Strengthening Fragile States: Evidence from Mobile Salary Payments in Afghanistan

Joshua Blumenstock Michael Callen Anastasiia Faikina
Stefano Fiorin Tarek Ghani*

September 10, 2024

Abstract

We conduct a randomized evaluation of a flagship Afghan government initiative designed to build state capacity. The program sought to modernize the tracking and payment of Ministry of Education employees through biometric registration and digital payments – technologies particularly promising in a context lacking a national ID system and financial infrastructure. The reform reduced payment delays, improved student learning, and deepened financial inclusion. Effects were more immediate and concentrated in urban areas (where the government had uncontested control). Such investments in basic state capacity might require that citizens are confident that the state will continue to exist, which, by definition, is not assured in fragile states. Consistent with this, implementation was most successful in districts with higher public confidence in, and consensus about, the government's ability to prevail against the Taliban. Collectively, these results indicate that technology can increase state capacity in fragile states and highlight the critical role of public confidence in state continuity to build such capacity.

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

Keywords: state capacity, fragile states, beliefs

JEL Classification: P00, O10, D70

1 Introduction

Modern states effectively organize and compensate their employees. Improving these capabilities is essential for fragile states to transition from collections of overlapping informal, personalized arrangements into structured, enduring organizations (Besley and Persson, 2009; Dincecco, 2015; Cantoni, Mohr and Weigand, Forthcoming). Two defining features of fragility, however, create obstacles. First, fragile states lack a monopoly on violence (Tilly, 1985; Olson, 1993; Besley and Persson, 2011*a*; Sanchez de la Sierra, 2020), leading to contested territories where the government has difficulty safely monitoring employees and paying their salaries (Dell, Lane and Querubin, 2018; Allen, Bertazzini and Heldring, 2023). Second, the risk of outright collapse, violent elite displacement, or other forms of radical institutional change (Fearon and Laitin, 2003; Collier, 2008; Acemoglu and Robinson, 2012) might deter necessary investments in state capacity – including basic functions like salary payments – that are expensive, coordination-intensive, and provide benefits over a long horizon (North and Thomas, 1973). Territorial contestation, low and stagnant capacity, high deprivation, and the risk of radical change are all hallmarks of states stuck in a 'fragility trap' (Besley, Collier and Khan, 2018). Addressing these issues requires states to build administrative and fiscal capacity, starting with systems to organize and compensate government employees.

Digital identification and payment systems offer promising solutions for building state capacity, especially in poor and conflict-affected countries where the absence of national ID systems and formal financial infrastructure makes it difficult for states to track and pay their employees.[1] Tracking and paying employees is a prerequisite to running a modern state. Yet in fragile contexts, records of government employees are often unreliable, and payments are highly irregular. In our baseline audit, only 41% of employees were present, and among them, only 30% had a government ID card. Additionally, 60% of employees experienced salary delays averaging 3.5 months (s.d. = 1.2 months). Digital technologies can make it easier and less risky to verify employees and to track and disburse salary payments relative to traditional analog systems, including paper-based employee rosters and the physical transportation of cash, with possible downstream impacts on public service delivery. Biometrics remove the need for a national ID system and can help eliminate ghost workers by creating a self-updating employee registry, while digital payments can reduce delays, minimize leakages,

[1] A substantial and growing literature explores the use of biometrics and digital payments to enhance state capacity in more stable developing countries, including functions such as providing social protection (Muralidharan, Niehaus and Sukhtankar, 2016, Forthcoming; Giné, Goldberg and Yang, 2012), tracking payments (Banerjee et al., 2020), and increasing state accountability (Duflo, Hanna and Ryan, 2012; Callen and Long, 2015; Callen et al., 2016; Dhaliwal and Hanna, 2017; Bossuroy, Delavallade and Pons, 2019).

and expand access to financial services in areas without banks. Reflecting growing interest, the World Bank has recently supported digital ID systems in nearly 60 countries and digital payments in 35, including in 10 fragile and conflict-affected states (World Bank, 2024). But the feasibility and impacts of digital technologies in fragile states are still not well understood.

This paper reports a randomized evaluation of a flagship Afghan government initiative – called the Mobile Salary Payment (MSP) reform – that used digital technologies to improve the identification and payment of civil servants. Initiated by the National Unity Government in 2016 and implemented between 2017-2021 by the Afghan Ministry of Education, the reform was seen as an investment in basic state capacity that could yield long-term benefits. In the context of this reform, we conducted an at-scale field experiment to address two broad questions. First, can biometric identification and mobile payments help fragile governments increase capacity and improve the organization of key services (in our case, education), including in violently contested territories? Second, do beliefs regarding the state's durability – and uncertainty about the same – impact the development of state capacity?

The Afghan government initiative we examine is an ideal context to study these questions. First, our field experiment involved approximately 34,000 public employees (mainly teachers) working in roughly 1,500 schools across three conflict-affected provinces that account for about 15% of Afghanistan's population. The study region thus includes both areas where the government had relatively uncontested control (Kandahar City, Jalalabad, and Kabul), and also substantial areas (about 1/3rd of our rural districts) contested variously by the Taliban, the Islamic State, and other violent actors. Second, the reform required substantial investment and coordination between a large number of employees, as well as the Ministry of Education (hereon, MoE), Ministry of Finance (MoF) and the Ministry of Communications and Information Technology (MCIT) at the federal, provincial, and local levels. Third, during our study, the US government was collecting (independently from our study) measures of citizens' beliefs about who would win the war.[2] The combination of a large-scale state-building reform and beliefs data provide a unique opportunity to study how beliefs about state durability shape the impact of complex reforms and whether the risk of radical institutional change deters the development of state capacity.

Our experimental design randomized the timing of the biometric registration process and the implementation of mobile money payments across three treatment arms: early registration and early mobile money payments (EE), early registration and delayed mobile money payments (ED), and delayed registration and delayed mobile money payment (DD or con-

[2]Despite playing a central role in models related to fragility and democratic consolidation, data reflecting beliefs about future institutions are rarely collected (cf. Callen, Weigel and Yuchtman, Forthcoming).

trol). The design allows us to study biometric registration *per se* using the ED arm as well as the full package of registration and payments using the EE arm. We combine administrative data from the government's payroll records and the Mobile Network Operator's transaction records with original surveys of MoE employees and students to produce three basic results.

First, biometric registration, alone, was not sufficient to perfectly identify public employees and remove "ghost workers" – a major concern for international donors paying government salaries – from the payroll. Superficially, we find that 2.8% of employees in the two early registration arms never register biometrically, despite official warnings that failing to do so would prevent them from being paid. Further investigation reveals this is an underestimate of ghost workers due to non-working "stand-ins" providing biometric information in order to receive salary payments. We construct a forensic "litmus test" survey that quizzes employees on basic facts that an active public employee should know (e.g., the name of the school's principal). By comparing the quiz performance of certified employees (who were verified to be present during a separate round of unannounced audit visits) with the performance of other biometrically registered employees, we estimate that the true share of employees who are ghost workers is likely between 8.5% and 20.4% of paid employees.[3]

Second, mobile payments improved the salary experience of employees, but the effects were neither immediate nor evenly distributed. Over the course of the reform, pervasive payment delays were reduced, from affecting 59% to 35% of all employees. We provide experimental evidence that this was driven by reductions in "first mile" delays (i.e., those occurring during the central salary approval process in Kabul): in treatment zones, first mile delays dropped from 45.4% to 14.6% after one year; in control zones, they dropped from 45.6% to 32.0%.[4] Treatment, therefore, reduced the share of employees whose salaries were held up due to first mile delays by 17.2pp in the first year.

Unexpectedly, also during the first year, treatment induced a large *increase* in "last mile" delays (i.e., those associated with local bank branches and mobile money agents) because the mobile network operator did not have enough agents to cash out salaries. After two years, however, delays were no longer distinguishable between the two groups and were substantially reduced relative to baseline. In urban areas (where the government had uncontested control and the MNO had more agents), treated employees reported better payment experiences and

[3]We separately elicit beliefs about the share of ghost workers in qualitative interviews with 76 education stakeholders, including NGO workers, journalists, local elites, senior and local government officials, school staff, and teachers. These respondents, on average, believe that 12% of teachers are ghosts, which is within our estimated range. Teachers, journalists, NGO workers, and local officials provide more accurate estimates than bureaucrats in Kabul, suggesting local stakeholders may have more complete information (Olken, 2009).

[4]All statistics presented here are based on comparisons of the EE treatment and the control group.

a reduction in any delays by 24.9pp (from a base of 34.5% in the urban control group). By contrast, in rural areas, treatment still elevated delays by 19.5pp (from a base of 19.8% in the rural control group), even after two years. Despite this divergence in salary experience, after two years, a majority of employees in both urban and rural areas supported expanding the reform nationally. On average, treated employees expressed a degree of support that is 25.1pp greater than the 71% of control employees who supported the reform.

Third, the MSP reform improved downstream outcomes. The ultimate objective of the reform was to improve student learning by increasing employee performance. We conducted a direct learning assessment in households with students to measure learning impacts. Estimates indicate that students living near treated schools increased combined learning scores by 0.147σ one year into the reform, reflecting progress in both math and reading, and that these gains were concentrated in cities. Administrative records reveal higher teacher retention rates in treated schools, but not higher attendance, pointing toward greater employee enthusiasm as a possible mechanism behind these gains.

In addition, mobile salary payments increased employee adoption and usage of mobile money, an indicator of deeper financial inclusion. Using survey data, we estimate that self-reported mobile money use for peer-to-peer transfers increased by 27.2pp relative to 6.7% usage in the control group. Further, estimates using administrative data indicate employees increased their mobile money balances by 0.4 percentage points more of their salary for every month they are on the platform – approximately 2.8pp after the median period of 7 months. More broadly, employees began to make additional deposits into these wallets and used them to purchase airtime, consistent with more financial inclusion.

The final section of the paper studies how the success of the MSP reform relates to citizens' confidence that the state would endure. This exercise is motivated by theoretical models of political development that explore how political coordination can be influenced by beliefs about the durability and the nature of future institutions (Besley and Persson, 2009; Bidner and Francois, 2013; Acemoglu and Robinson, 2019). Such considerations are particularly relevant to fragile states, which are defined by both territorial contestation and the risk of radical institutional change. In the context of the MSP reform, we find that the reform was most effective in districts where citizens believed that the Taliban would lose the war (even controlling for general beliefs about government effectiveness, perceptions and actual data on security, and other covariates). We estimate that for every percentage point increase in the share of citizens agreeing with the proposition that the Taliban would lose, the share of teachers promptly registering for the biometric system increased by 0.994 pp. Placebo exercises indicate that it is only beliefs right at the beginning of the reform (the last

time they were measured), and not beliefs in preceding years, that predict success. Consistent with theory and the results reported in Callen, Weigel and Yuchtman (Forthcoming), which link beliefs to engagement with the formal state, we find that this positive association between beliefs that the Taliban would lose and implementation are localized to districts where citizens *agree* about that fact.

To summarize, this paper makes three main contributions relative to existing research on the development of state capabilities in fragile states. First, we provide experimental evidence and a corresponding set of lessons on whether and how digital identification and payment systems can help build the foundational capabilities of being able to track and pay employees. Second, we show that building these capabilities can improve salary experience for government employees and so secure their support for continued reform, and consequently improve downstream outcomes (in our case, student learning). Third, we provide empirical evidence that uncertainty about the durability of the state can undermine investments in state capabilities and the provision of basic services. Our findings have particular relevance for policymakers given that progress on the Sustainable Development Goals has either stagnated or declined in over half of fragile states (OECD, 2022; FCDO, 2023), and building capacity in these states is essential for continued progress.

Our paper links most closely to the literature on state formation, which describes how incipient states consolidate a monopoly on violence, extend authority into contested territories, build basic capacities, and evolve into modern states. The narratives on how territorial security facilitates state development vary significantly: Tilly (1985) argues that a state's monopoly on violence is essential for performing basic functions; conversely, Scott (1998) suggests that state development requires a co-option strategy where the state makes citizens legible through measures like standardizing last names. Cantoni, Mohr and Weigand (Forthcoming) demonstrates that the establishment of professional tax administration bureaucracies in 16th to 18th-century Europe marked significant progress in state consolidation and territorial expansion. Additionally, detailed accounts by Sanchez de la Sierra (2020) and Henn et al. (2023) describe how non-state actors come to take on key state functions. Recent experiments, including those by Weigel (2020), Acemoglu et al. (2020), and Berman et al. (2019), indicate that by delivering services or levying taxes, states initiate processes that enhance their legitimacy. Finally, in studying the development of state capacity, particularly in contested territories, our study complements prior research documenting efforts of the central state to either formalize local institutions (Beath, Christia and Enikolopov, 2017; Casey, Glennerster and Miguel, 2012; Blattman et al., 2023; Hartman, Blair and Blattman, 2021; Karim, 2020) or integrate informal institutions into state functions (Balán et al., 2022;

Sanchez de la Sierra et al., 2022; Haim, Ravanilla and Sexton, 2021).

The rest of the paper proceeds as follows. Section 2 describes the institutional environment in Afghanistan during our study. Section 3 describes the MSP reform, our research design, and experimental data sources. Section 4 presents results on ghost employees, salary experience, and downstream outcomes. Section 5 examines the relationship between citizen beliefs and implementation of the MSP reform. Section 6 concludes.

2 Institutional Details

During its modern history, Afghanistan has struggled to develop basic institutions. Starting with the Soviet invasion in 1979, over four decades of internal and international conflict have made Afghanistan one of the poorest and weakest states in the world. Even during 20 years of unprecedented levels of international support from 2001 to 2021, the U.S.-backed government lacked many basic administrative competencies, including reliably identifying and paying its workers (Special Inspector General for Afghanistan Reconstruction, 2022).[5]

The establishment of the National Unity Government in 2014 marked the beginning of a period of renewed effort to build the Afghan state. The newly formed National Unity Government marked the first power transfer since the US-led invasion, and coincided with NATO's implementation of the final phase of its transition to Afghan security forces following the 2009 surge. This period sparked widespread optimism among Afghan citizens, as the majority believed for the first time that the government could defeat the Taliban insurgency, as shown in Figure 1.[6] Still, significant uncertainty remained: in May 2018, at the onset of the MSP reform, 47.74% of respondents in the country thought the National Army would defeat the Taliban, 19.51% believed they would lose, and 31.69% were not sure either way.

The MSP reform was a flagship initiative for the new government, beginning in 2017 with a pilot of 1,200 employees in the Ministry of Labor in Kabul. After the pilot's successful completion, the government sought to test whether digital technologies could effectively deliver public salaries at scale, including in contested territories. The government selected

[5]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 public payrolls to be as high as 40 percent. Outside our study, the best estimate of which we are aware is from Special Inspector General for Afghanistan Reconstruction (2022): 13.3% of the Afghan National Defense Security Forces, which includes the police and the army, were likely ghost workers in 2018. This 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).

[6]The data in the figure are from the Afghanistan Nationwide Quarterly Assessment Research (ANQAR) program — NATO's large-scale public opinion polling program. Sections 3 and 5 provide more detail.

the Ministry of Education (MoE) for the scale-up given its central importance to the country's development and the critical importance to international donors of ensuring their support to public salaries was well-managed. At the time, MoE employed roughly 80% of the country's civilian public servants and was the third largest ministry (behind defense and interior) in terms of expenditure with an annual budget of 35 billion AFN (450 million USD).[7]

In theory, digital identification and payment systems appear well-suited to address several challenges involved in processing public salaries in fragile states, all of which were relevant in Afghanistan. First, identifying and removing ghosts requires an authoritative record of who is being paid. However, ministries in Afghanistan maintain payroll records separately from employee rosters, where the former is updated much more frequently than the latter given its essential operational function and is vulnerable to manipulation.[8] Second, Afghanistan's Tazkira national identification system is highly incomplete and subject to duplicate identifiers.[9] Thus, a biometric identification system that could uniquely identify employees receiving payments represented an attractive substitute for an unreliable national registry.

