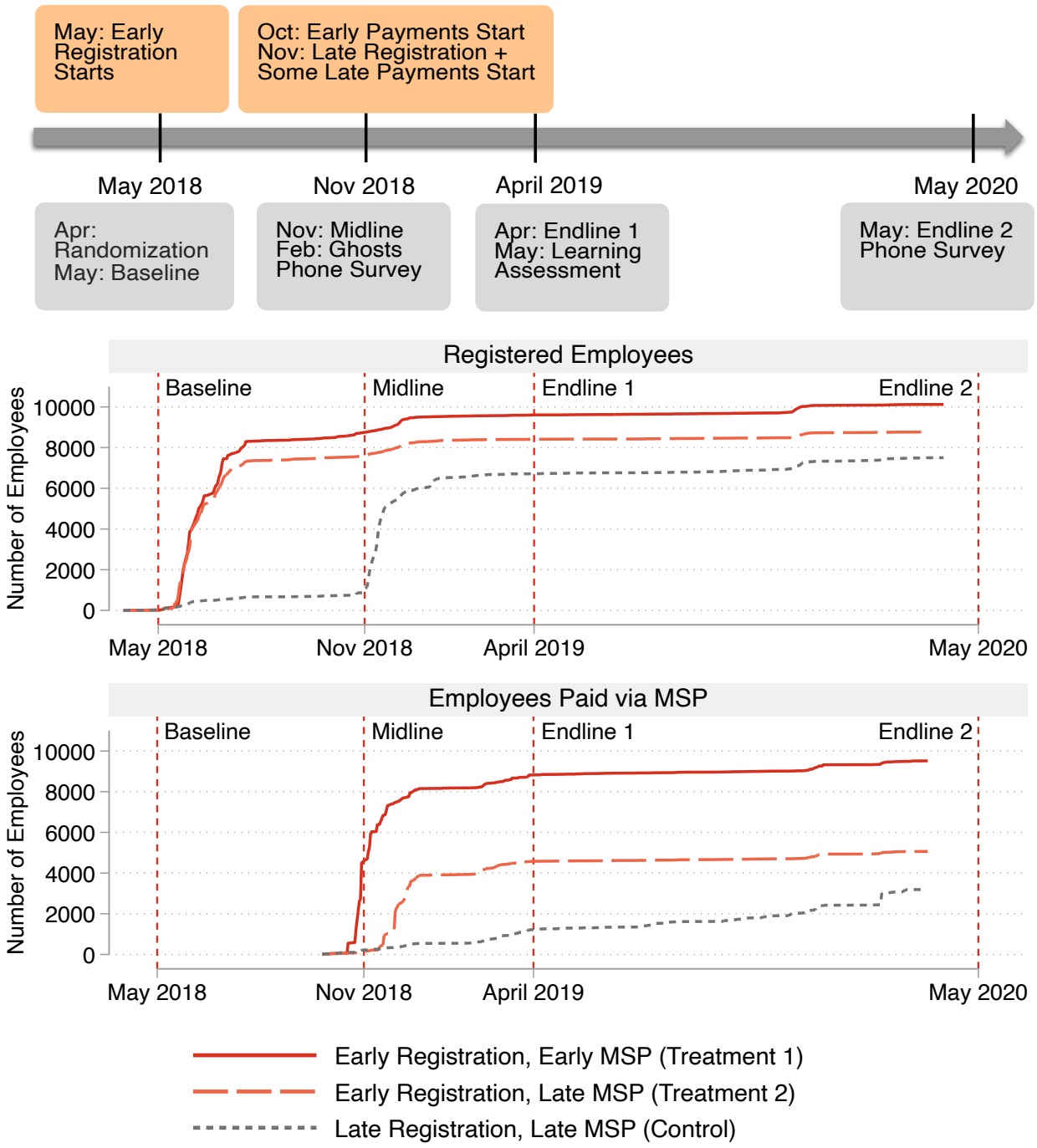


Panel B: Registration Zones and Treatment Status



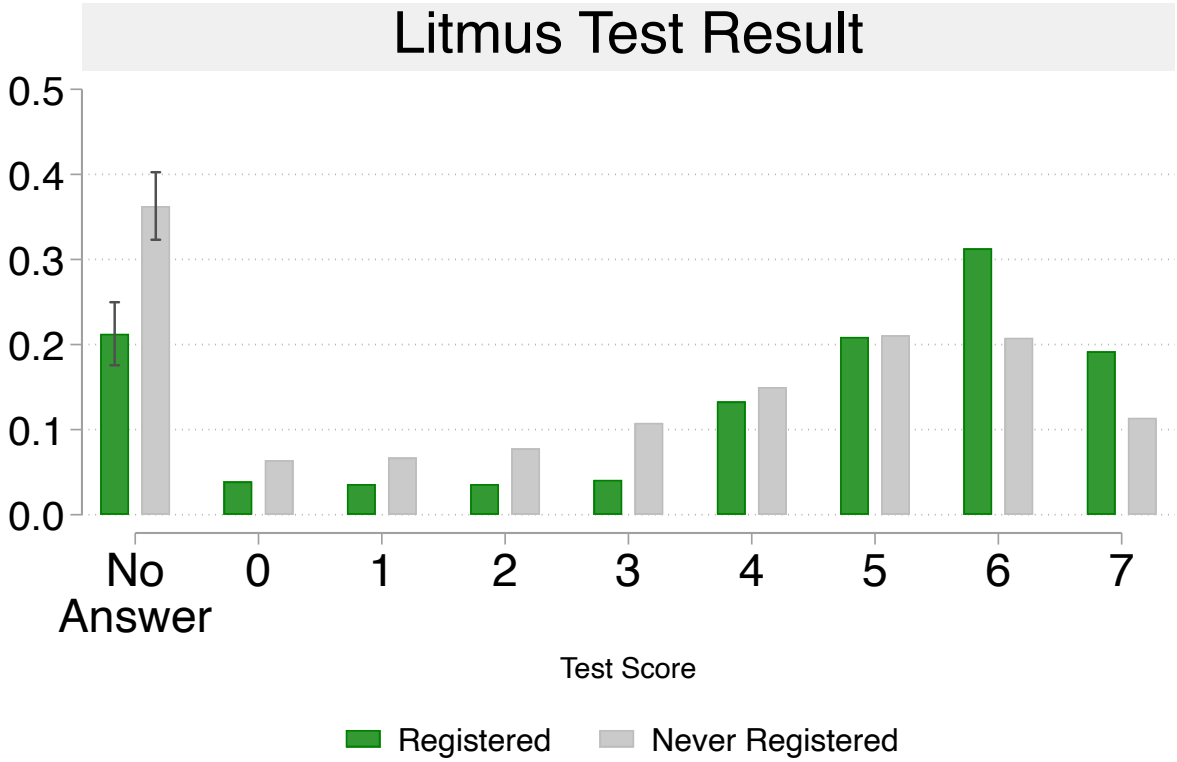
Notes: Panel A plots the map of Afghanistan, with three studied provinces highlighted in red. Panel B plots registration zones colored according to the treatment status in studied provinces. Registration zones are designed to encompass approximately 50 teachers and are created by a spatial grouping algorithm. 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



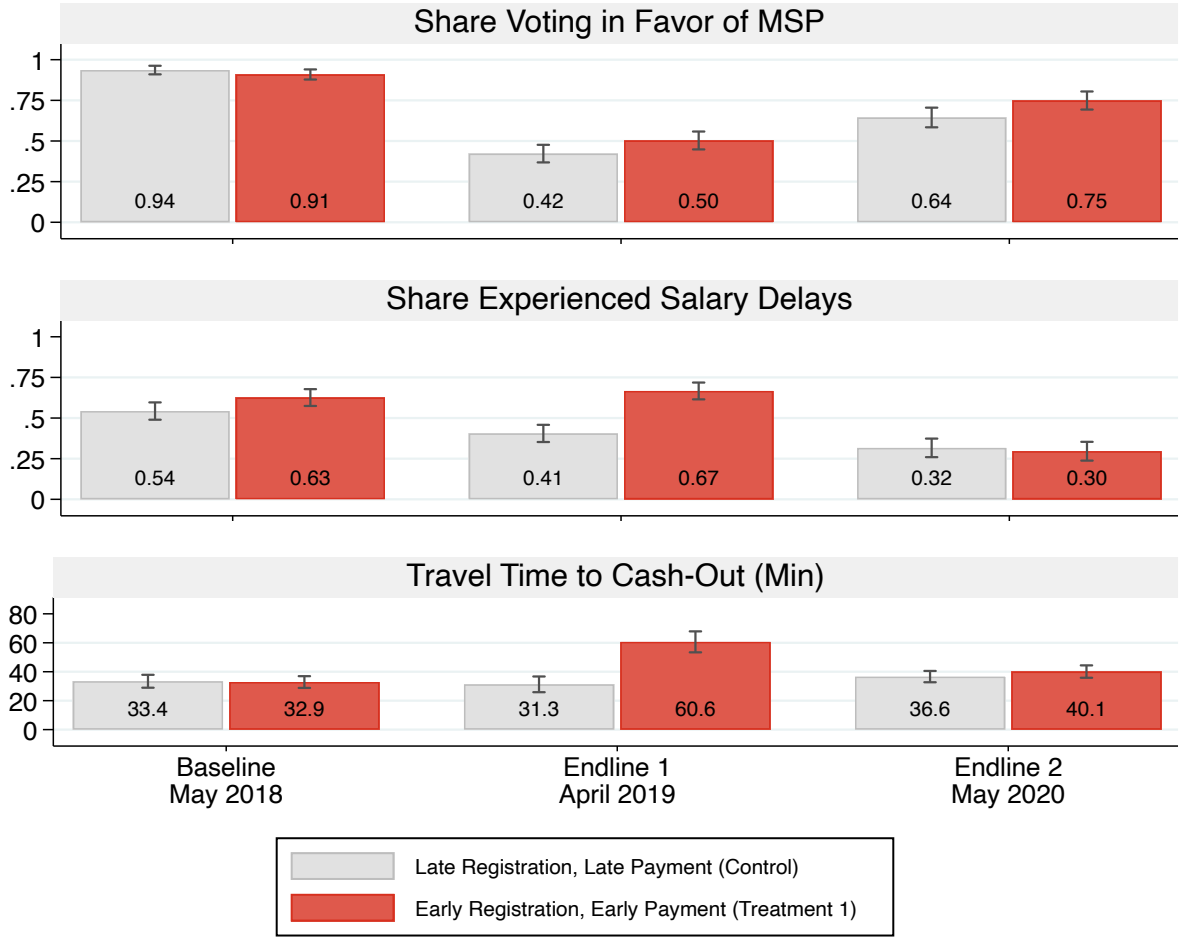
Notes: This figure plots the timeline of project implementation. The top panel presents the dates of the main milestones: the start of registration and payments in different treatment groups and administered surveys. The center panel presents the cumulative number of employees registered using administrative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



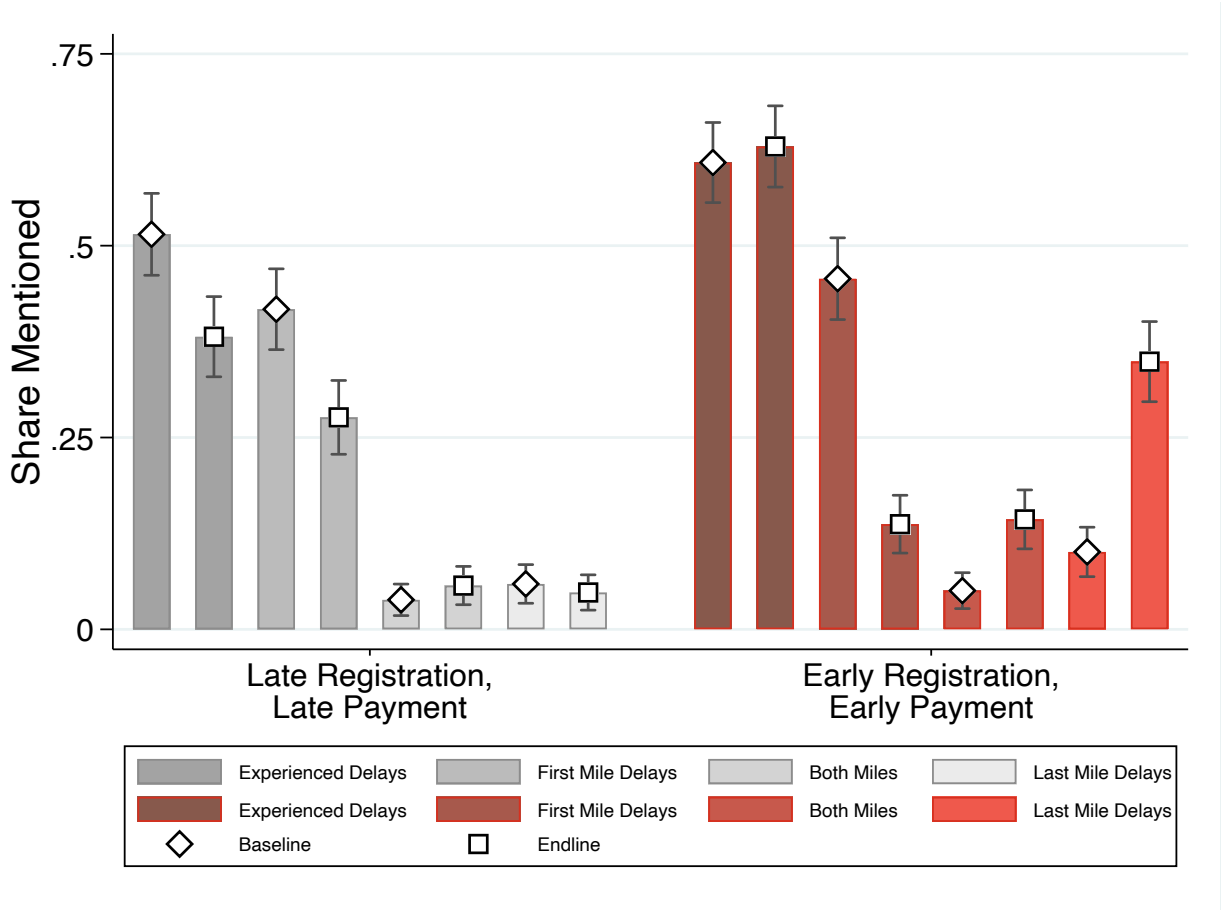
Notes: This figure presents the results designed to confirm whether employees knew basic facts about the schools where they were assigned. The sample consists of 2,545 employees from Early Registration zones, selected for the Litmus Test in February of 2019. The litmus test score is defined as the total 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 where 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.5 reports the corresponding treatment effects for both the Early Registration and Early Payment, Early Registration and Late Payment treatment arms estimated without stratum fixed effects for all outcomes specified in our pre-analysis plan and Appendix Table A.6 shows corresponding estimates with stratum fixed effects.