Third, public salary disbursements are often delayed, for two broad reasons in Afghanistan. First, the delay may originate from slow-moving 'first mile' approvals and administrative steps that must be taken in the MoE, the MoF, 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 little accountability to front line employees and lacked an effective protocol for quickly approving and processing salaries. A 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).[10] 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. At baseline, MoE employees report that 86.5% of delays are due to first mile actors and 22.9% to last mile

[7]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).

[8]A 2017 independent assessment noted that the Ministry of Education's (MoE) "payroll department is unable to provide an accurate figure on the number of people that it has paid over the past three months." Accounting for 68 percent of all civil servants, "of all Afghan ministries, it is MoE that has by far the most scope for using influence to place relatives and friends into jobs" (Monitoring, Committee et al., 2017).

[9]Tazkira cards are simple paper forms and when they are lost, citizens are issued a new one, with a new identification number. Consequently, it is possible for one person to have several identification numbers. Afghanistan also lacks a comprehensive government identification system.

[10]With mobile salary payments, employees immediately received an SMS payment confirmation, but still faced potential last-mile delays in withdrawing funds, which required accessing a mobile money agent. As we discuss below, these agents were concentrated in urban areas, leading to first-mile delays in rural areas.

actors. A fourth challenge is that anecdotal reports of leakage in public salary disbursement are common in fragile countries, particularly when physical cash must be transported.

In this context, financial development through traditional brick-and-mortar banks faces severe obstacles that mobile technologies may help address. Insecurity, limited economic activity, and low population density mean banks cannot operate profitably in most of the country. With a leaner and more scalable business model, mobile financial technologies can potentially leapfrog traditional finance in the same way that cellphones helped eliminate the need for landlines. Moreover, because mobile money is subject to substantial network effects (Katz and Shapiro, 1994), and the government is by far the largest employer, it is possible that moving government salary payment to mobile money could kickstart mobile financial development. The Afghan government thus hoped to catalyze a digital financial ecosystem by providing mobile money accounts to a large number of public servants.[11]

3 Research Design and Data

The research design aims to measure the extent to which the payroll reforms implemented by the Afghan government increased the state's capacity to identify and pay its employees. This section describes the experimental sample; provides details on the reforms; discusses treatment assignment and treatment compliance; and describes the administrative and survey data used to evaluate the reform.

3.1 Experimental Sample

Our experimental sample consists of the 34,422 MoE employees who appear at least once on official government payroll records between March 2017 and February 2020, ostensibly because they worked in one of 1,530 schools in Kandahar, Nangarhar, and Parwan provinces (spanning a total of 42 districts). These schools were divided into 401 experimental registration zones using information on schools' location, number of employees, and security. In consultation with the implementing partners, about 300 other institutions were immediately deemed to be inappropriate for the experiment because they belonged to other zones that were less secure and consisted in large part of non-typical schools (e.g., madrassas and vocational training institutes) or schools with less than ten employees. Subsequently, the mobile

[11] 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.

network operator appears to have been similarly unable to successfully operate in another 341 schools (where less than 10% of employees registered despite attendance rates and support for the reform being at least as high as everywhere else during baseline unannounced audit visits). These schools are less secure than others, and indeed the operator mentioned security as the reason for not working there for most of them. About half of them are in the control group, possibly reflecting both a worsening of the security situation and the limited capacity of the operator to register employees in control zones while implementing the full reform in the treatment zones. We remove these schools from the sample.[12] Nevertheless, our final sample contains some insecure regions, some non-traditional schools, and some small schools. Approximately 12% of the teachers in our sample are female and over half of the schools taught girls.

3.2 Details of the MSP Reform

The MSP reform involves three stages:

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

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

[12]Earlier versions of this paper included the 341 schools where the mobile network operator was non-functional within the experimental sample. At the time, we were unaware that these schools had been excluded by the operator. The original analyses retaining these schools are available in the CEPR working paper at `https://cepr.org/publications/dp18254`, or, identically at `https://tinyurl.com/2023Archive`. Including these schools introduces noise to the treatment indicator; notably, our estimates of impact on learning outcomes are much clearer when these schools, which were never effectively part of the experimental sample, are excluded.

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

3.3 Assignment to Treatment

The research design randomized the timing of the registration process and the implementation of mobile money payments at the registration zone level. At the beginning of the reform, the 401 experimental zones were randomly assigned to one of the three treatment arms, with treatment assignment stratified at the district level:

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

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

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

Figure 2 provides a map of the registration zones by status across our three provinces and Appendix Figure A.1 provides a satellite photograph of two registration zones. Appendix Table A.1 checks for balance using the survey data described in Section 3.5 below.

3.4 Treatment Compliance and Estimation

Figure 3 reports the timeline of the reform and data collection (top panel), the number of employees registered at each point in time for each treatment arm (middle panel), and the number paid digitally in each arm (bottom panel). The majority of employees in the EE and ED group were registered between May 2018 and July 2018. Most of the registration in the control group took place between November 2018 and December 2018, with the pace slowing in 2019 due to capacity constraints.

Employees in the EE group started to be paid by MSP in October 2018 in line with the research design. However, because there were insufficient mobile money agents, employees

were transition more gradually to mobile payments than planned; many employees in the EE group received their first mobile payment either in November or December 2018. During these months, about half of the employees in the ED group were also transitioned to mobile payments. This happened earlier than planned as the government wished to show progress on increasing the number of employees paid via mobile money. Only a small minority of employees in the control group were paid via mobile payments during the study period, starting mostly in February and April 2019. After April 2019, the number of employees paid via mobile money did not increase in the ED and control group, reflecting the persisting capacity constraints of the mobile network operator. In April 2019 (when our first endline survey occurred), 72% of employees in the EE group, 44% in the ED group, and 8% in the control group reported being paid by mobile money. In May 2020 (our second endline survey), these numbers increased to 81%, 49%, and 33%.

While compliance with the research design schedule was far from perfect, the randomization generated variation in the number of employees impacted by the reform across treatment groups, as displayed in Figure 3 and Appendix Table A.2. To account for imperfect compliance, in our main specification we estimate Treatment-on-the-Treated (ToT) effects by instrumenting an indicator for whether the respondent received their salary via mobile money with treatment dummies using this 2SLS specification:

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

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

3.5 Mobile Salary Payments Outcome Data

We use a combination of administrative and survey data to assess the impacts of the reform.

Payroll records: The government provided payroll records for all employees in our sample. These are generated when schools submit monthly requests for salaries to the MoF, and contain detailed information about schools, employees, and their salaries.

Registration data and mobile money transactions: The mobile network operator provided registration data and complete mobile money transaction records.

Audit surveys: These measure teacher attendance. Sampled schools were visited without prior warning, and all present teachers were briefly interviewed. We report results from two rounds of unannounced in-person surveys: May 2018 (baseline) and April 2019 (endline 1). We attempted to conduct these surveys in one randomly-selected school in each of the 401 registration zones, but due to security issues, the survey company could visit only 375 schools at baseline and 362 schools at endline 1.[13]

Detailed surveys: These surveys capture teachers' payment experience and support for the reform for a random sample of up to three teachers among those present during the audit. In total, we conducted in-person interviews with 1,005 teachers at baseline and 974 teachers at endline 1. We conducted another survey in May 2020 (endline 2) using a shortened survey instrument over the phone due to Covid-19.[14] In Appendix Table A.3 we report demographics for all teachers paid at least once before the reform, along with means for our survey samples and differences for our treatment arms. Our survey samples are representative of the population of teachers, and are very similar across time even though we do not interview the same teachers in baseline and endline 1, and only interview a subset of endline 1 respondents in endline 2. As such, our analysis in each wave, and comparisons of treatment impacts between wave, are likely not contaminated by sample selection bias.[15]

Employee verification 'litmus test': In addition, we conducted a litmus test survey in February–March 2019 that was designed to check whether individuals registered during the in-person drives were in fact legitimate employees. This phone survey conducted with 2,663 employees included a mix of employees present during audit visits, absent during those visit, and all employees who never registered (see Appendix C for sampling details). The survey asked for basic details about the employee's workplace and job that we would expect any genuine employee to be able to answer; questions included the school's geographic location, the principal's identity, and the rank of their position. We use this survey to help identify, ghosts (who could not be found), and stand-ins.

Learning assessment: In May 2019, we conducted a learning assessment of 1,001 chil-

[13]If a school could not be visited, we did not substitute another school in that zone. Our survey firm used extensive local contacts to enable interviews in some Taliban-controlled regions. Some surveys were successful in areas where the mobile network operator could not carry out the reform. We have data from 45 such schools in the baseline and 42 such schools in endline 1.

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

[15]Appendix Table A.4 also confirms that response rates do not differ across treatment arms for employees invited to participate in the survey.

dren aged 6–10 from a random sample of households who lived in the proximity of the schools selected for the in-person teachers' surveys. One child in each household was randomly chosen to take the learning assessment, which tested both literacy and numeracy skills.

Qualitative expert interviews: Finally, in 2019, we conducted semi-structured interviews with 76 key stakeholders (including teachers, government officials, journalists, and NGO workers) to gather their beliefs about the existence of ghost workers.

3.6 Beliefs About State Durability

We are interested in whether the success of the MSP reform relates to citizens' beliefs about state durability, which we extract from the Afghanistan Nationwide Quarterly Assessment Research (ANQAR) program. The survey includes a broad range of questions on politics and economics and covers approximately 13,000 Afghans in a repeated cross-section from 2008 until 2020, and is representative at the district level. We use the question "Do you think the National Security Forces (Army and Police) will be able to defeat the Opposing Government Elements (Taliban/Mukhal-feen-e dawlat) in the next few years?" Responses are on a five-point Likert scale, encoding a latent subjective probability that the Afghan National Army would defeat the Taliban.[16] As such, this question plausibly captures citizens' beliefs about future political institutions.[17] The institutional arrangement if the National Army prevailed would resemble that outlined in the constitution, including a commitment to regular elections, and protections for the rights of women and minorities. The Opposing Government Elements never strayed from their commitment to reimposing an Islamist theocracy. The data therefore provide measurement of beliefs about whether the government would endure.[18]

4 Results

This section (i) reports the impacts of biometric registration on removing "ghost workers" from the payroll, (ii) documents how mobile salary payments affected the salary experience of employees, and (iii) explores how the MSP reform affected downstream outcomes including

[16]The possible responses to this five-point question were 1 = "certainly not defeat", 2 = "most likely not defeat", 3 = "maybe", 4 = "most likely defeat", and 5 ="certainly defeat".

[17]In addition, because it regards a binary outcome, and is measured on a five-point scale, it encodes information about both the first and second moment of citizens' beliefs about state durability.

[18]In Section 5, we discuss how this measure links to theories that rely on beliefs about future institutions shape coordination and investments in state capacity, check the validity of the measure, discuss its specific link to the decision problem faced by MoE employees when deciding to register, and examine whether it predicts teachers' decision to enter the biometric registry.

student learning and financial inclusion.

4.1 Biometric Registration Impacts on "Ghosts" and "Stand-ins"

In implementing a biometric registration process, the Afghan government sought to create a single, reliable, biometric roster of employees. Government officials hoped that requiring employees to appear in person to enroll would immediately reduce the number of non-existent ghost workers on the payroll. Biometric verification was also required to withdraw salaries, providing an additional check against ghost workers.

In practice, and despite the threat that failure to register would result in losing salary, not all teachers registered: only 81% of employees eventually registered. Table 1 shows that registration was highest in the EE and EL registration zones (where registration took place before the mobile network operator had to manage the additional task of cashing out salaries), among more educated employees, for employees working at the schools designated as registration centers, and for employees living near primary roads and in districts that were not contested by insurgent groups.[19] Figure 4 provides a histogram of the share of teachers that registered in different zones in Panel A and a map of registration levels in Panel B. The figure reveals both substantial dispersion in the share of employees registered in a school, and also spatial correlation in the success of registration during the in-person drives. We will revisit the predictors of registration in Section 5, using a similar regression specification, when we consider the role of beliefs about state durability in predicting the decision to participate in the reform.

In the early registration (EE and EL) zones – where the MNO was most successfull at organizing registration activities – the majority of employees registered during the first registration wave, with only 2.8% failing to register eventually. We interpret this number as a conservative (lower bound) estimate of the number of ghosts on the payroll prior to the reforms (additional details on this calculation are in Appendix C). Their salaries amount to about AFN 3.3 million (USD 43,000 using the 2018/19 exchange rate) or 2.6% of the total monthly wage bill in these schools.

However, not all employees who failed to register were actually removed from the payroll, reflecting decisions made by the MoE adjudication committee tasked with eliminating ghost employees. Using payroll records, we can track whether MoE continued to request salaries

[19] Appendix Table A.5 reports similar analysis (and finds similar results), changing the dependent variable to be a dummy variable equal to one for employees registered at any time. Appendix Table A.6 also reports wildcluster bootstrap p-values; the only meaningful difference relative to Table 1 is that the estimate on the "contested district" dummy is no longer statistically significant.

for employees who never registered. Of the 2.8% of employees in early registration zones who failed to register, 89.6% remained on the payroll one month after the start of the reform, and 56.9% remained after 12 months. As a result, the corresponding payroll savings were limited to only 1% (AFN 1.3 million or USD 17,000) of the total monthly wage bill in the EE and EL zones. Even so, back of the envelope calculations indicate the MSP reform had a return on investment of 50% by year 2.[20]

Stand-in Employees: To game biometric registration, corrupt actors could send stand-ins to register, thereby maintaining salary lines that were previously associated with nonexistent ghosts. To better quantify the extent of this problem, we conducted a litmus test described in Section 3.5 that asked employees seven simple questions about their job, which any legitimate employee should be able to answer easily: (1) the name of the school in which they work; (2) the district in which their school is located; (3) the employee's rank and (4) position; (5) the principal's name; (6) the headmaster's name; and (7) the total number of employees working in the school.

As summarized in Figure 5, employees who never registered were less capable of answering these questions than employees who registered. To start, while only 21% of registered employees were not available to take the test, this share is 36% for employees who never registered ($p < 0.01$), corroborating the argument that those who never registered are more likely to be entirely fictitious. Among employees who took the test, those who did not register are 7.8 percentage points (s.e.= 2 pp) less likely to score seven, 11.3 percentage points (s.e. = 2 pp) less likely to get a score of six, and 23.2 percentage points (s.e. = 3 pp) more likely to score three or less.

In Appendix C, we conduct further bounding exercises to account for imperfections in our litmus test. The idea behind these exercises is simple: the lower the scores of registered employees with respect to the scores of known true employees (who were certified to be present during a separate round of unannounced audit visits), the higher the number of stand-ins. We find that estimates of stand-ins range from 5.9% to 18.1% of registered employees (5.7% to 17.6% of all employees), depending on how strictly we mark the litmus test. Adding these potential stand-ins to the 2.8% of unregistered employees, we estimate that between 8.5% and 20.4% of all paid employees might be ghosts. If they were all removed from the

[20]The MSP reform entailed a 195 AFN registration fee per employee enrolled (with total registration costs amounting to AFN 2.7 million) and a 100 AFN monthly fee per individual salary transfer (for a total of AFN 16.7 million calculated using mobile money transaction data). Savings from salaries not paid to unregistered employees removed from the payroll totaled AFN 29.2 million. Both savings and costs are calculated from the start of the reform in May 2018 until March 2020, the last month for which data is available.

payroll, the government would save AFN 26 million (USD 0.34 million) per month in the early registration zones.