Figure 5: Separating First Mile and Last Mile Delays



Notes: This figure depicts the cause of delays. First mile delays are those due to the Ministry of Education, the Ministry of Finance, or the Central Bank (Da Afghanistan Bank). Last mile delays are those due to New Kabul Bank or Motameds. 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 in May 2020.

Table 1: Treatment on the Treated Estimates of Impacts on Payment Quality

	Favor MSP (1)	Experienced Delay (2)	Travel Time (3)	Money Transfer (4)
Panel A. Year 1 Outcomes				
Mobile Salary Payment (=1)	0.178***	0.401***	43.208***	0.293***
	(0.067)	(0.073)	(8.496)	(0.048)
Control Mean	0.422	0.405	31.276	0.063
Observations	950	969	922	965
R squared	0.334	0.269	0.273	0.149
Panel B. Year 2 Outcomes				
Mobile Salary Payment (=1)	0.263***	–0.059	3.755	
	(0.080)	(0.081)	(5.937)	
Control Mean	0.645	0.316	36.628	
Observations	710	733	723	
R squared	0.268	0.119	0.090	

Notes: This table reports on the impacts of mobile money salary payments on measures of payment quality. Favor MSP is a dummy variable equal to one if the respondent indicates they would support scaling the reform across the Ministry of Education. 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. 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. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors clustered at the registration zone (treatment unit) level.

Table 2: Impacts on Payment Quality by Region and Security

	Favor MSP (1)	Experienced Delay (2)	Travel Time (3)	Money Transfer (4)
Panel A. Year 1 Outcomes				
MSP × Urban	0.306***	0.239**	33.788***	0.299***
	(0.089)	(0.102)	(8.192)	(0.063)
MSP × Rural Controlled	–0.019	0.557***	54.817**	0.293***
	(0.116)	(0.093)	(22.532)	(0.087)
MSP × Rural Contested	–0.024	0.811***	54.480**	0.281*
	(0.220)	(0.289)	(21.332)	(0.162)
Insecure p-value	0.985	0.403	0.991	0.952
Urban p-value	0.027	0.022	0.381	0.952
Control Mean	0.422	0.405	31.276	0.063
Observations	950	969	922	965
R squared	0.013	0.064	0.080	0.066
Panel B. Year 2 Outcomes				
MSP × Urban	0.236**	–0.265**	–0.275	
	(0.115)	(0.116)	(7.145)	
MSP × Rural Controlled	0.287***	0.097	–1.226	
	(0.110)	(0.135)	(12.086)	
MSP × Rural Contested	0.282	0.316	18.862	
	(0.210)	(0.193)	(17.619)	
Insecure p-value	0.982	0.351	0.348	
Urban p-value	0.749	0.042	0.946	
Control Mean	0.645	0.316	36.628	
Observations	712	735	725	
R squared	0.087	–0.033	0.015	

Notes: This table reports on the impacts of mobile money salary payments on measures of payment quality. Favor MSP is a dummy variable equal to one if the respondent indicates they would support scaling the reform across the Ministry of Education. 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. Money transfer is a dummy variable equal to one if the respondent indicates make a mobile money transfer to someone else in the previous month. Robust standard errors clustered at the registration zone (treatment unit) level. Urban zones are those that have a population density of more than 300 inhabitants per squared kilometer. Controlled zones are those that can be enumerated by outsiders. Contested zones are those that require the permission of a local commander in order to survey. There were no contested urban zones during this study. Endline 1 outcomes were collected during April 2019 and Endline 2 outcomes are collected in May 2020.

Table 3: Impacts on Categories of Mobile Money Use

	Deposit	Transfer to Cust. Wallet	Own Airtime Top-up	Other Airtime Top-up	Balance
	(1)	(2)	(3)	(4)	(5)
Panel A. OLS					
MSP Months	0.004***	0.003**	0.001***	0.002***	0.012***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.002)
Constant	–0.010	–0.013	–0.006***	–0.014***	0.089***
	(0.008)	(0.015)	(0.001)	(0.003)	(0.020)
1-Month Mean	0.002	0.003	0.000	0.000	0.142
	(0.002)	(0.003)	(0.000)	(0.000)	(0.019)
Observations	15,928	15,928	15,928	15,928	15,928
R^2	0.005	0.016	0.022	0.040	0.120
Panel B. TOT					
MSP Months	0.004***	0.004	0.001***	0.002***	0.006**
	(0.001)	(0.003)	(0.000)	(0.000)	(0.003)
Observations	15,928	15,928	15,928	15,928	15,928
R^2	0.003	0.001	0.011	0.016	0.006

Notes: The following outcomes are measured as the total value of transactions conducted in October 2018 – 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. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors clustered at the registration zone (treatment unit) level are reported in parentheses.

Table 4: Impacts on Categories of Mobile Money Use by Security

	Deposit (1)	Transfer to Cust. Wallet (2)	Own Airtime Top-up (3)	Other Airtime Top-up (4)	Balance (5)
Panel A. OLS					
MSP Months × Urban	0.004***	0.003*	0.001***	0.002***	0.008***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)
MSP Months × Rural Controlled	0.002*	–0.000	–0.000	–0.000	0.030***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.007)
MSP Months × Rural Contested	0.003**	0.004*	0.001***	0.001***	0.021***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.005)
1-Month Mean	0.002	0.003	0.000	0.000	0.142
Unsafe p-value	0.344	0.079	0.000	0.000	0.259
Observations	15,928	15,928	15,928	15,928	15,928
R squared	0.005	0.017	0.023	0.042	0.122
Panel B. TOT					
MSP Months × Urban	0.004***	0.004	0.001***	0.002***	0.005*
	(0.001)	(0.003)	(0.000)	(0.001)	(0.003)
MSP Months × Rural Controlled	–0.009	0.001	0.000	0.000	0.030
	(0.012)	(0.001)	(0.000)	(0.000)	(0.030)
MSP Months × Rural Contested	0.008***	0.001	0.001	0.001	0.006
	(0.002)	(0.004)	(0.001)	(0.001)	(0.008)
Unsafe p-value	0.164	0.854	0.493	0.254	0.433
Observations	15,928	15,928	15,928	15,928	15,928
R squared	0.001	0.001	0.011	0.017	0.008

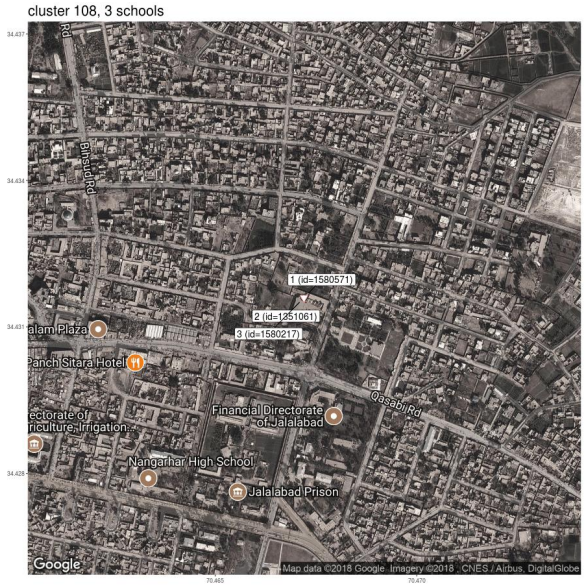
Notes: Urban zones are those that have a population density of more than 300 inhabitants per squared kilometer. Controlled zones are those that can be enumerated by outsiders. Contested zones are those that require the permission of a local commander in order to survey. There were no contested urban zones during this study. The following outcomes are measured as the total value of transactions conducted in October 2018 – 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. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors clustered at the registration zone (treatment unit) level are reported in parentheses.