Qualitative Evidence: To provide some qualitative context for these estimates, we collaborated with Integrity Watch Afghanistan – one of Afghanistan's premier NGOs focused on governance – to conduct semi-structured interviews with 116 key informants with domain knowledge of public education in the three experiment provinces. This sample includes seven local elites, 16 teachers, 15 local education officials, 40 provincial and district officials, 11 senior officials in Kabul, 12 journalists, five ex-government officials, and 10 NGO workers. We asked these respondents,"For every 100 employees on the MOE payroll, how many would you guess are ghost workers?" This is a challenging quantity to estimate, and 40 of our 116 respondents indicated that they had no clear estimate. The remaining 76 on average estimate that 11.82% (median = 9.75%) of teachers are ghosts (standard deviation = 15.02pp). This estimate is within the above estimated range of 8.4% to 20.4% of all employees.[21]

4.2 Mobile Salary Payment Impacts on Employees

The MSP reform's second component was the transition of teacher salary payments from a hybrid bank and trusted agent system to an integrated mobile money system.[22] Policymakers hoped that this reform would reduce leakage of salary payments, particularly because of potential problems related to physically transporting cash (Special Inspector General for Afghanistan Reconstruction, 2022). In practice, however, we find little evidence of leakage both before and after the reform based on employees' self-reported payments to receive their salary.[23] On average, control group employees report paying less than AFN 19 (USD 23 cents) to receive their salary in year one, which amounts to less than 1% of the average monthly salary. This observation is corroborated by our key informant interviews. After two years, treatment reduced leakage essentially to zero (Table 2 column 5, panel B).

[21]Interestingly, the estimates of senior government official (whose average estimate was 6%) were less accurate than those of stakeholders in the field, including journalists (13.9%); NGO workers (13.8%); teachers (12.3%), and provincial and district officials (13.4%).

[22]In Appendix Table A.7 we check whether registration *per se* impacted the outcomes discussed in this section, and do not find evidence that it did.

[23]We initially planned to measure leakage as the difference between the official salary displayed in the payroll records for an employee in a given month and the corresponding amount that the employee reported receiving in our survey for that month. However, due to the frequent delays in salary payment documented below (and the fact that often salaries for multiple months were disbursed at the same time), this simple mapping of monthly salaries proved challenging: employees could report the amount of the last salary they received, but they could not always say to which month of work that salary referred.

While the reform did not significantly reduce leakage, it did significantly impact delays in salary payment. These delays resulted from a combination of "first mile" delays in the capital due to bureaucratic approval processes at MoE, MoF and Kabul Bank, as well as "last mile" delays in getting the final payment to the employee.[24] Before the reform, 59% of employees in the control group and 63% of employees in the treatment group experienced delays (Appendix Table A.1). In the subsequent two years of the reform, delays were substantially reduced for all employees — including in the control group. After one year, 44% of control employees experienced delays with treatment increasing delays by a substantial 38.6 pp of employees (Table 2 column 2, panel A). After two years, delays fell further in the control group to 28.4% of employees, with treatment having a modest and insignificant reduction of 5.4 pp (Table 2 column 2, panel A).

Our analysis indicates that the reductions in delays – for both the treatment and control groups – were due to improvements in centralized bureaucratic (first mile) obstacles to salary delivery. This can be seen by comparing pairs of columns in Figure 6.[25] At baseline, both the treatment and control groups identify first mile delays as the major source of concern, but few report last mile delays. At endline, first mile delays were reduced in the treatment group by 30.8 percentage points ($p < 0.01$); in the control group, the share experiencing first-mile delays fell from 45.6% to 32.0% (13.6 pp; $p < 0.01$). The fact that first mile delays fell more in the treatment group is consistent with the Treasury Department's initially focus on improving salary approval protocols for treated employees.

By contrast, the payment reform initially *increased* last mile delays in the treated group, from a baseline of 10.8% to 36.2% in April 2019 (Figure 6). This is largely due to teething problems faced by the mobile network operator, which lacked capacity to quickly cash out digital payments in rural areas during the first year of the reform. For instance, Appendix Figure A.2 highlights that there were fewer active mobile money agents when the program launched, and many of those agents had very little prior experience. However, by the end of the second year, delays were reduced substantially for all employees (Figure 6). The same pattern is reflected in travel time, with increases due to treatment in year one, and no difference between treatment and control in year two (Table 2 column 4).

For the reform to succeed, it needed the support of employees, particularly given the

[24]In addition to the slow government process of approving salaries, government salaries provide a key source of liquidity for Kabul Bank, providing incentives to move slowly.

[25]The figure is based on survey questions that separately asked employees about their experience of first and last mile delays. It is plausible that employees could separate the two: in the control group employees or their proxy trusted agent would be told at a bank that the salary was not ready; in the treatment group employees would receive an SMS notification when funds posted, whether or not an agent was available to cash them out.

opposition from those benefiting from the existing system. The government was especially focused on whether employees would vote to expand the reform nationwide.[26] At baseline, before any reforms were implemented, support was very high among employees (94% in the control group), likely due to a pervasive dissatisfaction with the status quo ante payment system (only 26.9% were satisfied in the control group at baseline, while 59.6% experienced delays; see Table 2, Panel A). However, as shown in Table 2 (Panel A, column 5), support for the reform dropped during the first year of the reform, due to the implementation issues discussed earlier.[27] As these initial problems were resolved, support for the reform rebounded (Panel B of Table 2). First-hand experience with mobile salary payments increased support, particularly in the second year as the mobile network operator increased capacity.

Heterogeneity Between Urban and Rural Areas The difficulty of implementing the reform, particularly in rural areas, did not come entirely as a surprise. Indeed, there was debate among government stakeholders over whether the reform should begin in the cities or in the countryside, in part given the contested security situation in rural areas.[28] The government selected Kandahar, Nangarhar, and Parwan as the three provinces for the experiment precisely because they contained major government-controlled cities (Kandahar City, Jalalabad, and peri-urban areas of Kabul) and rural areas actively contested by the Taliban, such that the experiment could provide evidence to help resolve this debate.[29]

In this regard, the results are clear: the reform ultimately proceeded much more effectively in cities. In Table 3, we find that in the first year, the reform had most and statistically insignificant impacts on payments to receive transfer (Panel A, column 1), delays and travel times increased everywhere quite substantially (columns 2 and 3), with both increases larger in rural areas. Correspondingly, payment satisfaction was done, especially in rural areas

[26]Employees were asked, "If you were asked to vote on whether the entire ministry should switch to mobile salary payments, would you be in favor or against such a change? Please note this is a hypothetical question and you will not actually be asked to vote on this decision." The government intended this three-province pilot to be a test run for nationwide scaling, and considered responses to this question when considering the value of the reform. Employees were aware that their responses could influence whether the program continued, but it was made clear to employees that the question was hypothetical.

[27]As can be seen in Appendix Figure A.3, the most common reasons to support or oppose the reform were issues of accessibility and timeliness, highlighting the central importance of program implementation.

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

[29]We define urban zones as those with a population density of greater than 300 inhabitants per sq km.

(panel A, column 4). Yet, support for expanding the reform was up in urban areas, where the reform was showing signs of improvement even by the end of the first year (panel A, column 5), and mobile money transactions (excluding the salary transfers themselves), were up everywhere.[30]

In year 2, matters improved. Payments to receive salary were down everywhere (panel B, column 1), delays were reduced in urban areas, but remained (less) elevated in rural areas (panel B, column 2). Correspondingly, treatment had much more positive impacts on satisfaction in urban than in rural areas (panel B, column 4), and there was substantial support to scale the reform once again throughout the sample (panel B, column 5).

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

In the appendix, we examine the effectiveness of the reform in government-controlled and insurgent-controlled areas. We gauge government control in one of two ways: either by assessing whether non-local survey enumerators are permitted to conduct surveys without seeking permission from local commanders, or relying on an independent assessment of Taliban control from the Long War Journal.[31] According to both criterion, all urban areas are classified as being under government control. However, we can differentiate rural areas based on whether, relatively speaking, they were under government control or contested/controlled by the Taliban. The results are presented in Appendix Tables A.8 and A.9, respectively. Overall, our findings indicate that the reform's progress was concentrated in areas under government control compared to those under insurgent control.

4.3 Impacts on Student Learning and Financial Inclusion

Finally, we examine the downstream impacts of the MSP reform, focusing on student learning and employee financial inclusion. We find evidence of increased learning, mainly localized in urban areas, where the payment experience improved more quickly. We also find that treated employees were more likely to use their mobile wallets for deposits, airtime purchases, and peer-to-peer transfers.

[30] Appendix Figure A.4 shows using mobile account data that rural users took longer relative to urban users to access their salaries prior to endline 1, but that outcomes in the two groups converged by late-2019.

[31] For the former, we adopt the method in Wright (2023) and thank Austin Wright for sharing these data.

4.3.1 Impacts on Student Learning

The ultimate objective of the reform was to improve student learning through better employee performance. Since the government did not plan to collect reliable and systematic learning outcomes in our schools, we conducted our own learning assessment in May 2019, one year after the reform began (described in Section 3.5).[32] Given pandemic-related restrictions on fieldwork, we were only able to collect learning outcomes after the first year. Our sample is 828 school-going children, aged 6-10 who have completed an average of 2.25 years of education. We complement this data with administrative records of employee retention and attendance to understand mechanisms linking the reform to improved learning.

Students living near treated schools perform better in learning assessments. Table 4 reports the results.[33] Across the sample, combined learning scores increase by 0.147σ ($p < 0.1$; Panel A, Column 1), which is reflected in both increased math and reading scores. Consistent with employee support for scaling the reform increasing more quickly in cities, education impacts are larger in urban areas. Math scores increase by 0.335σ ($p < 0.01$), and reading scores increase by 0.295σ ($p < 0.05$). By contrast, and likewise consistent with the more substantial issues affecting the reform in rural areas, point estimates on learning outcomes in rural areas are negative, though statistically insignificant. Panel B of Table 4 reports corresponding reduced form estimates, which also indicate increases in learning, particularly in cities.

How is it possible that learning gains realized in a single year? In Appendix Table A.11, we estimate the impact of the reform on teacher retention (noting that estimates in Table 2 and 3 indicate that treated employees are more supportive of the reform). We focus specifically on the 28,099 employees who received at least one salary payment at some point before the learning assessment. This reflects the stock of employees who could possibly teach a student in the relevant timeframe. Among this group, about 14% of the employees in control registration zones joined during the first year of the experiment (after registration and before the learning assessment). The share of newly joined employees is lower by 2.14 pp ($p < 0.05$) in EE zones and 1.88 pp ($p < 0.01$) in EL zones. This, in turn, is reflected

[32] We thank Dana Burde and Joel Middleton for sharing with us the learning assessment methodologies they validated for use in Burde and Linden (2013), Burde, Middleton and Samii (2017), and Burde, Middleton and Wahl (2015). We validate our learning assessments in Appendix Table A.10, finding that learning correlates in the predicted direction with years of education, gender, and whether children live in a house with a maintained road, access to water, and access to electricity.

[33] Panel A presents ToT estimates and Panel B reports the reduced form. To obtain ToT estimates, we define a student as treated if at least 50% of the teachers in the school that they attend received a payment via mobile money prior to the learning assessment, and we instrument this with the school's randomly assigned treatment status.

in employees in treatment schools having worked for 0.55 years longer in EE zones and 0.48 years more in EL zones, relative to a control mean of 10.725 years. Consistent with the pattern of increased learning, retention is greater in urban treatment zones, though not significantly different from estimates on the rural sample. While not by any means dispositive, this pattern is consistent with greater enthusiasm among employees, hopeful the reform would eventually lead to improvements. By contrast, in both administrative records, and in our primary audits, we do not find evidence of increased teacher attendance (see Appendix Tables A.12, A.13 and A.14).

4.3.2 Impacts on Employee Financial Inclusion

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

The survey evidence discussed above indicates that the reform increased peer-to-peer transfers by 27.2pp from a base of 6.7%. To further understand if and how the salary reforms impacted how employees used their mobile wallets, we draw on rich administrative data from the mobile phone company. We focus our analysis on estimating how the length of exposure to mobile salary payments increased the use of other mobile financial services, using the following dosage-response specification:[34]

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

In the above, $Months_{izd}$ indicates the number of months that employee i from registration zone z in district d had received mobile payments at the time of the last recorded transaction. To facilitate interpretation, we normalize the outcomes Y_{izd} by the total monthly salary received in the month prior to the start of the reform. We estimate (2) using transaction-level data from the nearly 15,000 employees that were paid via mobile money from October 2018 to February 2020. Our results, in Table 5 show the effect (β) of an additional month of receiving mobile salaries, both using an OLS specification (Panel A) and an instrumental

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

variables specification (Panel B) in which we instrument for $Months_{izd}$ using the randomly assigned treatment dummies.[35]

The first column of Table 5 shows the impact of the reform on mobile money deposits, which we consider to be a particularly important measure of adoption since it requires employees to add money to their wallet beyond what they receive for their salaries. In the first month of use, employees deposit a statistically insignificant 0.2 percentage points of their salary (shown in the '1-Month Mean' row). The ToT estimates (Panel B) indicate this increases by 0.4 percentage points for every month they are on the platform. The median time on the platform observed in our data is 7 months; this implies an increase in deposits of 2.8 percentage points on average. Transfers to others via mobile money (column 2) go up by a similar amount (though this is not significant after instrumenting). Airtime purchases (column 3) and transfers of airtime to others (a basic version of a mobile money transfer that can be done to subscribers without a mobile money wallet) also increases considerably (column 4). Balances go up by 1.2 percentage points, or by 8.4 percentage points at the median (though this effect is smaller and not significant after instrumenting).[36]

5 Does State Fragility Deter State Capacity?

5.1 Fragility, Capacity, and the Role of Beliefs

The reform we study was implemented by a fragile state that wanted to increase its capacity.[37] However, building state capacity demands effort and coordination, with delayed benefits.

In other words, the decision to make such investments may depend on beliefs about the state's durability. If the state faces a risk of radical transformation, this uncertainty will manifest in citizens' beliefs about future institutions and could, in turn, deter the efforts needed to build capacity. Similarly, the effort to build capacity might depend on higher-order beliefs. When citizens trust that the state will endure and are confident that others share this belief, coordinated investments – like the major reform we study – are more likely to succeed. Conversely, disagreement about the state's future could deter investments in

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

[36]Table A.15 reports results from a similar specification that interacts $Months_{izd}$ with whether the sample is urban or rural. The ToT estimates (Panel B) indicate that the financial inclusion benefits of the MSP reform are generally not statistically distinguishable between urban or rural areas.

[37]Definitions of state fragility often highlight systemic risks such as potential state failure, violent displacement of the ruling regime (Besley and Persson, 2010, 2011*b*), or other radical institutional changes, coupled with very low capacity. The OECD, for instance, describes fragility as "exposure to risk and insufficient coping capacities of the state to mitigate those risks" (OECD, 2022).

state capabilities.

Thus, the coordination of investments in state capacity by state actors and citizens plausibly depends on both the *level* of their beliefs about the state's durability and the *degree of consensus* regarding future institutions. This dependency — linking low state capacity with the risk of radical institutional change — could help explain the enduring nature of fragility as both a "syndrome" and a "trap" (Besley, Collier and Khan, 2018).

These broad considerations are reflected in the employees' decision to advance this reform. We can test two hypotheses with our data related to first-order and second-order beliefs, respectively. First, whether teachers register for the system depends on current perceptions of the Afghan government's durability. Second, this decision is influenced by the extent of consensus in these beliefs.

5.2 Beliefs and Employees' Decision to Register

The success of the reform depended directly on teachers registering for the biometric system, and the share signing up provided the government a signal of the degree of support for the reform. Signing up provides an example of a complex decision to participate in advancing state capacity, a choice likely partly driven by beliefs about the Taliban's potential takeover. Indeed, many teachers chose not to sign up, as shown in Table 1. The MoE indicated that registration was necessary to continue receiving salary, promising benefits such as more timely disbursements, the advantages of a mobile money wallet, and participation in a significant reform aimed at improving the ministry. However, it was also quite plausible the reform would not be seen through, that the requirement to sign up would not be enforced (which proved to be true), and that salary disbursements might face delays due to teething issues. Crucially, it was unlikely that the Taliban would continue the reform if they came to power (they immediately cancelled the reform after taking control).