Appendix

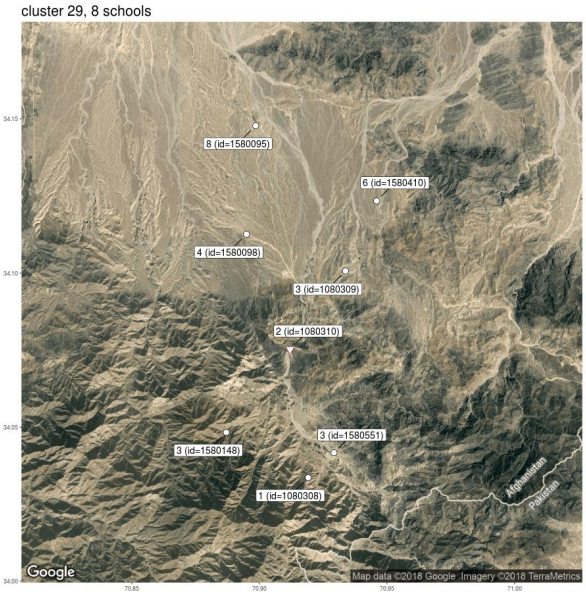
A Supplementary Figures and Tables

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

Panel A: Urban Registration Zone

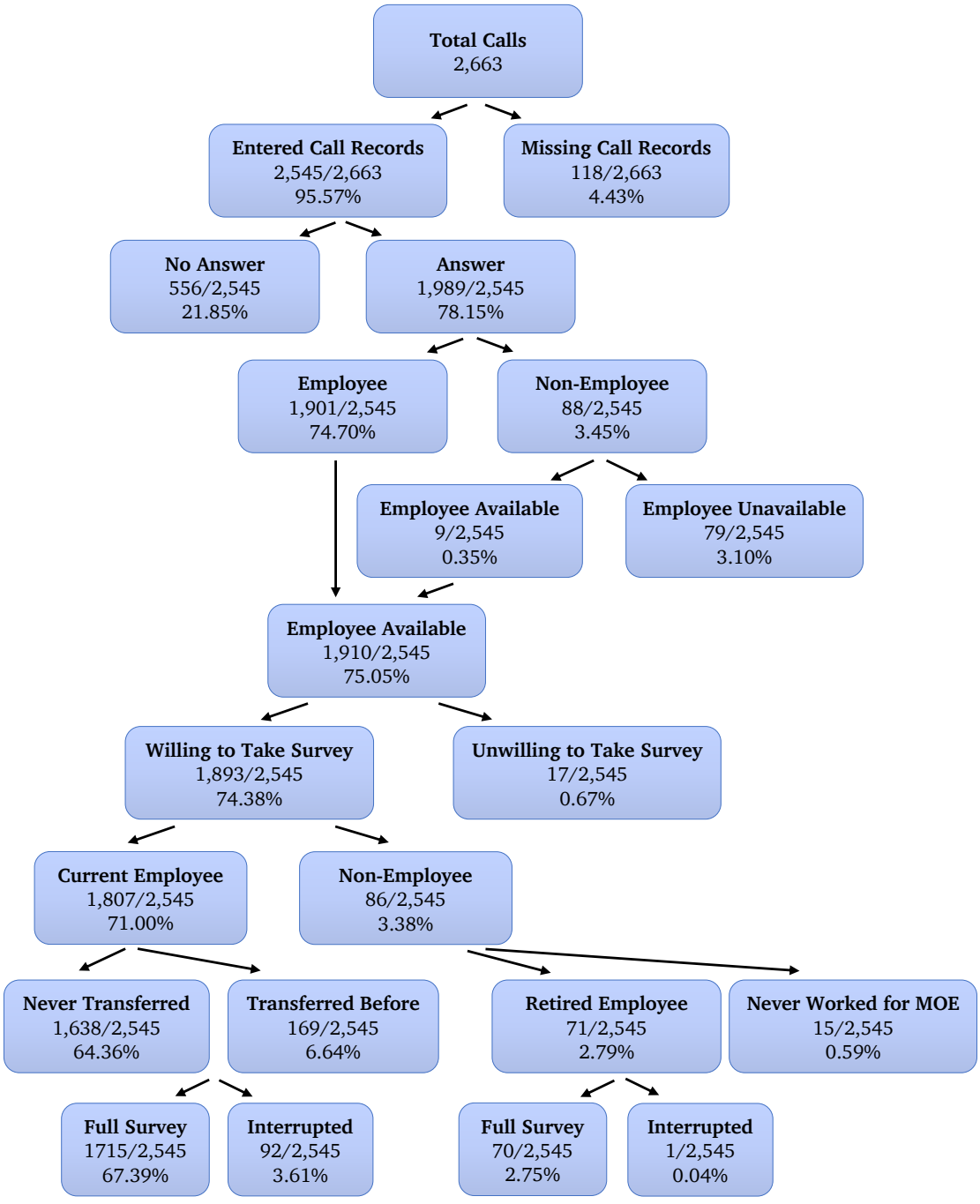


Panel B: Rural Registration Zone



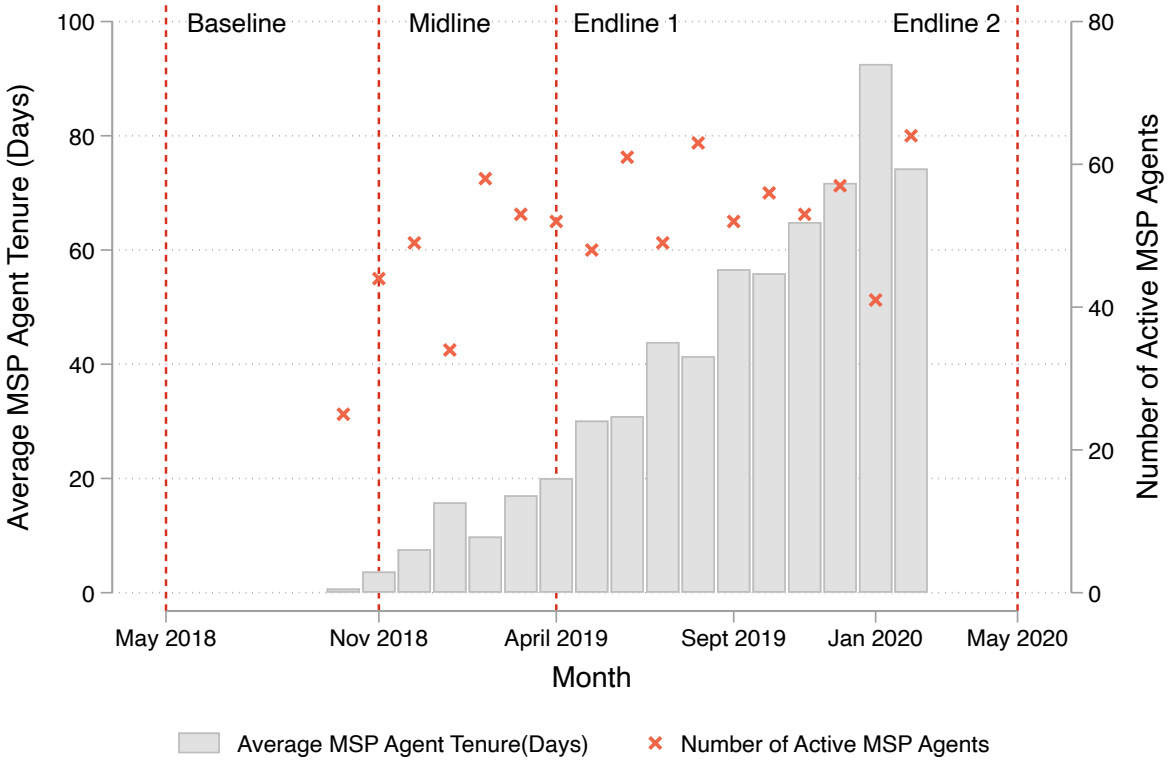
Notes: This figure shows examples of two registration zones located in urban and rural areas of Nangarhar Province. White windows display the location and identifiers of the schools located in these zones.

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: Treatment Effects on Education Quality at Endline 1 (1st Year)

	Employees' Attendance		**Students' Leaning Assessment**		
	Midline	Endline	Total Score	Math Score	Reading Score
	(1)	(2)	(3)	(4)	(5)
Panel A. ITT					
Early	–0.018	–0.036	0.049	0.049	0.044
	(0.029)	(0.029)	(0.082)	(0.084)	(0.082)
Late	–0.034	–0.018	0.062	0.091	0.039
	(0.030)	(0.028)	(0.082)	(0.085)	(0.080)
Control	0.518***	0.457***	0.000	0.000	–0.000
	(0.028)	(0.027)	(0.061)	(0.062)	(0.061)
Early + Late = Control	0.315	0.281	0.432	0.334	0.552
Observations	10,906	10,731	1,101	1,101	1,101
R-squared	0.077	0.066	0.094	0.098	0.086
Panel B. TOT					
MSP	–0.023	–0.041	0.068	0.073	0.059
	(0.033)	(0.034)	(0.106)	(0.109)	(0.106)
Observations	10,906	10,731	1,101	1,101	1,101
R-squared	0.075	0.064	0.092	0.096	0.085

Notes: This table reports treatment effects for the outcomes measuring the quality of education during the first year of the reform, controlling for the district (strata) fixed effects. Panel A presents the results for attendance rates of MoE employees at the Midline (November 2018) and Endline 1 (April 2019). The sample consists of MoE employees working in the visited schools who received positive salary payments at baseline according to the administrative payroll data (M41 forms). The employee is present if the survey team found him at the school during the unannounced visit. Endline ToT is the treatment-on-the-treated effect obtained by instrumenting the payment via MSP with the treatment group assignment. Panel B presents the results for students' learning outcomes based on the independent learning assessment. The sample consists of 1,001 children aged 6-10 who live in the catchment area of 367 experimental schools. The test scores are standardized relative to the outcomes in the control group. Test score ToTs are the treatment-on-the-treated effects obtained by instrumenting the share of employees paid via MSP with the treatment group assignment. LL is the Late Registration, Late Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and EL is the Early Registration, Late Payment (Treatment 2) group. Robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.

Table A.2: Registration Outcomes

Months Paid	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 registered for a mobile money wallet during the main registration wave (before July 18, 2018). Alternatively, the employee is "Registered With Delay" if he registered after the main registration wave (after July 18, 2018).

Table A.3: Balance of Nonresponse Rates for Main Outcomes

Outcome	Sample Mean (1)	LL Mean (2)	EE Mean (3)	EL Mean (4)	P-value LL=EE=EL (5)	N Obs. (6)
Panel A: Baseline – May 2018						
Share Voting in Favor of MSP	0.02	0.02	0.02	0.01	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	0.50	1005
	(0.08)	[0.00]	[0.01]	[0.00]		
Share Experienced Salary Delays	0.00	0.00	0.00	0.01	0.24	1005
	(0.06)	[0.00]	[0.00]	[0.01]		
Payment to Receive Salary (Afg)	0.01	0.02	0.01	0.01	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)	[0.03]	[0.02]	[0.02]		
Panel B: Endline 1 – April 2019						
Share Voting 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]		
Share Experienced Salary Delays	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						
Share Voting 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]		
Share Experienced Salary Delays	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. LL is the Late Registration, Late Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and EL is the Early Registration, Late Payment (Treatment 2) group. Standard errors in parentheses and robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.

Table A.4: First Stage Estimates: Self-Reported MSP Payment

Outcome	LL Mean (1)	EE Effect (2)	EL Effect (3)	EE - EL Difference (4)	N Obs. (5)
Panel A: Endline 1 – April 2019					
Share Self-Reported MSP	0.08	0.63***	0.37***	0.27***	970
	[0.02]	[0.04]	[0.04]	[0.05]	
Panel B: Endline 2 – May 2020					
Share Self-Reported MSP	0.33	0.49***	0.15***	0.32***	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 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. LL is the Late Registration, Late Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and EL is the Early Registration, Late Payment (Treatment 2) group. Robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.

Table A.5: Effects on Salary Payment Experience – Without Stratum FEs

	LL Mean (1)	EE Effect (2)	EL Effect (3)	ToT Effect (4)	N Obs (5)
Panel A. Baseline					
Share Voting in Favor of MSP	0.936	–0.027	–0.026		985
	(0.019)	(0.029)	(0.028)		
Share Very Satisfied with Pay System	0.297	–0.020	–0.041		999
	(0.034)	(0.046)	(0.047)		
Share Experienced Salary Delays	0.543	0.083	–0.017		1,001
	(0.039)	(0.053)	(0.056)		
Payment to Receive Salary (Afg)	23.533	0.806	–3.109		990
	(4.167)	(5.371)	(5.160)		
Travel Time to Cash-Out (Min)	33.412	–0.541	–0.983		928
	(3.192)	(4.370)	(4.297)		
Travel Time to Cash-Out (Afg)	58.723	3.015	3.131		935
	(7.834)	(10.727)	(11.060)		
Panel B. Endline 1					
Share Voting in Favor of MSP	0.422	0.081	0.052	0.128	950
	(0.040)	(0.057)	(0.057)	(0.088)	
Share Very Satisfied with Pay System	0.628	–0.257***	–0.146***	–0.401***	966
	(0.037)	(0.053)	(0.055)	(0.085)	
Share Experienced Salary Delays	0.405	0.261***	0.118**	0.404***	969
	(0.037)	(0.052)	(0.057)	(0.080)	
Payment to Receive Salary (Afg)	22.311	–3.735	–0.452	–5.589	959
	(2.626)	(4.110)	(3.986)	(6.240)	
Travel Time to Cash-Out (Min)	31.276	29.355***	20.677***	46.084***	922
	(3.241)	(6.298)	(7.224)	(9.423)	
Travel Time to Cash-Out (Afg)	45.724	82.644***	56.556***	129.270***	932
	(6.776)	(16.315)	(15.226)	(24.966)	
Panel C. Endline 2					
Share Voting in Favor of MSP	0.645	0.104**	0.011	0.233***	712
	(0.035)	(0.047)	(0.051)	(0.090)	
Share Very Satisfied with Pay System	0.727	–0.022	–0.022	–0.041	736
	(0.031)	(0.045)	(0.043)	(0.092)	
Share Experienced Salary Delays	0.316	–0.021	–0.015	–0.039	735
	(0.030)	(0.044)	(0.043)	(0.089)	
Payment to Receive Salary (Afg)	12.479	–8.062***	0.986	–18.396***	738
	(1.853)	(2.266)	(3.281)	(3.874)	
Travel Time to Cash-Out (Min)	36.628	3.469	0.163	7.535	725
	(2.185)	(3.264)	(2.976)	(6.632)	
Travel Time to Cash-Out (Afg)	53.323	34.496***	15.404*	68.318***	635
	(5.485)	(10.957)	(8.443)	(24.244)	

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. LL is the Late Registration, Late Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and EL is the Early Registration, Late Payment (Treatment 2) group. 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.

Table A.6: Effects on Salary Payment Experience – With Stratum FEs

	LL Mean (1)	EE Effect (2)	EL Effect (3)	ToT Effect (4)	N Obs (5)
Panel A. Baseline					
Share Voting in Favor of MSP	0.936	–0.027	–0.024		984
	(0.019)	(0.027)	(0.024)		
Share Very Satisfied with Pay System	0.297	–0.022	–0.041		998
	(0.034)	(0.042)	(0.044)		
Share Experienced Salary Delays	0.543	0.085**	0.005		1,000
	(0.039)	(0.042)	(0.043)		
Payment to Receive Salary (Afg)	23.533	1.089	–3.956		989
	(4.167)	(3.735)	(4.216)		
Travel Time to Cash-Out (Min)	33.412	–2.378	–1.838		927
	(3.192)	(3.647)	(3.348)		
Travel Time to Cash-Out (Afg)	58.723	–1.433	0.076		934
	(7.834)	(6.045)	(5.976)		
Panel B. Endline 1					
Share Voting in Favor of MSP	0.422	0.110**	0.076*	0.178***	950
	(0.040)	(0.044)	(0.045)	(0.068)	
Share Very Satisfied with Pay System	0.628	–0.252***	–0.145***	–0.397***	966
	(0.037)	(0.049)	(0.048)	(0.078)	
Share Experienced Salary Delays	0.405	0.254***	0.146***	0.401***	969
	(0.037)	(0.048)	(0.049)	(0.075)	
Payment to Receive Salary (Afg)	22.311	–3.208	–1.644	–4.966	959
	(2.626)	(3.402)	(2.996)	(5.109)	
Travel Time to Cash-Out (Min)	31.276	27.193***	21.824***	43.208***	922
	(3.241)	(5.880)	(6.052)	(8.684)	
Travel Time to Cash-Out (Afg)	45.724	80.058***	58.801***	125.785***	932
	(6.776)	(15.592)	(13.549)	(23.500)	
Panel C. Endline 2					
Share Voting in Favor of MSP	0.645	0.122***	0.013	0.263***	712
	(0.035)	(0.043)	(0.040)	(0.082)	
Share Very Satisfied with Pay System	0.727	–0.003	–0.015	–0.001	736
	(0.031)	(0.041)	(0.040)	(0.081)	
Share Experienced Salary Delays	0.316	–0.033	–0.018	–0.064	735
	(0.030)	(0.042)	(0.040)	(0.083)	
Payment to Receive Salary (Afg)	12.479	–7.787***	1.252	–17.563***	738
	(1.853)	(2.033)	(2.549)	(3.955)	
Travel Time to Cash-Out (Min)	36.628	1.661	–0.244	3.677	725
	(2.185)	(3.185)	(2.999)	(6.110)	
Travel Time to Cash-Out (Afg)	53.323	28.542***	13.395	53.554***	635
	(5.485)	(9.217)	(8.200)	(17.552)	

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. LL is the Late Registration, Late Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and EL is the Early Registration, Late Payment (Treatment 2) group. 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.

Table A.7: Spatial Treatment Externalities on Salary Payment Experience (within 2 km)

Outcome	LL Mean (1)	EE Effect (2)	EL Effect (3)	# EE Neighb Schools 2 km (4)	# EL Neighb Schools 2 km (5)	# Total Neighb Schools 2 km (6)	N Obs. (7)
Panel A: Endline 1 – April 2019							
Share Voting in Favor of MSP	0.42	0.11**	0.09**	0.00	-0.02	0.00	950
	[0.04]	[0.04]	[0.04]	[0.01]	[0.02]	[0.01]	
Share Very Satisfied with Pay System	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]	
Share Experienced Salary Delays	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]	
Payment to Receive Salary (Afg)	22.31	-4.05	-1.04	0.64	-0.73	-0.09	959
	[2.63]	[3.02]	[3.00]	[1.20]	[0.95]	[0.55]	
Travel Time to Cash-Out (Min)	31.28	28.10***	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							
Share Voting in Favor of MSP	0.65	0.11***	0.01	0.01	0.00	-0.00	712
	[0.04]	[0.04]	[0.04]	[0.01]	[0.02]	[0.01]	
Share Very Satisfied with Pay System	0.73	-0.01	-0.02	0.01	0.01	-0.00	736
	[0.03]	[0.04]	[0.04]	[0.01]	[0.01]	[0.01]	
Share Experienced Salary Delays	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]	
Payment to Receive Salary (Afg)	12.48	-8.72***	1.54	0.83	-0.49	-0.11	738
	[1.85]	[2.25]	[2.59]	[0.85]	[0.90]	[0.39]	
Travel Time to Cash-Out (Min)	36.63	3.35	-0.24	-1.35	0.42	-0.03	725
	[2.19]	[3.27]	[3.05]	[1.10]	[1.00]	[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. LL is the Late Registration, Late Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and EL is the Early Registration, Late Payment (Treatment 2) group. # EE (EL) Neighb Schools 2 km is the number of neighboring schools within a two-kilometer radius located in EE (EL) treatment zone, # Total Neighb 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.