Several potential benefits of the reform, such as the network externalities from a large-scale rollout of mobile money, and even the sustainability of the reform itself, plausibly depended on the overall share of participating teachers. As such, employees' beliefs about how many of their colleagues would register also likely impacted their own decision to register.

For these reasons, the rate of employee registration is a key metric of interest when considering the role of beliefs. It reveals whether employees perceived the net present benefits of registering to outweigh the costs. This calculus, in turn, could reflect employees' first- and second-order beliefs about the durability of the Afghan state.[38]

[38]We focus on registration during the drive rather than at any time, as late registrants are more likely to

5.3 Proxying for Employees' Beliefs

Ideally, testing whether beliefs about state durability influence participation would involve directly and incentive-compatibly measuring employees' beliefs about state durability.[39] Instead, we use the ANQAR question about who would win the war, discussed in Section 3.6, as a proxy. We proxy for employees' first-order beliefs using the beliefs of citizens living in their same district. This requires an assumption that teachers are similar to the general population in terms of these beliefs, which is supported by the data.[40] Furthermore, assuming teachers understand the distribution of beliefs in their community, we proxy for their second-order beliefs by using the dispersion of beliefs across respondents in their district. There are a substantial number of respondents in each district-year, allowing the standard deviation to provide a plausible substitute for employees' beliefs about the spread in the distribution.[41]

Validating the Measure of Beliefs: Four features of this measure indicate it captures beliefs about state durability. First, Afghans exhibit substantial uncertainty about the country's political future (see Figure 1), consistent with when these measures were taken. Second, these beliefs respond to news shocks (Figure 1). Third, when the ANQAR data are collapsed into a district-panel spanning 2008-2018, these beliefs show sensible correlations with a broad range of security measures, police and army presence, governance quality, and economic conditions, both nationwide and in our experimental sample during the relevant period (Appendix Figure A.6). Fourth, these measures predict the likelihood of citizens using formal state structures instead of informal institutions for dispute resolution (Callen et al., 2023).[42]

be stand-ins. Additionally, a link between late registration and beliefs about state durability could reflect corrupt actors seeking a stream of salary payments for stand-ins.

[39]This would need to overcome potential social desirability bias, as asking government employees about their employer's durability could influence their responses.

[40]Appendix Figure A.5 shows that the belief distribution for government employees in our experimental region in 2018 is comparable to that of the general population.

[41]Our measure encodes a subjective probability that the army would prevail (a binary event) on a five-point scale, providing information about citizens' confidence and first-order uncertainty. The standard deviation reflects disagreement among citizens. In Appendix Table A.16, we analyze the average response, the implied average first-order confidence, and the standard deviation among respondents, finding the reform most effective where citizens were confident, certain, and agreed on the state's durability.

[42]Beliefs about state durability are likely to influence whether citizens use state-run institutions for dispute adjudication, as the binding nature of court decisions depends on the continued existence and power of enforcement of these courts. The results in Callen et al. (2023), which examine the use of formal versus informal institutions and beliefs about state durability, closely align with the findings reported here.

5.4 Analysis

To test whether beliefs about who would win the war predict participation in the reform, we estimate the specification:

$$\begin{aligned}\text{Registered During Drive}_{id} = \beta_1 &\text{Taliban will lose [mean]}_d \\ &+ \beta_2(\text{Taliban will lose [mean]}_d \times \text{Taliban will lose [std dev.]}_d) \\ &+ \beta_3 \mathbf{\Gamma X}_d + \varepsilon_{id}\end{aligned}$$

where Registered During Drive_{id} is a dummy variable equal to one if employee i in district d registered during the in-person drive at their school, Taliban will lose $(\text{mean})_d$ is the average share of ANQAR respondents across all 2018 survey rounds who indicate that they believe that the Taliban will lose in the next few years and $\mathbf{X}_d$ is a vector of individual- and district-level covariates. In some specifications, $\mathbf{X}_d$ includes beliefs measures from years prior to the reform to provide a placebo, and also contains the same vector of individual and district-level covariates used to predict registration in Table 1. To test the importance of consensus of these beliefs, we interact the mean belief with the within-district standard deviation of responses to the five-point Likert scale question.[43]

Table 6 reports the corresponding estimates. We see that a having a higher share of respondents who believe the Taliban will lose positively predicts whether employees appear during the in-person registration drive (columns 1 and 3). Similar measures of beliefs from prior years do not substantially explain whether employees appear for registration (column 2). Above, we argue that because investments of this sort require coordination, their success will depend not only confidence that the state will endure, but also relative consensus about that fact. We test this in columns 4 and 5 by interacting the level of beliefs with the within-district standard deviation during the first year of the reform (2018). We see that the relationship between a belief that the Taliban would lose and whether an employee registers is indeed localized to districts where there is agreement about that fact. This is also depicted graphically in Figure 7.

Estimates reported in this table are estimated using only employees who registered during the initial in-person drive. We do this to help with the concerns that corrupt officials might later register stand-ins to preserve places on the payroll – which might be more likely in places where they believed the state would endure (and so provide a reliable stream of

[43] Among the 28,945 employees recorded in payrolls during the 12 months before the reform, 19,626 worked in schools located in districts with ANQAR data where more than 10% of employees registered.

rents). In Appendix Table A.17 we report estimates from a similar specification, checking whether employees appeared at any point for registration. This specification is subject to the concern about stand-ins, but also provides an additional test of the same two hypotheses. We find that a similar set of results obtain.[44]

The potential conceptual link between beliefs about state durability and effort to enroll in the MSP system we outline in subsection 5.2 is causal. If employees are confident and agree that the state will endure, this should encourage action and coordination to build state capacity. In contrast, the estimates reported in Table 6 cannot be determined to be causal. It is the case that beliefs contemporary to the reform predict registration, while beliefs in prior years do so less and, and adding a covariates for a broad range of plausible confounds (beliefs about contemporary state effectiveness, measures of police and army presence, and other measures that strongly predict registration) in addition to beliefs from years prior to the reform does not move estimates.[45] In addition, the pattern of beliefs predicting registration in districts with relative consensus about state durability depicted in Figure 7 is consistent with the conceptual arguments above.[46]

6 Conclusion

The effort to build democratic institutions in Afghanistan from 2001 to 2021 was an ambitious project aimed at building an effective and accountable state in a challenging environment. The MSP reform we analyze was a critical component of this broader state-building project, and was essential for developing the two foundational capabilities of identifying and paying employees. This reform took place at a key moment of public optimism as the newly formed National Unity Government marked the first power transfer since the US-led invasion. Simultaneously, NATO executed the final phase of its transition to Afghan security forces

[44]In Appendix Tables A.22, A.23, and A.24, for completeness, we check for heterogeneous treatment impacts, dividing the sample at the median of beliefs that the Afghan government would win the war, and checking impacts on payment experience, student learning, and mobile money use.

[45]One exception is that adding province fixed effects eliminates the relationship between beliefs and registration (Appendix Table A.18). However, because the question concerns beliefs about the federal government and participation in a federal program, it is reasonable to consider between-province variation.

[46]Although the first and second moments of an individual's beliefs about the war's outcome are inherently linked through the Likert scale's encoding of latent subjective probabilities, this connection does not necessarily extend across different employees. In Appendix Table A.16, we examine whether registration rates are driven by an individual's level of confidence and certainty, as captured by their Likert scale responses, or by the overall degree of social confidence, measured by the standard deviation of responses within a district. The highest registration rates occur in areas where individuals are confident in a government victory and where there is relatively little disagreement among respondents

following the 2009 surge. Against this backdrop, the Afghan government faced the urgent need to stand on its own, develop capacity, and address a potent insurgency if it wanted to survive.

This paper makes two primary contributions. First, working across three Afghan provinces with around 34,000 employees and contested by multiple insurgent groups, we provide experimental evidence on whether and how technology can improve a government's ability to identify and pay its workers. Prior to the reform, these processes were abysmal, as the government possessed three inconsistent employee rosters and roughly 55% of employees experienced substantial salary delays. In the initial year, the reform exacerbated these issues, increasing both the delays and the travel times required to retrieve salaries. After two years, delays were reduced across the sample, and were reduced particularly in government-held cities (though they remained substantial and elevated in rural areas). At the same time, the reform showed signs of promise. By the end of the reform, most employees – and especially those who had used the system – supported scaling it nationwide. This support impacted policy: just after this study concluded in 2021, the government decided to scale up the reform to eight additional provinces. It also increased financial uptake, as employees began making peer-to-peer transfers, deposits into their mobile wallets, and progressively accumulating a savings balance. More speculatively, there is some evidence that the reform began to improve education. Taken together, these results make clear that technology was not an instant panacea, but that given sufficient time and focused in the right areas – like where the state has control – it can assist building state capacity in fragile states.

Second, we provide evidence on how the risk of radical institutional change constrains state development. This reform, aimed at building essential capacity, showed preliminary progress only after two years. It was an investment requiring coordination and adjustment costs, intended to yield long-term benefits if the state endured. The worth of such an investment depends on whether people believe the state will continue, which was uncertain. Indeed, local beliefs about who might win the war robustly predict where the reform took root. This suggests that a defining feature of fragility – the risk of radical institutional change – can impede the development of state capacity. This insight is actionable and consistent with the view that a monopoly on violence is a prerequisite to state development (Tilly, 1985): escaping the fragility trap requires creating confidence and consensus that the state will endure. A corollary insight is that measuring beliefs about future institutions provides an empirical measure of the degree of fragility and potentially whether a state is at a critical juncture. Our second contribution, therefore, is to provide evidence that uncertainty about the durability of the state can undermine investments in state capabilities and the provision

of basic services.

The two goals of this paper are to provide evidence on *what* helps to build capacity and, more conceptually, *why* fragility can stand in the way. Our setting carried all the hallmarks of fragility: stagnant state capacity, pronounced discrepancies between the *de facto* and *de jure* territorial extents of the state, alongside widespread vulnerability and deprivation. Globally, state fragility is stopping progress on vital development goals, like ending poverty, eliminating hunger, and providing access to basic education. As emphasized in a recent report by the Foreign, Commonwealth and Development Office, "We have to understand the importance of fragility and conflict in shaping development outcomes. Conflict has stalled or even reversed progress in too many places, with humanitarian needs at their highest since 1945. Soon, half the world's poorest will reside in fragile or conflict-affected states, characterized by inadequate state capacity, limited respect for the rule of law, high levels of corruption, and persistent ethnic or political divisions" (FCDO, 2023). We contribute to this conversation by providing evidence that technology can help, with important caveats, and that the risk of radical institutional change provides a major obstacle to the development of state capacity.

References

Acemoglu, Daron, Ali Cheema, Asim I Khwaja, and James A Robinson. 2020. "Trust in State and Nonstate Actors: Evidence from Dispute Resolution in Pakistan." *Journal of Political Economy*, 128(8): 3090–3147.

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

Acemoglu, Daron, and James A. Robinson. 2019. *The Narrow Corridor: States, Societies, and the Fate of Liberty.* Penguin.

Allen, Robert C, Mattia C Bertazzini, and Leander Heldring. 2023. "The economic origins of government." *American Economic Review*, 113(10): 2507–2545.

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

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.

Beath, Andrew, Fotini Christia, and Ruben Enikolopov. 2017. "Can development programs counter insurgencies?: Evidence from a field experiment in Afghanistan."

Berman, Eli, Michael Callen, Clark Gibson, and James Long. 2019. "Elections and Government Legitimacy in Afghanistan." *Journal of Economic Behavior & Organization*, , (168): 292 – 317.

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. 2010. "State Capacity, Conflict, and Development." *Econometrica*, 78(1): 1 – 34.

Besley, Timothy, and Torsten Persson. 2011*a*. "The logic of political violence." *The quarterly journal of economics*, 126(3): 1411–1445.

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

Besley, Timothy, Paul Collier, and Adnan Khan. 2018. *Escaping the Fragility Trap: Report of the Commission on Fragility, Growth, and Development.* International Growth Center (IGC).

Bidner, Chris, and Patrick Francois. 2013. "The emergence of political accountability." *The Quarterly Journal of Economics*, 128(3): 1397–1448.

Blattman, Christopher, Gustavo Duncan, Benjamin Lessing, and Santiago Tobón. 2023. "Civilian alternatives to policing: Evidence from Medellín's community problem-solving intervention Operación Convivencia."

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

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

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

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

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

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

Callen, Michael, Jonathan L. Weigel, and Noam Yuchtman. Forthcoming. "Experiments About Institutions." *Annual Review of Economics*.

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

Cantoni, Davide, Cathrin Mohr, and Matthias Weigand. Forthcoming. "The Rise of Fiscal Capacity: Administration and State Consolidation in the Holy Roman Empire." *Econometrica*.

Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping institutions: Evidence on aid impacts using a preanalysis plan." *The Quarterly Journal of Economics*, 127(4): 1755–1812.

Collier, Paul. 2008. *The bottom billion: Why the poorest countries are failing and what can be done about it.* Oxford University Press, USA.

Dell, Melissa, Nathan Lane, and Pablo Querubin. 2018. "The historical state, local collective action, and economic development in Vietnam." *Econometrica*, 86(6): 2083–2121.

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

Dincecco, Mark. 2015. "The rise of effective states in Europe." *The Journal of Economic History*, 75(3): 901–918.

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

Fearon, James D, and David D Laitin. 2003. "Ethnicity, insurgency, and civil war." *American political science review*, 97(1): 75–90.

Foreign, and Commonwealth Development Office (FCDO). 2023. "International development in a contested world: ending extreme poverty and tackling climate change."

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

Haim, Dotan, Nico Ravanilla, and Renard Sexton. 2021. "Sustained government engagement improves subsequent pandemic risk reporting in conflict zones." *American Political Science Review*, 115(2): 717–724.

Hartman, Alexandra C, Robert A Blair, and Christopher Blattman. 2021. "Engineering informal institutions: Long-run impacts of alternative dispute resolution on violence and property rights in Liberia." *The Journal of Politics*, 83(1): 381–389.

Henn, Soeren J, Christian Mastaki Mugaruka, Miguel Ortiz, and David Qihang Wu. 2023. "Monopoly of Taxation Without a Monopoly of Violence: The Weak State's Trade-Offs From Taxation." *Review of Economic Studies*.

Karim, Sabrina. 2020. "Relational state building in areas of limited statehood: Experimental evidence on the attitudes of the police." *American Political Science Review*, 114(2): 536–551.

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.

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

Monitoring, Independent Joint Anti-Corruption, Evaluation Committee, et al. 2017. "Ministry-wide vulnerability to corruption assessment of the Ministry of Education." *Kabul, Afghanistan*.

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

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

North, Douglass Cecil, and Robert Paul Thomas. 1973. *The rise of the western world: A new economic history*. Cambridge University Press.

Olken, Benjamin A. 2009. "Corruption Perceptions and Corruption Reality." *Journal of Public Economics*, 93((7-8)): 950 – 964.

Olson, Mancur. 1993. "Dictatorship, democracy, and development." *American political science review*, 87(3): 567–576.

Organization for Economic Cooperation and Development (OECD). 2022. *States of Fragility 2022*.

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

Sanchez de la Sierra, Raúl, Kristof Titeca, Haoyang Stan Xie, Albert Malukisa Nkuku, and Aimable Amani Lameke. 2022. "The Real State: Inside the Congo's Traffic Police Agency."

Scott, James C. 1998. *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press.

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

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

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

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

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

World Bank. 2024. *Identification for Development (ID4D) and Digitalizing G2P Payments (G2Px) 2023 Annual Report.* World Bank Group.

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

Figures and Tables

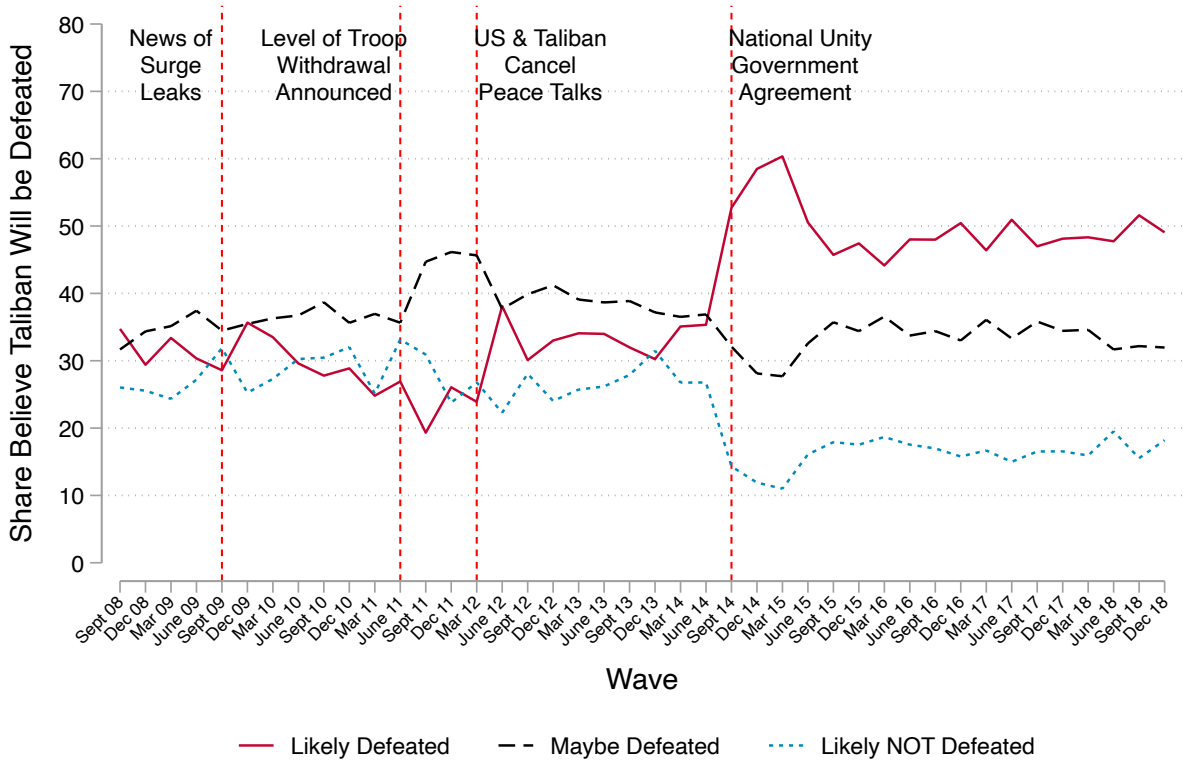
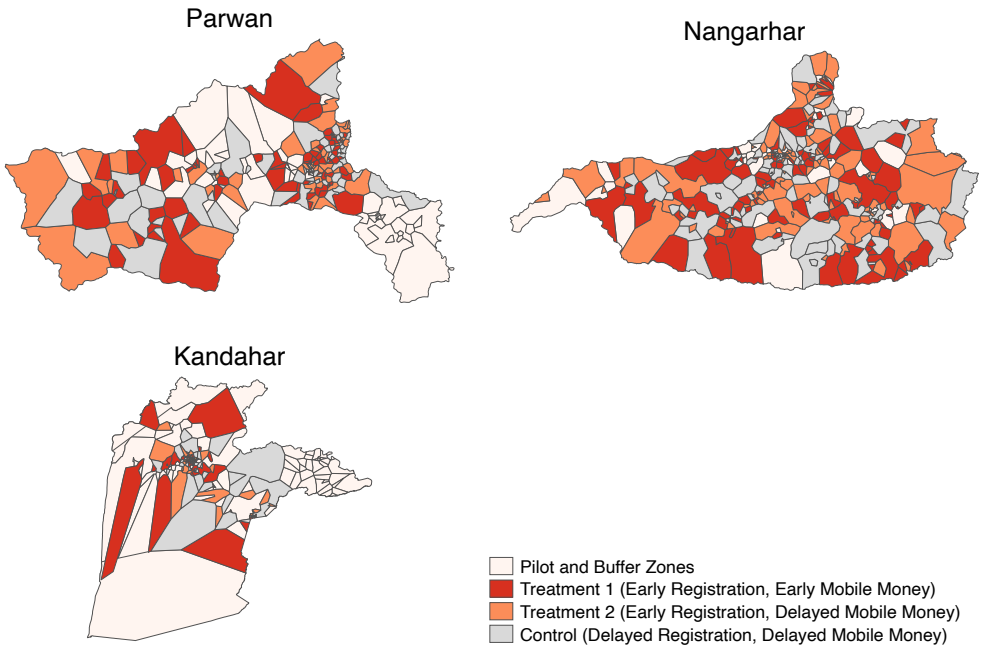


Figure 1: Evolution of Beliefs Over Conflict's Outcome

Notes: This figure plots the share of respondents to the question "Do you think the National Army will be able to defeat the Anti-Government Elements (Mukhalafeen-e dawlat) in the next few years?" across all respondents surveyed as part of the ANQAR survey. Those who answered "most likely defeat" and "certainly defeat" are grouped together in red, those who answered "certainly not defeat" and "most likely not defeat" are grouped together in blue, and those who answered "maybe defeat" are shown in black.



Panel A: Studied Provinces of Afghanistan



Panel B: Registration Zones and Treatment Status

Figure 2: Map of Provinces and Registration Zones

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

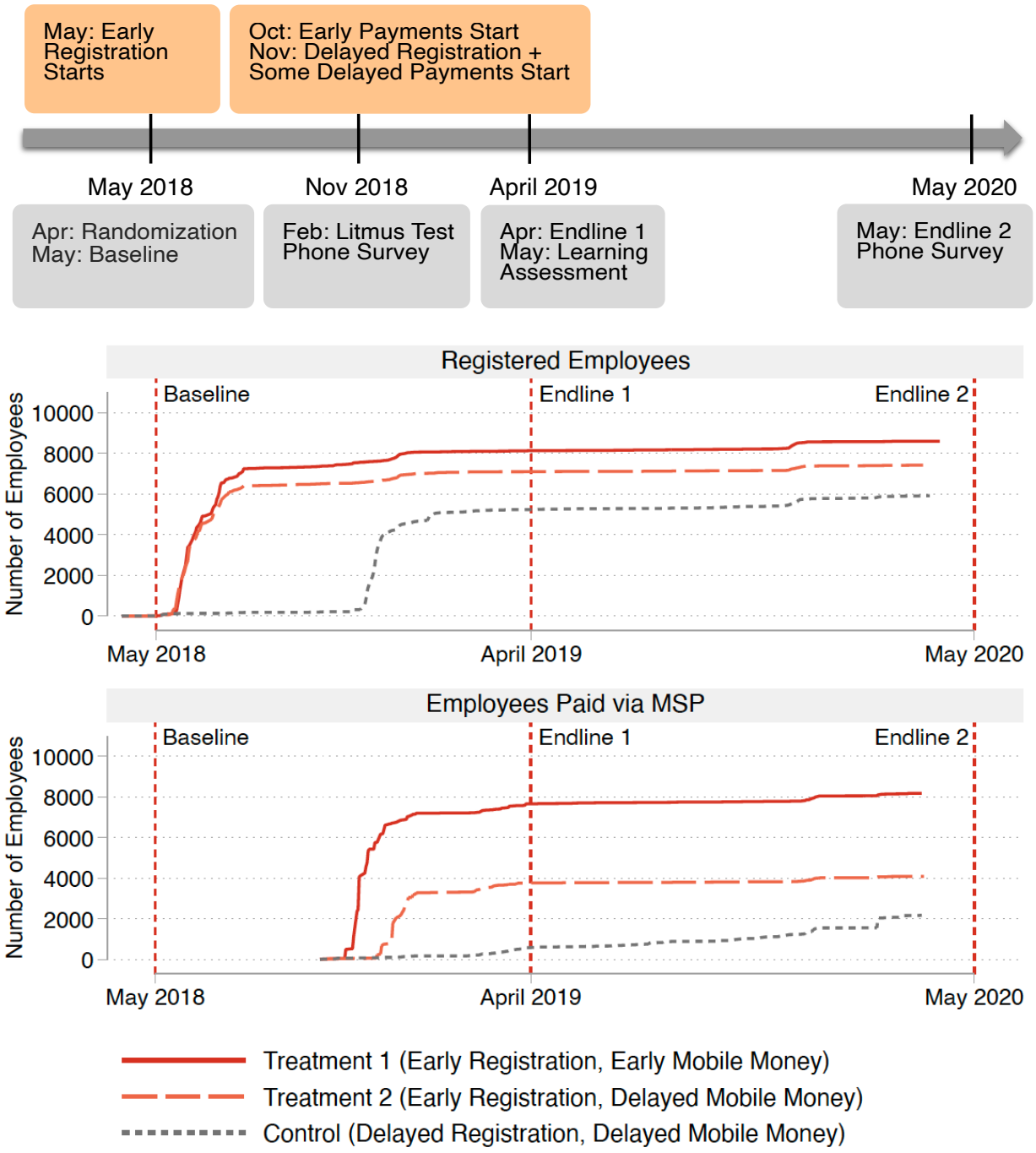
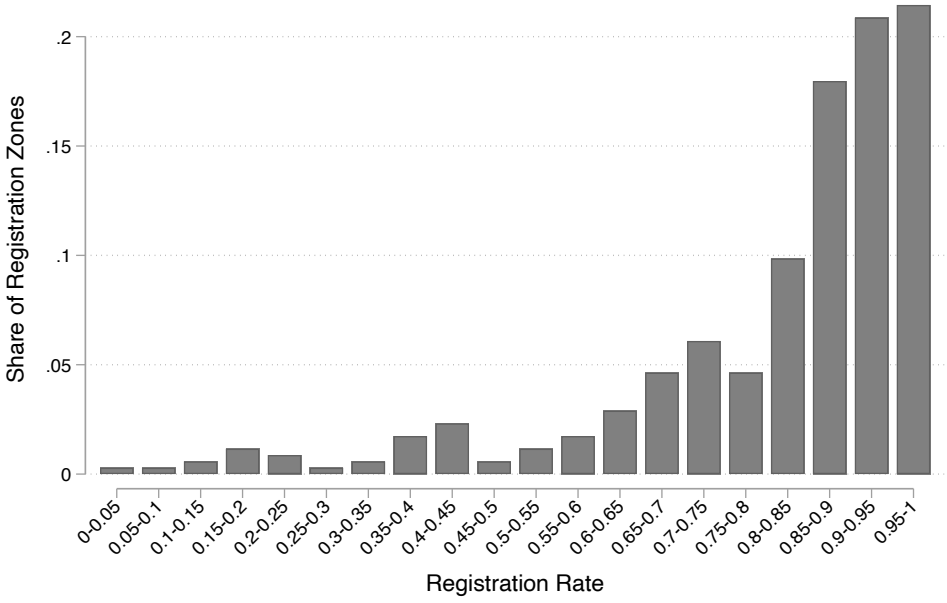
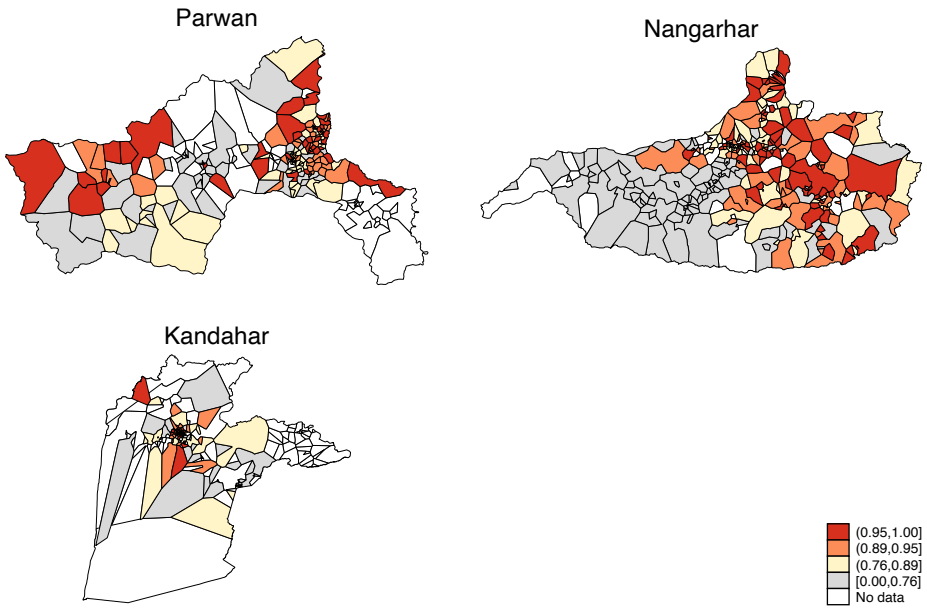


Figure 3: Project Implementation Timeline and Treatment Compliance

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.



Panel A: Histogram



Panel B: Map of Registration Levels by Zone

Figure 4: Registration During Drive Levels by Registration Zone

Notes: This figure shows the share of registration zones by registration rates during the drive (Panel A), and across the three provinces of the study (Panel B). We include in this figure zones identified by the mobile network operator as too unsafe for in-person registration drives (as some employees managed to register in these zones using alternative registration protocols).

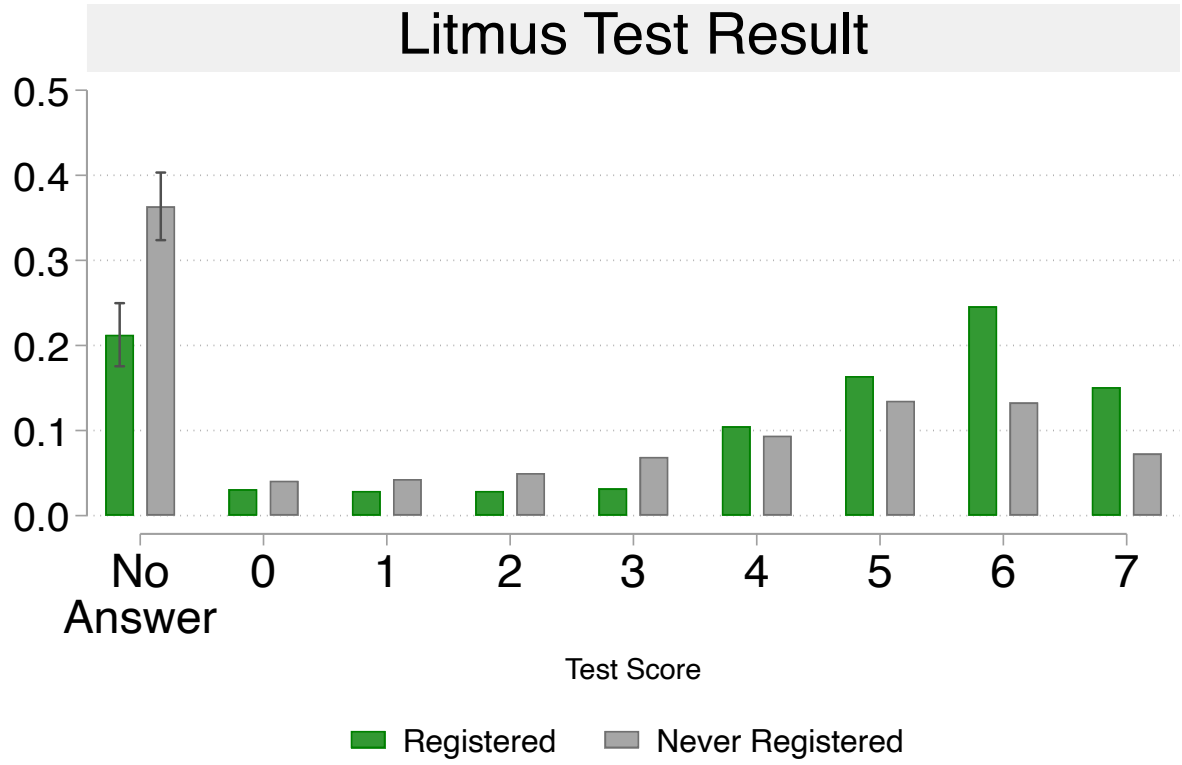


Figure 5: Litmus Test Results

Notes: This figure presents the results of the 'litmus test' survey designed to confirm whether employees knew basic facts about the schools where they were assigned. The litmus test score is defined as the total number of correct answers to the following seven questions: school district and name, employee's rank and position, principal's and headmaster's names, and the total number of employees. Employees were interviewed on the phone based on a phone number registry that pre-dated the experiment. No Answer corresponds to employees who did not answer the phone or were unavailable to respond to the survey.