Table A.8: First Stage Estimates with Strata Fixed Effects

Outcome	LL Mean (1)	EE Effect (2)	EL Effect (3)	EE - EL Difference (4)	F-stat (5)	N Obs. (6)
Panel A: Endline 1 – April 2019						
Share Registered	0.66	0.26***	0.25***	0.01	24.74	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	0.65	0.29***	0.25***	0.02	25.13	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 emoloyees 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. LL is the Late Registration, Late Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and EL is the Early Registration, Late Payment (Treatment 2) group. Robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.

Table A.9: Registration Effects on the Main Outcomes (LL Control Group, Unpaid)

Outcome	Unregistered Mean (1)	Registration Effect (2)	N Obs. (3)
Panel A: Endline 1 – April 2019			
Share Voting in Favor of MSP	0.35	0.13	296
	[0.06]	[0.09]	
Share Satisfied with Payment System	0.90	0.02	305
	[0.03]	[0.08]	
Share Experienced Salary Delays	0.36	0.05	307
	[0.06]	[0.08]	
Payment to Receive Salary (Afg)	32.41	-1.30	305
	[4.69]	[9.29]	
Travel Time to Cash-Out (Min)	31.69	-1.44	290
	[4.26]	[6.46]	
Travel Cost to Cash-Out (Afg)	36.68	25.28	289
	[9.50]	[24.66]	
Panel B: Endline 2 – May 2020			
Share Voting in Favor of MSP	0.46	0.01	158
	[0.04]	[0.14]	
Share Satisfied with Payment System	1.00	-0.01	172
	[0.00]	[0.01]	
Share Experienced Salary Delays	0.46	-0.07	172
	[0.06]	[0.14]	
Payment to Receive Salary (Afg)	21.95	-3.06	172
	[3.30]	[6.32]	
Travel Time to Cash-Out (Min)	37.58	1.42	171
	[3.89]	[7.99]	
Travel Cost to Cash-Out (Afg)	68.80	-21.31	147
	[10.69]	[29.48]	

Table A.10: Heterogeneity Treatment Effects – Sec. Measure: LWJ

	Favor MSP (1)	Experienced Delay (2)	Travel Time (3)	Money Transfer (4)
Panel A. Year 1 Outcomes				
MSP × Urban	0.306***	0.239**	33.788***	0.299***
	(0.089)	(0.102)	(8.192)	(0.063)
MSP × Rural × Secure	0.035	0.584***	56.993***	0.365***
	(0.131)	(0.099)	(21.733)	(0.088)
MSP × Rural × Insecure	–0.085	0.722***	53.187*	0.164
	(0.172)	(0.227)	(27.762)	(0.135)
Insecure p-value	0.579	0.579	0.914	0.213
Urban p-value	0.088	0.015	0.318	0.541
Control Mean	0.422	0.405	31.276	0.063
Observations	950	969	922	965
R squared	0.010	0.062	0.079	0.066
Panel B. Year 2 Outcomes				
MSP × Urban	0.236**	–0.265**	–0.275	
	(0.115)	(0.116)	(7.145)	
MSP × Rural × Secure	0.303**	–0.002	8.982	
	(0.123)	(0.133)	(13.479)	
MSP × Rural × Insecure	0.288	0.390**	7.366	
	(0.196)	(0.193)	(16.331)	
Insecure p-value	0.949	0.096	0.939	
Urban p-value	0.691	0.136	0.544	
Control Mean	0.645	0.316	36.628	
Observations	712	735	725	
R squared	0.087	–0.049	0.005	

Notes: Robust standard errors clustered at the registration zone (treatment unit) level.

Table A.11: Heterogeneity Analysis Mobile Money Use, Standardised by Salary – By Safety & Rural Type, LWJ Measure

	Deposit (1)	Transfer to Cust. Wallet (2)	Own Airtime Top-up (3)	Other Airtime Top-up (4)	Balance (5)
Panel A. ITT					
MSP Months × Urban	0.004***	0.003*	0.001***	0.002***	0.008***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)
MSP Months × Rural × Insecure	0.000	0.001	0.000**	0.000**	0.022***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.007)
MSP Months × Rural × Secure	0.004***	0.004	0.001***	0.001***	0.023***
	(0.001)	(0.003)	(0.000)	(0.000)	(0.004)
Constant	–0.005	–0.010	–0.005***	–0.012***	0.066***
	(0.007)	(0.013)	(0.001)	(0.002)	(0.024)
1-Month Mean	0.002	0.003	0.000	0.000	0.142
	(0.002)	(0.003)	(0.000)	(0.000)	(0.019)
Insecure p-value	0.032	0.187	0.112	0.006	0.946
Observations	15,928	15,928	15,928	15,928	15,928
R squared	0.005	0.017	0.022	0.042	0.122
Panel B. TOT					
MSP Months × Urban	0.004***	0.004	0.001***	0.002***	0.005*
	(0.001)	(0.003)	(0.000)	(0.001)	(0.003)
MSP Months × Rural × Insecure	0.003	–0.008	–0.001	–0.001	–0.025
	(0.006)	(0.008)	(0.001)	(0.001)	(0.037)
MSP Months × Rural × Secure	0.007***	0.005**	0.001**	0.002**	0.010
	(0.002)	(0.002)	(0.000)	(0.001)	(0.008)
Insecure p-value	0.514	0.130	0.017	0.021	0.365
Urban p-value	0.233	0.969	0.470	0.519	0.596
Observations	15,928	15,928	15,928	15,928	15,928
R squared	0.002	0.000	0.004	0.016	–0.004

Robust standard errors clustered at the registration zone (treatment unit) level. The following outcomes are measured as the total value of transactions conducted in October 2018 – 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 at the time of the last MoE payment for each employee. All variables in AFN.

Table A.12: Treatment Effects on Government Support at the Endline 1 (April 2019)

Outcome	LL Mean (1)	EE Effect (2)	EL Effect (3)	ToT Effect (4)	N Obs. (5)
Government Support Index (Equal-Weighted)	0.00	0.02	0.00	0.03	929
	[0.03]	[0.05]	[0.05]	[0.07]	
Favorable of Government (=1)	0.75	0.06*	0.03	0.09*	929
	[0.03]	[0.04]	[0.04]	[0.05]	
Government Improves Living Conditions (=1)	0.82	-0.01	-0.02	-0.02	929
	[0.03]	[0.03]	[0.04]	[0.05]	
Will Vote in 2019 Election (=1)	0.94	-0.00	-0.01	-0.00	929
	[0.01]	[0.02]	[0.02]	[0.03]	
Very Important to Report to ANDSF (=1)	0.92	-0.02	-0.00	-0.03	929
	[0.02]	[0.02]	[0.02]	[0.04]	
Government Agency Dispute Resolution (=1)	0.13	0.02	0.02	0.03	929
	[0.02]	[0.03]	[0.03]	[0.05]	

Notes: This table reports treatment effects for the outcomes measuring government support at the endline 1 (April 2019), controlling for the district (strata) fixed effects. The sample consists of 929 MoE employees who participated in the full endline 1 survey, self-reported the payment system, and successfully responded to all of the above outcomes. LL is the Late Registration, Late Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and EL is the Early Registration, Late Payment (Treatment 2) group. ToT is the treatment-on-the-treated effect obtained by instrumenting the self-reported payment via MSP with the treatment group assignment. Robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.

Table A.13: Effects on Financial Inclusion at Endline 1 (April 2019)

Outcome	LL Mean (1)	EE Effect (2)	EL Effect (3)	ToT Effect (4)	N Obs. (5)
P2P Transactions Index	0.00	0.39 [0.005]	0.35 [0.006]	0.63 [0.002]	965
Money Transfer via Mobile Phone (=1)	0.06	0.18 (0.001)	0.17 (0.001)	0.29 (0.001)	965
Money Transfer via Other Means (=1)	0.24	0.02 (0.742)	0.01 (0.756)	0.03 (0.471)	965
Savings Index	0.00	0.08 [0.673]	0.07 [0.409]	0.13 [0.595]	960
Saves Typically (=1)	0.36	0.02 (0.688)	0.07 (0.306)	0.05 (0.739)	960
Saves: Mobile Money Account (=1)	0.00	0.02 (0.244)	0.02 (0.244)	0.04 (0.065)	960
Saves: Other than Mobile Money (=1)	0.18	0.00 (0.871)	0.01 (0.719)	0.00 (1.000)	960
Savings: Mobile Money Account (AFG)	0	106 (0.414)	47 (0.306)	164 (0.504)	960
Savings: Other than Mobile Money (AFG)	3547	864 (0.799)	-1109 (0.688)	1030 (1.000)	960
Financial Empowerment Index	0.00	-0.09 [0.794]	-0.13 [0.440]	-0.16 [0.712]	952
Empowered in Small Purchases (=1)	0.92	-0.03 (0.567)	-0.05 (0.330)	-0.05 (0.566)	952
Empowered in Expensive Purchases (=1)	0.97	-0.01 (0.567)	-0.01 (0.567)	-0.02 (0.566)	952
Consumption Index	0.00	-0.01 [0.874]	0.02 [0.956]	-0.01 [0.893]	959
No Hunger	0.88	0.04 (0.524)	0.06 (0.201)	0.07 (0.648)	959
No Unpaid School Fees	0.82	-0.01 (0.912)	-0.01 (0.912)	-0.02 (0.822)	959
No Default	0.64	0.00 (0.912)	0.02 (0.912)	0.00 (0.957)	959
No Unpaid Medical Emergency	0.74	-0.05 (0.647)	-0.05 (0.524)	-0.08 (0.648)	959
Understands Salary Calculation (=1)	0.80	-0.03 [0.827]	0.01 [0.943]	-0.04 [0.875]	963
Borrowing Index	0.00	0.05 [0.827]	0.02 [0.956]	0.07 [0.875]	948
Outstanding Loan (=1)	0.58	0.07 (1.000)	0.03 (1.000)	0.11 (1.000)	948
Overdue Loan (=1)	0.52	0.04 (1.000)	0.00 (1.000)	0.05 (1.000)	948
Overdue Loan Amount (AFG)	25662	-1033 (1.000)	-128 (1.000)	-1517 (1.000)	948
Principal Loan Amount (AFG)	30773	-1971 (1.000)	1636 (1.000)	-2478 (1.000)	948

Notes: This table reports treatment effects for the financial inclusion outcomes at the endline 1 (April 2019), controlling for the district (strata) fixed effects. The sample consists of 970 MoE employees who participated in the full survey and self-reported the payment system in the Endline 1 survey. LL is the Late Registration, Late Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and EL is the Early Registration, Late Payment (Treatment 2) group. ToT is the treatment-on-the-treated effect obtained by instrumenting the self-reported payment via MSP with the treatment group assignment. Romano-Wolf p-values adjusting for FWER between indexes in square brackets. Sharpened q-values adjusting for FDR within indexes in brackets. Standard errors are clustered at the registration zone (treatment unit) level.

Table A.14: Effects on Mobile Money Wallet Use

Outcome	1 Month Mean (1)	# Months OLS (2)	# Months 2SLS (3)	N Obs. (4)
Panel A: Year 1, October 2018 – May 2019				
Deposit (=1)	0.01	0.00***	0.00	13402
	[0.01]	[0.00]	[0.00]	
Transfer to Customer Wallet (=1)	0.01	0.00	0.00	13402
	[0.01]	[0.00]	[0.00]	
Airtime Top-up to Own Number (=1)	0.01	0.02***	0.02**	13402
	[0.01]	[0.00]	[0.01]	
Airtime Top-up to Another Number (=1)	0.02	0.02***	0.02*	13402
	[0.01]	[0.00]	[0.01]	
Pre-Pay Balance > 0 (=1)	0.43	0.03***	0.04***	13402
	[0.05]	[0.01]	[0.01]	
Pre-Pay Balance > 100 (=1)	0.24	-0.02***	-0.00	13402
	[0.05]	[0.00]	[0.01]	
Pre-Pay Balance > 1000 (=1)	0.24	-0.02***	-0.01	13402
	[0.05]	[0.00]	[0.01]	
Pre-Pay Balance $1,000 - 5,000$ (=1)	0.00	-0.00***	-0.01**	13402
	[0.00]	[0.00]	[0.00]	
Panel B: Year 2, June 2019 – February 2020				
Deposit (=1)	0.00	0.00***	0.00***	15384
	[0.00]	[0.00]	[0.00]	
Transfer to Customer Wallet (=1)	0.00	0.00*	0.00	15384
	[0.00]	[0.00]	[0.00]	
Airtime Top-up to Own Number (=1)	0.02	0.01***	0.01***	15384
	[0.01]	[0.00]	[0.00]	
Airtime Top-up to Another Number (=1)	0.03	0.01***	0.01***	15384
	[0.01]	[0.00]	[0.00]	
Pre-Pay Balance > 0 (=1)	0.60	0.01***	-0.00	15384
	[0.08]	[0.00]	[0.00]	
Pre-Pay Balance > 100 (=1)	0.13	0.01***	0.00	15384
	[0.02]	[0.00]	[0.00]	
Pre-Pay Balance > 1000 (=1)	0.13	0.01***	0.00	15384
	[0.02]	[0.00]	[0.00]	
Pre-Pay Balance $1,000 - 5,000$ (=1)	0.00	0.00	0.00	15384
	[0.00]	[0.00]	[0.00]	

Notes: This table reports the effects of an additional month of MSP payments on mobile money wallet use, controlling for the district (strata) fixed effects. The sample consists of 15,951 employees who received at least one payment via MSP. 2SLS instruments the number of months paid via MSP with the treatment group assignment. The following outcomes are a dummy equal to one if an employee conducted at least one transaction during the first or second year: 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 at the time of the last MoE payment each year for every employee. Robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.

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:[21]

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.

[21] 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.

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:

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

where Y_{iszd} is the payment experience outcome for employee i from school s located in registration zone z and district d, $\# \, EE \; Neighbors_{szd}^{2km}$ and $\# \, EL \; Neighbors_{szd}^{2km}$ are the number of neighboring schools within a two-kilometer radius from school s located in EE and EL treatment zones, $\# \, Total \; Neighbors_{szd}^{2km}$ 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.7 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.6. 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.