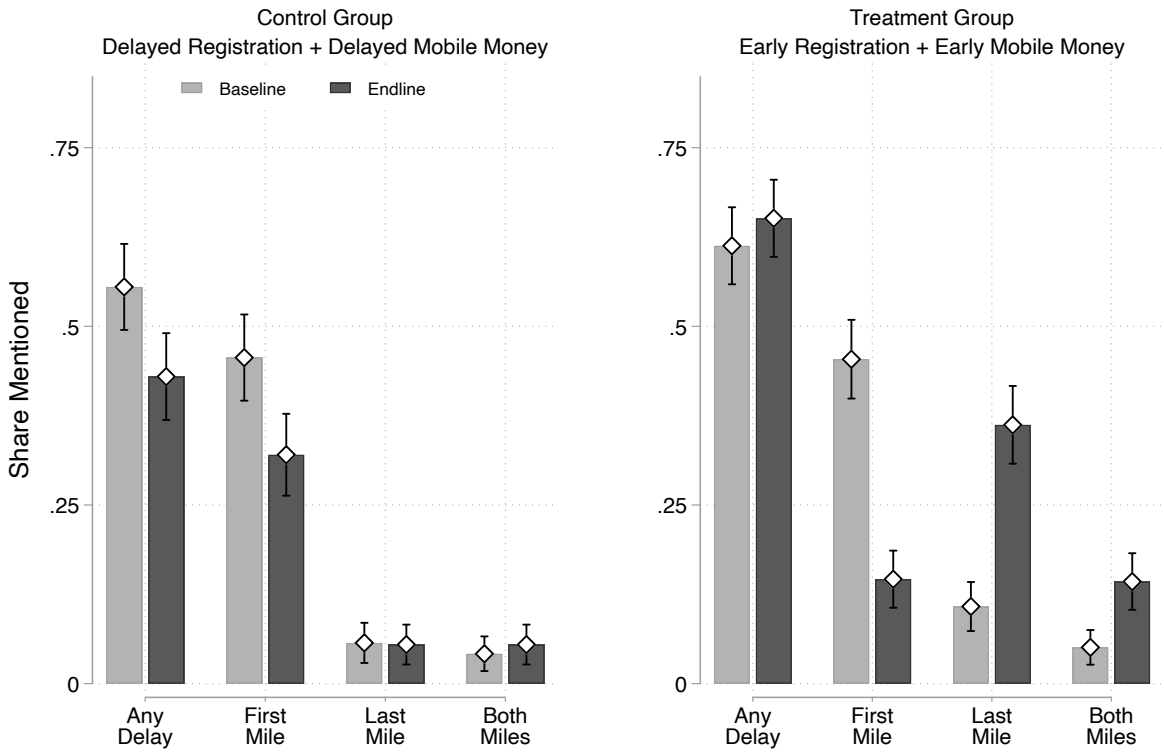


Figure 6: Separating First Mile and Last Mile Delays

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

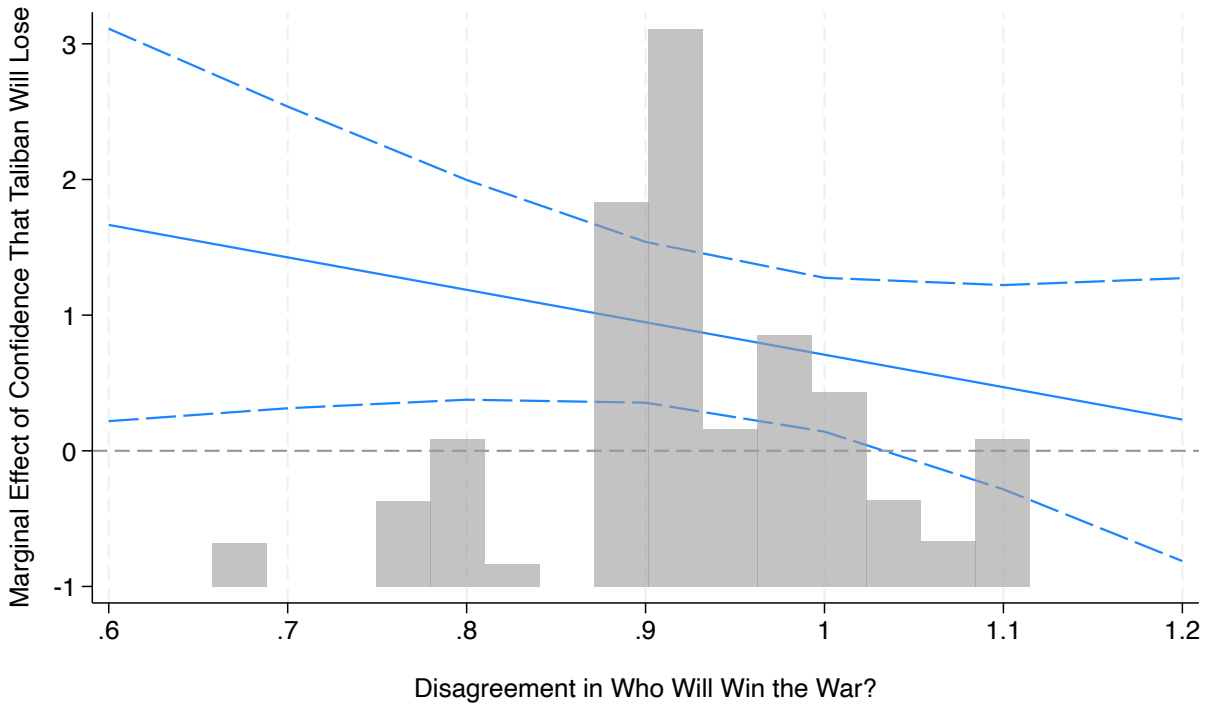


Figure 7: The Relationship Between Confidence and Registration by the Degree of Consensus

Notes: This figure depicts the relationship between the share of citizens in a district who believe the Taliban would lose the war in 2018 – the outset of the MSP reform – and the share of eligible teachers who appeared to give their biometric data in person (y-axis) against the degree of disagreement in who will win the war in the district (x-axis). These estimates are derived from Column 5 in Table 6.

Table 1: Predictors of Registration During On-site Drive

Dependent Variable	Registered During Drive (=1)			
	(1)	(2)	(3)	(4)
Early Reg., Early Mobile Money	0.203***			0.202***
	(0.038)			(0.036)
Late Reg., Early Mobile Money	0.206***			0.203***
	(0.037)			(0.038)
Based at Registration School (=1)		0.048**		0.046**
		(0.022)		(0.021)
Female		0.008		0.012
		(0.021)		(0.022)
Has At Least A Bachelor's		0.066**		0.062*
		(0.026)		(0.031)
Government is Effective (Overall, 2018, %)			–0.125	–0.218
			(0.135)	(0.131)
Pashtun (2018, %)			0.020	–0.002
			(0.039)	(0.033)
Contested District (=1)			–0.122*	–0.128**
			(0.069)	(0.061)
Urban (=1)			0.043	0.016
			(0.033)	(0.031)
On Primary Road (=1)			0.061	0.065*
			(0.041)	(0.033)
Police Capable of Protection (2018, %)			0.050	0.036
			(0.158)	(0.149)
Poor Cell Coverage (%)			0.096	0.073
			(0.076)	(0.062)
Constant	0.662***	0.765***	0.752***	0.669***
	(0.037)	(0.037)	(0.127)	(0.116)
Observations	19,626	19,626	19,626	19,626
R squared	0.055	0.010	0.026	0.087
# Districts	31	31	31	31
Mean Dependent Var	0.809	0.809	0.809	0.809

Notes: This table reports on the predictors of whether employees appeared for in-person registration for the new salary system. The dependent variable is a dummy variable equal to one for employees who registered during the registration drive. Standard errors clustered at the district level are reported in parentheses. Covariates in column 2 are measured at the individual level and in column 3 are measured at the district level except for poor cell coverage, which is measured at the school level. The sample includes all three treatment groups.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 2: Estimates of Impact on Employees' Payment Experience

	Payment to Receive Transfer (1)	Experienced Delay (2)	Travel Time to Cash-Out Salary (Min) (3)	Very Satisfied with the Pay System (4)	Vote in Favor of MSP (5)	Conducted Mobile Money Transaction (6)
Panel A. Year 1 Outcomes						
Mobile Salary Payment (=1)	–7.925	0.386***	39.839***	–0.371***	0.114*	0.272***
	(4.921)	(0.076)	(9.400)	(0.079)	(0.068)	(0.048)
Observations	838	848	810	845	829	844
R squared	0.249	0.251	0.284	0.209	0.343	0.161
Control Mean (Year 1)	19.094	0.441	31.219	0.598	0.461	0.067
Control Mean (Baseline)	23	0.596	37.449	0.269	0.980	
# Reg. Zones	310	310	303	310	308	310
Panel B. Year 2 Outcomes						
Mobile Salary Payment (=1)	–18.620***	–0.054	2.169	–0.015	0.251***	
	(3.827)	(0.083)	(5.934)	(0.080)	(0.083)	
Observations	644	642	632	643	623	
R squared	0.248	0.134	0.118	0.152	0.287	
Control Mean (Year 1)	10.178	0.284	36.629	0.761	0.710	
# Reg. Zones	286	287	283	287	282	

Notes: This table reports ToT estimates of impacts of mobile salary payments on payment quality and mobile money use. Corresponding ITT estimates are reported in Appendix Table A.21. Favor MSP is a dummy variable equal to one if the respondent indicates they would support scaling the reform across the Ministry of Education. Satisfied with Pay System is a dummy equal to one if the respondent indicates a very high level of satisfaction. Experienced Delay is a dummy variable equal to one if the respondent reports their salary being delayed. Travel Time is the time to convert the mobile money payment to cash in minutes. Payment to Receive measures what respondents report paying to receive their salary in Afghanis. For comparison, the average (net) salary in the sample is 8560 Afg. Money Transfer is a dummy variable equal to one if the respondent indicates making a mobile money transfer to someone else in the previous month. This outcome is not recorded in year 2 because the survey was abbreviated due to the pandemic. Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3: Estimates of Impacts on Payment Experience for Urban and Rural Employees

	Payment to Receive Transfer (1)	Experienced Delay (2)	Travel Time to Cash-Out Salary (Min) (3)	Very Satisfied with the Pay System (4)	Vote in Favor of MSP (5)	Conducted Mobile Money Transaction (6)
Panel A. Year 1 Outcomes						
β_1: MSP × Urban	–8.923	0.223**	32.790***	–0.259**	0.223**	0.304***
	(7.485)	(0.104)	(8.515)	(0.105)	(0.090)	(0.065)
β_2: MSP × Rural	–6.533	0.639***	48.990**	–0.548***	–0.064	0.223***
	(5.582)	(0.110)	(20.079)	(0.123)	(0.107)	(0.076)
Observations	838	848	810	845	829	844
R squared	0.007	0.072	0.063	0.022	0.013	0.062
Control Mean (Urban)	17.020	0.503	26.243	0.586	0.490	0.073
Control Mean (Rural)	22.136	0.350	38.175	0.618	0.415	0.058
p-value $\beta_1 = \beta_2$	0.798	0.006	0.458	0.074	0.041	0.416
# Reg. Zones	310	310	303	310	308	310
Panel B. Year 2 Outcomes						
β_1: MSP × Urban	–13.112***	–0.249*	–0.746	0.220*	0.224*	
	(4.976)	(0.129)	(7.451)	(0.115)	(0.129)	
β_2: MSP × Rural	–25.817***	0.195**	6.208	–0.320***	0.284***	
	(6.280)	(0.097)	(10.343)	(0.116)	(0.109)	
Observations	644	642	632	643	623	
R squared	0.067	–0.031	0.002	–0.048	0.100	
Control Mean (Urban)	4.914	0.345	37.035	0.733	0.776	
Control Mean (Rural)	17.716	0.198	36.038	0.802	0.620	
p-value $\beta_1 = \beta_2$	0.114	0.006	0.586	0.001	0.721	
# Reg. Zones	286	287	283	287	282	

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

Table 4: Estimates of Impact on Student Learning

	Combined Score			Math Score			Reading Score		
	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. 2SLS Estimates									
Treated School (=1)	0.147*	0.315***	–0.093	0.138	0.335***	–0.142	0.156*	0.295**	–0.044
	(0.085)	(0.110)	(0.121)	(0.094)	(0.116)	(0.141)	(0.095)	(0.122)	(0.140)
R^2	0.326	0.300	0.378	0.327	0.287	0.397	0.262	0.250	0.290
Panel B. Reduced Form									
EE (=1)	0.135*	0.279***	–0.097	0.127	0.287**	–0.155	0.143	0.271**	–0.040
	(0.081)	(0.105)	(0.118)	(0.088)	(0.111)	(0.136)	(0.090)	(0.117)	(0.136)
ED (=1)	0.128	0.264**	–0.094	0.141	0.331***	–0.169	0.114	0.197*	–0.018
	(0.081)	(0.104)	(0.124)	(0.090)	(0.110)	(0.144)	(0.087)	(0.112)	(0.133)
R^2	0.333	0.321	0.376	0.333	0.312	0.396	0.268	0.265	0.289
Untreated School Mean	0.050	0.040	0.040	0.050	0.020	0.090	0.040	0.050	0
p-value EE = ED = 0	0.922	0.887	0.972	0.873	0.709	0.915	0.720	0.505	0.856
FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Reg. Zones	318	189	129	318	189	129	318	189	129
# Students	822	483	339	822	483	339	822	483	339

Notes: This table reports the impacts of the Mobile Salary Payment Reform on student learning outcomes measured in the assessment described in Section 3.5. Scores are standardized using the control group mean and standard deviation. The sample comprises students aged between 6 and 10 years who attended a public school eligible for the Mobile Salary Payments reform within the last three months. Treated School is a dummy variable equal to one for students who attend a school where at least 50% of teachers were paid using mobile money by April 2019, the month before the learning assessment. DD is the Delayed Registration, Delayed Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and ED is the Early Registration, Delayed Payment (Treatment 2) group. Covariates include years of education, years of age, a dummy variable equal to one if a student lives in a high socio-economic status household, and gender. High SES households are those that are on a maintained road with access to water and to electricity. Standard errors clustered at the registration zone level are reported in parentheses. For each column, the sample is the same across panels.
Levels of significance: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: Impacts on Mobile Money Use from Administrative Data

	Deposit (1)	Transfer to Cust. Wallet (2)	Own Airtime Top-up (3)	Other Airtime Top-up (4)	Balance (5)
Panel A. OLS					
MSP Months	0.004***	0.003**	0.001***	0.002***	0.012***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.002)
1-Month Mean	0.002	0.000	0.000	0.000	0.137
Observations	14,753	14,753	14,753	14,753	14,753
R^2	0.005	0.015	0.022	0.042	0.126
# Reg. Zones	284	284	284	284	284
Panel B. 2SLS					
MSP Months	0.004***	0.004	0.001***	0.002***	0.005*
	(0.001)	(0.003)	(0.000)	(0.000)	(0.003)
Observations	14,753	14,753	14,753	14,753	14,753
R^2	0.002	0.001	0.011	0.017	0.005
# Reg. Zones	284	284	284	284	284

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

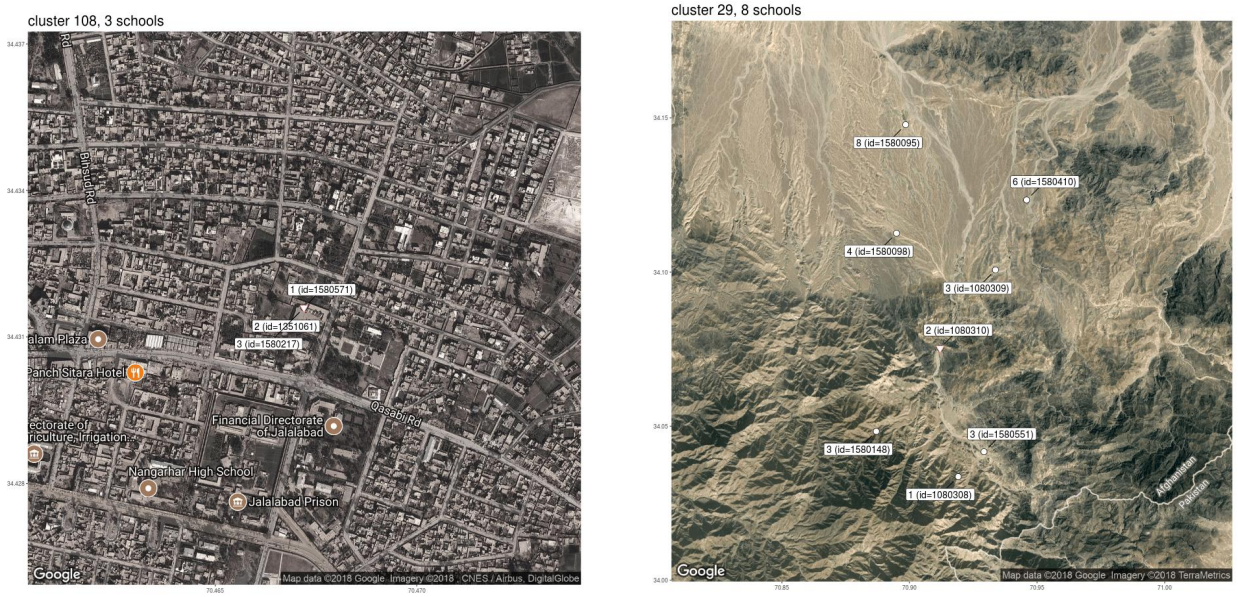
Table 6: Beliefs About State Durability and Registration During On-site Drive

Dependent Variable	Registered During Drive (=1)				
	(1)	(2)	(3)	(4)	(5)
Taliban Will Lose at Baseline (2018, %)	0.994***		0.918**	6.488***	6.868***
	(0.343)		(0.402)	(2.212)	(2.179)
SD of Beliefs				5.397**	5.826***
				(2.181)	(2.010)
Levels of Beliefs × SD of Beliefs				–5.510**	–6.207***
				(2.224)	(2.135)
Taliban Will Lose Pre-Reform (2017, %)		0.413	–0.044		–0.041
		(0.300)	(0.290)		(0.265)
Taliban Will Lose Pre-Reform (2016, %)		0.036	–0.579*		–0.610**
		(0.348)	(0.326)		(0.241)
Taliban Will Lose Pre-Reform (2015, %)		0.444	0.463		0.491*
		(0.366)	(0.299)		(0.268)
Police Capable of Protection (2018, %)			0.048		0.056
			(0.142)		(0.126)
Government is Effective (Overall, 2018, %)			–0.389***		–0.418***
			(0.129)		(0.127)
Pashtun (2018, %)			–0.014		–0.022
			(0.039)		(0.040)
On Primary Road (=1)			0.081**		0.055**
			(0.030)		(0.024)
Urban (=1)			0.018		0.052
			(0.029)		(0.036)
Contested District (=1)			–0.122**		–0.098**
			(0.045)		(0.043)
Bad Cell Coverage (%)			–0.001		0.023
			(0.067)		(0.076)
Based at Registration School (=1)			0.046**		0.048**
			(0.021)		(0.021)
Female			0.011		0.016
			(0.022)		(0.021)
Has At Least A Bachelor's			0.059*		0.058*
			(0.031)		(0.031)
Early Reg., Early Mobile Money			0.198***		0.196***
			(0.035)		(0.035)
Late Reg., Early Mobile Money			0.195***		0.193***
			(0.036)		(0.036)
Constant	–0.111	0.009	0.099	–5.486**	–5.493***
	(0.328)	(0.349)	(0.308)	(2.191)	(1.934)
Observations	19,626	19,626	19,626	19,626	19,626
R squared	0.013	0.011	0.095	0.020	0.100
# Districts	31	31	31	31	31
WC Bootstrap p-value	0.031		0.128	0.028	0.049
Mean Dependent Var	0.809	0.809	0.809	0.809	0.809

Notes: This table reports on the relationship between citizen's beliefs about who will win the war and the likelihood that employees appeared for in-person registration for the new salary system. The dependent variable is a dummy variable equal to one for employees who registered during the registration drive. Data on beliefs are from ANQAR survey. Data for a given year include all the survey waves in said year. Baseline measures are from the 2018 survey rounds and measures pre-reform are from the 2015-2017 rounds. The sample includes teachers from all three treatment arms who were paid at least once before the reform began, but excludes unsafe schools and schools where less than 10% of employees paid in the 12 months prior to the reform registered. Standard errors clustered at the district level are reported in parentheses. Wild Cluster (WC) Bootstrap p-value calculated for Tabliban Will Lose at Baseline (2018, %) coefficient.
Levels of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Appendix - For Online Publication

A Appendix Figures



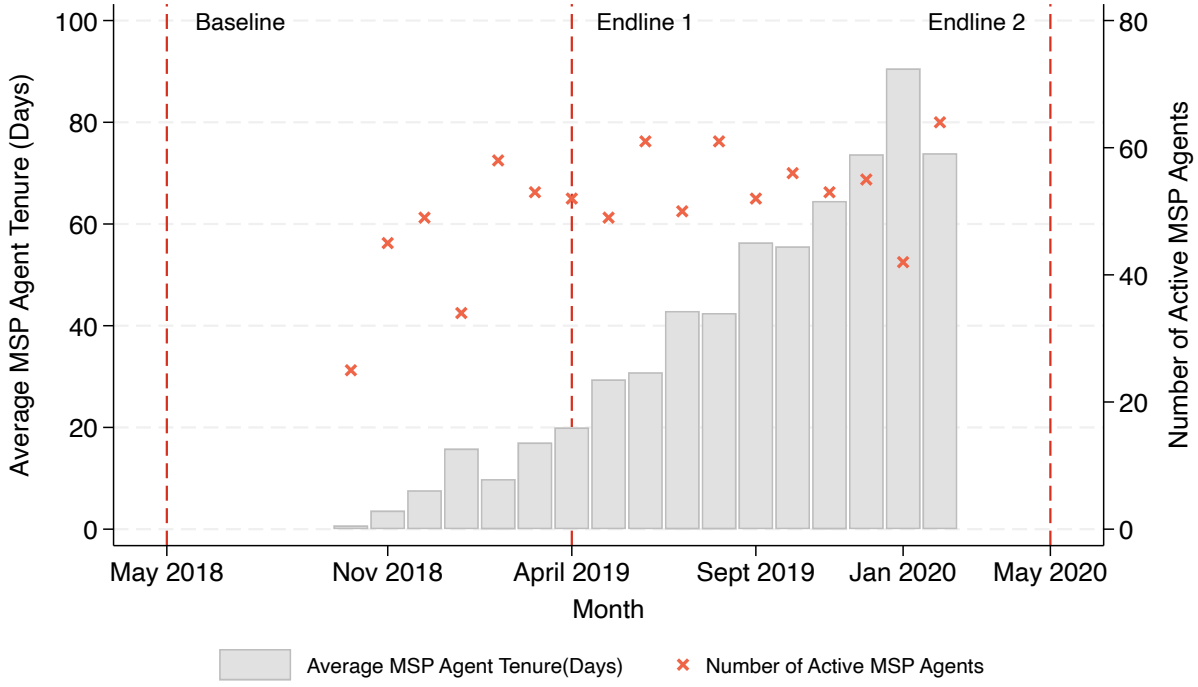
Panel A: Urban Registration Zone

Panel B: Rural Registration Zone

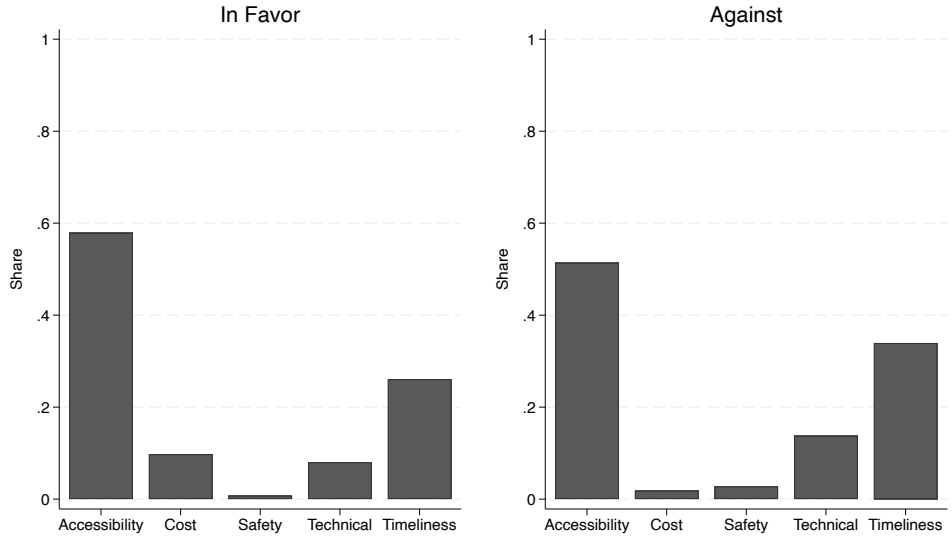
Figure A.1: Example of Two Registration Zones in Nangarhar

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

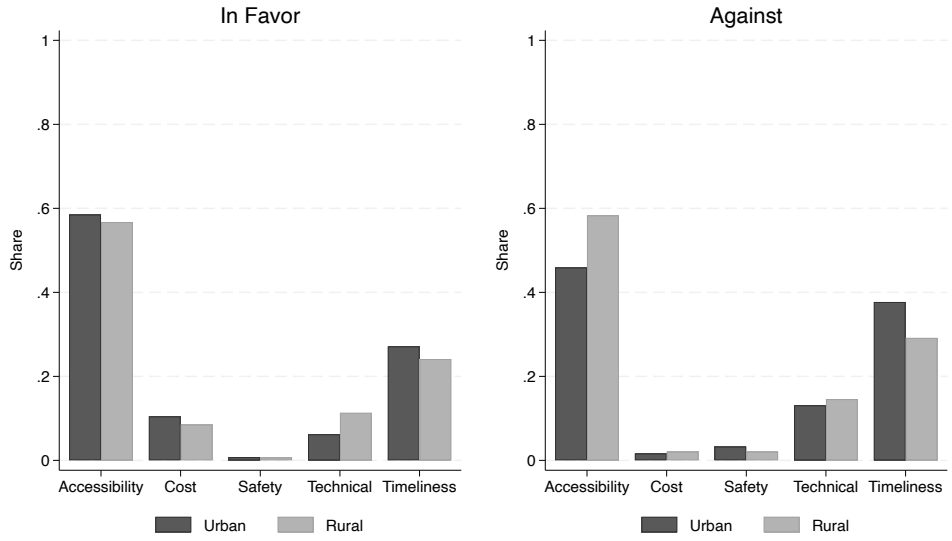
Figure A.2: Tenure of MSP Agents



Notes: This figure shows the average tenure (in days) and the number of active MSP agents between October 2018 and February 2020. The MSP agent is active on a particular day (month) if the agent has at least one 'Customer to Merchant Withdrawal - cashout' transaction on that day (month) in the transaction data. The agent tenure is the number of days the agent was active in a particular month. Vertical lines mark surveys.



Panel A: Full Sample



Panel B: Urban-Rural Split Samples

Figure A.3: Reasons for MSP Reform Support/Opposition

Notes: This figure shows self-reported reasons for voting in favor or against scaling the MSP reform in Year 2. Open-ended responses were hand-coded into one or more of five categories: Accessibility (e.g. access to mobile money agents), Cost (e.g. withdrawal fees and leakage), Safety (e.g. security of withdrawal location), Technical (e.g. SIM or network functionality), and Timeliness (e.g. payment delays). Panel A shows full sample of coded responses, split by In Favor (n=399) and Against (n=109). Panel B further divides the sample into four groups: In Favor - Urban (n=258), In Favor - Rural (n=141), Against - Urban (n=61), and Against - Rural (n=48).

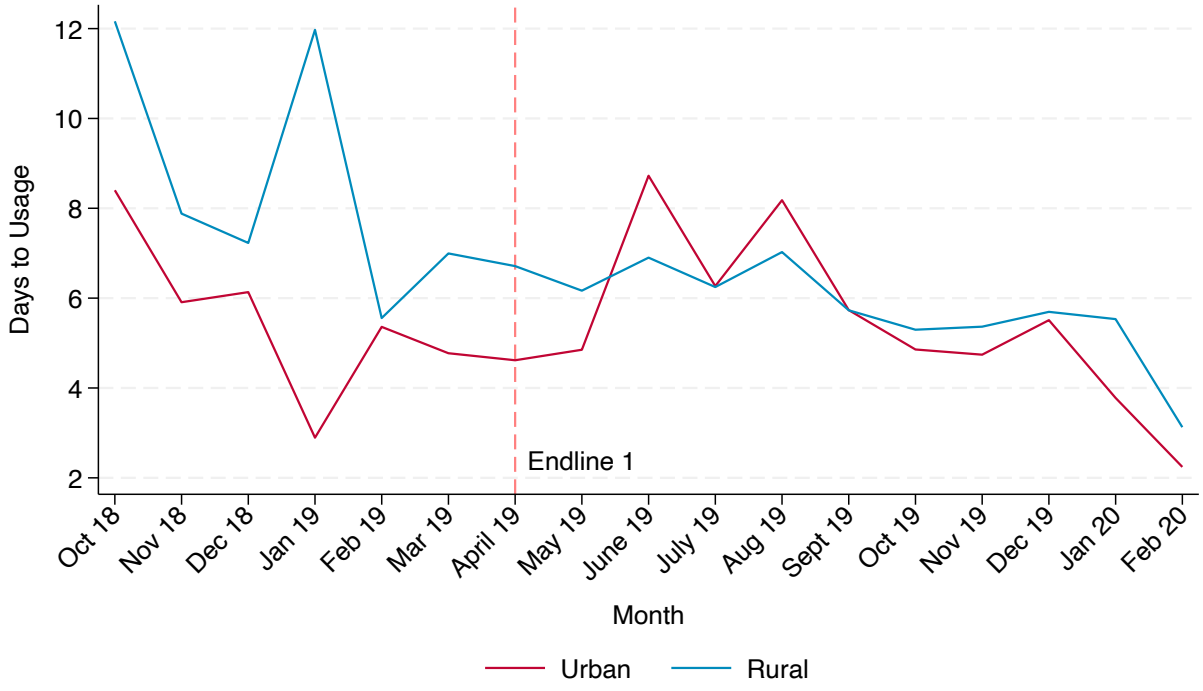
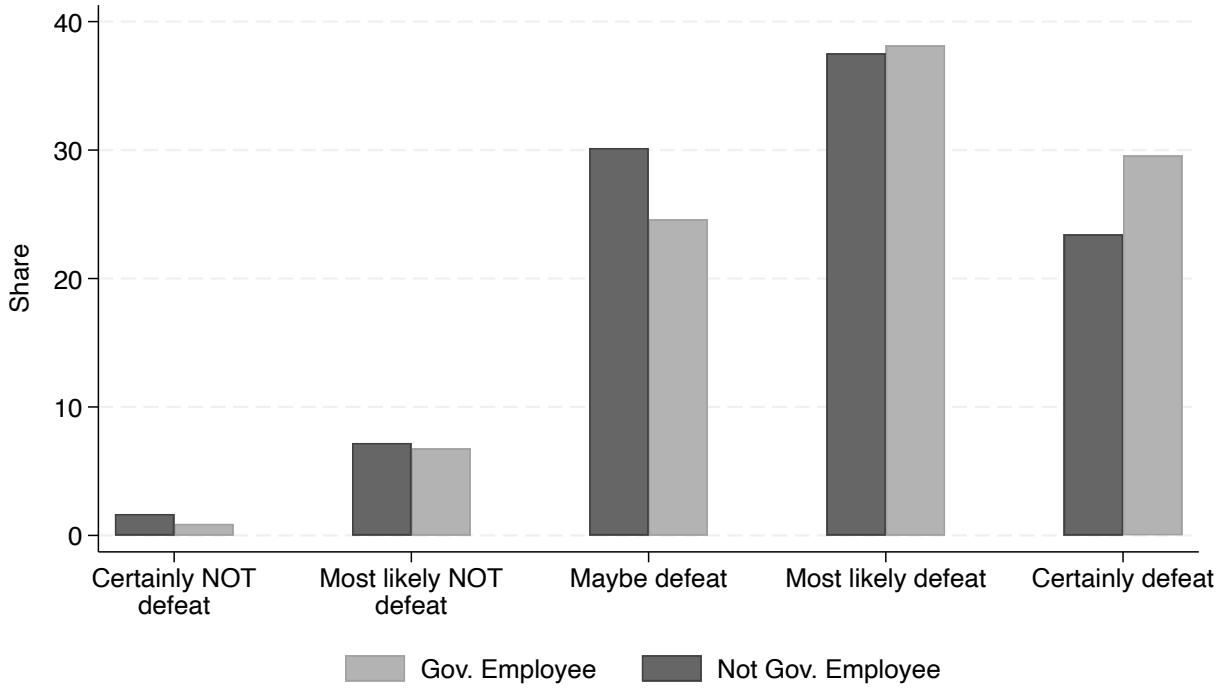


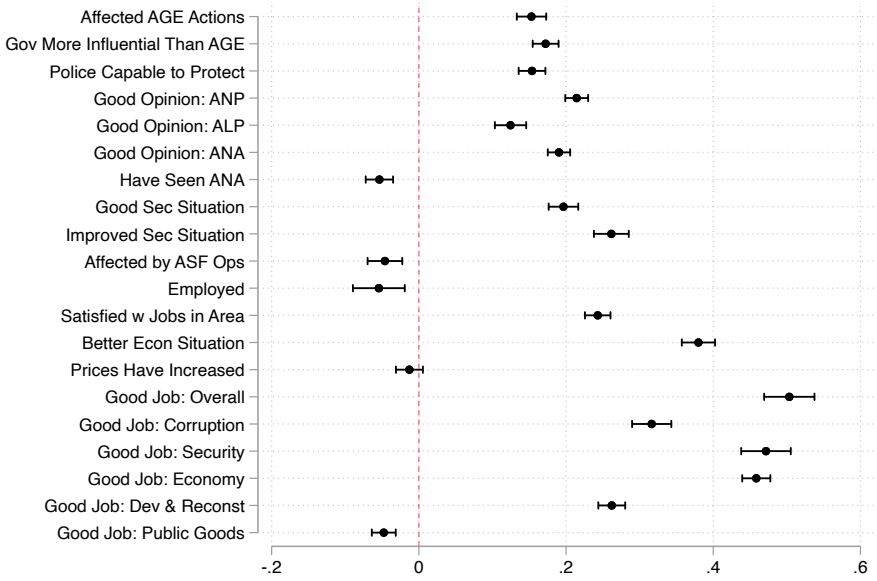
Figure A.4: Days to Mobile Account Usage

Notes: Average days between mobile salary payment and first cash withdrawal or peer-to-peer transfer, calculated separately for urban/rural samples. Vertical line marks date of endline 1 surveys.

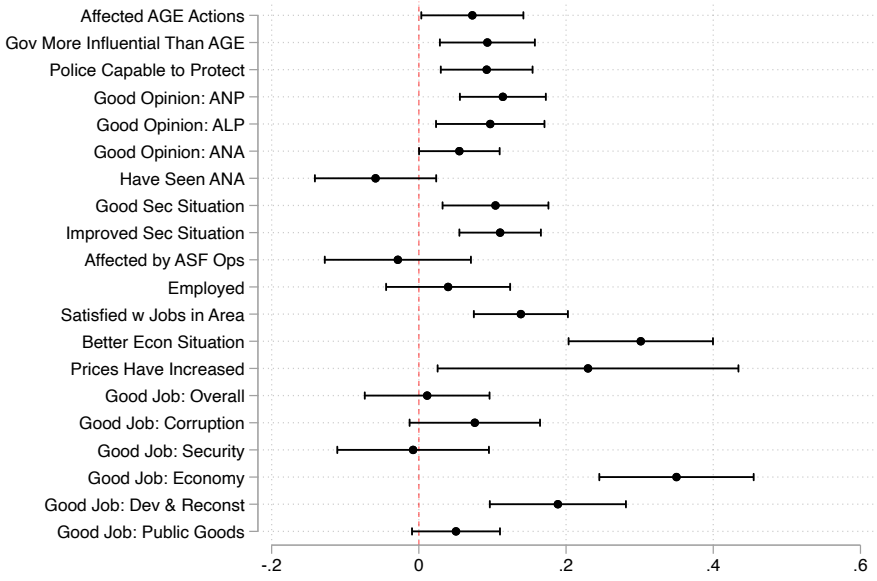
Figure A.5: Beliefs of Taliban Defeat by Occupation



Notes: The Figures uses data from the four waves of the ANQAR survey in 2018. The sample is restricted to those districts with individuals in the EE and ED groups, for which the control variables in Table 6 are not missing. It shows, for those individuals working for the government and for those working for a different entity, the share of respondents who selected each of the five possible answers to the question "Do you think the National Army will be able to defeat the Anti-Government Elements (Mukhalafeen-e dawlat) in the next few years?", excluding those who refused to answer or answered that they did not know.



Panel A: All Waves, All Districts



Panel B: 2015-2018, Intervention's Outcome

Figure A.6: Predicting Beliefs over Conflict's Outcomes

Notes: This figure shows coefficients of regressions of the share of respondents stating that it is likely that the National Army will defeat the Taliban in the coming years at the district level on survey wave fixed effects, district fixed effects, and a given potential predictor. Each row corresponds to a different regression, using one potential predictor at the time. Standard errors are robust, and 95% confidence intervals are also shown. Panel A shows the results using all survey waves and all districts, Panel B shows the results using only the waves in years 2015-2018, in districts in which the intervention took place.

B Appendix Tables

Table A.1: Balance Checks for Survey Outcomes

	DD Mean (1)	EE Effect (2)	ED Effect (3)	N Obs (4)
Panel A. Without Stratum FEs				
Payment to Receive Salary (Afg)	24.469	0.306	–3.587	873
	(5.112)	(6.252)	(6.016)	
		[0.987]	[0.814]	
Experienced Delay	0.592	0.040	–0.069	883
	(0.043)	(0.057)	(0.060)	
		[0.781]	[0.394]	
Travel Time to Cash-Out (Min)	31.488	–0.777	–3.645	822
	(3.443)	(4.525)	(4.076)	
		[0.987]	[0.594]	
Very Satisfied with Pay System	0.286	0.010	–0.014	881
	(0.038)	(0.049)	(0.051)	
		[0.987]	[0.984]	
Vote in Favor of MSP	0.938	–0.035	–0.033	869
	(0.022)	(0.033)	(0.031)	
		[0.448]	[0.448]	
Panel B. With Stratum FEs				
Payment to Receive Salary (Afg)	24.469	–1.450	–6.397	872
	(5.112)	(4.271)	(4.818)	
		[0.901]	[0.263]	
Experienced Delay	0.592	0.062	–0.022	882
	(0.043)	(0.045)	(0.045)	
		[0.259]	[0.889]	
Travel Time to Cash-Out (Min)	31.488	–3.074	–4.760	821
	(3.443)	(3.826)	(3.420)	
		[0.688]	[0.259]	
Very Satisfied with Pay System	0.286	0.002	–0.018	880
	(0.038)	(0.047)	(0.049)	
		[0.962]	[0.901]	
Vote in Favor of MSP	0.938	–0.047	–0.041	868
	(0.022)	(0.031)	(0.027)	
		[0.201]	[0.200]	

Notes: This table checks balance on the main outcomes reflecting employees' salary experience. DD is the Delayed Registration, Delayed Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and ED is the Early Registration, Delayed Payment (Treatment 2) group. Robust standard errors are clustered at the registration zone (treatment unit) level. FWER-adjusted p-values within each panel are reported in squared brackets (following Romano & Wolf, 2005, using 1000 repetitions).

Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

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

	DD Mean (1)	EE Effect (2)	ED Effect (3)	EE - ED Difference (4)	F-stat (5)	N Obs. (6)
Panel A. Endline 1 – April 2019						
Share Registered	0.85	0.13***	0.12***	0.00	6.03	849
	[0.03]	[0.04]	[0.04]	[0.02]		
Share Self-Reported MSP	0.1	0.68***	0.4***	0.29***	106.1	849
	[0.03]	[0.05]	[0.05]	[0.05]		
Panel B. Endline 2 – May 2020						
Share Registered	0.85	0.15***	0.11***	0.03	8.15	645
	[0.04]	[0.04]	[0.04]	[0.02]		
Share Self-Reported MSP	0.41	0.5***	0.14***	0.35***	60.22	645
	[0.05]	[0.05]	[0.04]	[0.05]		

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

Table A.3: Representativity of Survey Samples and Balance on Demographics

	Overall Mean (1)	Sample Mean (2)	DD Mean (3)	EE Effect (4)	ED Effect (5)	p-value DD=EE=ED (6)	N Obs. (7)
Panel A. Baseline – May 2018							
Male	0.86	0.85	0.85	–0.03	0	0.97	887
	(0.34)	(0.36)	[0.03]	[0.04]	[0.04]		
Age	40.29	40.23	40.75	–0.58	–1.51	0.51	887
	(13.72)	(12.88)	[0.94]	[1.08]	[1.04]		
College Graduate	0.22	0.32	0.29	–0.01	0.05	0.19	887
	(0.42)	(0.47)	[0.03]	[0.04]	[0.04]		
Pashtun	0.68	0.59	0.56	–0.03	–0.04**	0.61	731
	(0.47)	(0.49)	[0.05]	[0.03]	[0.02]		
Tajik	0.28	0.35	0.38	0.04	0.05**	0.71	731
	(0.45)	(0.48)	[0.05]	[0.03]	[0.02]		
Other Ethnicity	0.04	0.05	0.06	–0.02	–0.01	0.87	731
	(0.2)	(0.22)	[0.02]	[0.02]	[0.02]		
Employee Total Salary Payment	9279.8	9522.22	8608.59	414.16	1534.1	0.18	230
	(7618.2)	(6175.06)	[373.15]	[645.84]	[1103.98]		
Panel B. Endline 1 – April 2019							
Male		0.83	0.83	–0.04	0.01	0.87	849
		(0.38)	[0.04]	[0.05]	[0.05]		
Age		38.57	37.71	0.91	1.46	0.56	848
		(12.68)	[0.96]	[1.09]	[1.05]		
College Graduate		0.31	0.32	–0.04	–0.05	0.85	849
		(0.46)	[0.04]	[0.05]	[0.04]		
Pashtun		0.58	0.58	–0.04**	–0.04**	0.94	705
		(0.49)	[0.05]	[0.02]	[0.02]		
Tajik		0.36	0.38	0.02	0.04**	0.93	705
		(0.48)	[0.05]	[0.03]	[0.02]		
Other Ethnicity		0.06	0.04	0.03	0	0.48	705
		(0.23)	[0.02]	[0.02]	[0.02]		
Employee Total Salary Payment		9604.66	8859.73	743.71	1036.56	0.17	136
		(3022.29)	[390.69]	[738.25]	[823.24]		
Panel C. Endline 2 – May 2020							
Male		0.87	0.88	–0.05	0	0.93	645
		(0.34)	[0.03]	[0.05]	[0.05]		
Age		38.09	37.3	0.33	0.75	0.69	645
		(12.89)	[1.19]	[1.29]	[1.21]		
College Graduate		0.33	0.34	–0.05	–0.03	0.88	645
		(0.47)	[0.04]	[0.05]	[0.05]		
Pashtun		0.67	0.68	–0.04**	–0.02	0.83	531
		(0.47)	[0.05]	[0.02]	[0.03]		
Tajik		0.29	0.29	0	0.03	1	531
		(0.45)	[0.05]	[0.03]	[0.02]		
Other Ethnicity		0.04	0.03	0.04**	0	0.34	531
		(0.2)	[0.01]	[0.02]	[0.02]		
Employee Total Salary Payment		9621.37	8583.36	1196.79*	1438.58*	0.05*	102
		(3007.04)	[322.11]	[718.35]	[830.58]		

Notes: This table shows the balance of demographic and income measures for respondents active at baseline, endline 1 and endline 2, controlling for the district (strata) fixed effects. The employee is present if the survey team found him at school during the unannounced visit. DD is the Delayed Registration, Delayed Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and ED is the Early Registration, Delayed Payment (Treatment 2) group. Column 1 shows the overall mean among all teachers who were paid at least once prior to the start of the reform (a total of 24768 teachers). Standard deviations in parentheses and robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.

Table A.4: Balance of Nonresponse Rates for Main Outcomes

	Sample Mean (1)	DD Mean (2)	EE Mean (3)	ED Mean (4)	p-value D=EE=ED (5)	N Obs. (6)
Panel A. Baseline – May 2018						
Payment to Receive Salary (Afg)	0.02	0.02	0.01	0.02	0.89	887
	(0.12)	[0.01]	[0.01]	[0.01]		
Share Experienced Salary Delays	0.00	0.00	0.00	0.01	0.25	887
	(0.07)	[0.00]	[0.00]	[0.01]		
Travel Time to Cash-Out (Min)	0.07	0.06	0.09	0.07	0.67	887
	(0.26)	[0.02]	[0.03]	[0.02]		
Share Very Satisfied with Pay System	0.01	0.00	0.01	0.00	0.52	887
	(0.08)	[0.00]	[0.01]	[0.00]		
Share Voting in Favor of MSP	0.02	0.02	0.02	0.02	0.88	887
	(0.14)	[0.01]	[0.01]	[0.01]		
Panel B. Endline 1 – April 2019						
Payment to Receive Salary (Afg)	0.01	0.01	0.02	0.01	0.6	849
	(0.11)	[0.01]	[0.01]	[0.01]		
Share Experienced Salary Delays	0.00	0.00	0.00	0.00	0.32	849
	(0.03)	[0.00]	[0.00]	[0.00]		
Travel Time to Cash-Out (Min)	0.05	0.04	0.04	0.06	0.58	849
	(0.21)	[0.02]	[0.02]	[0.02]		
Share Very Satisfied with Pay System	0.00	0.01	0.00	0.00	0.76	849
	(0.07)	[0.01]	[0.00]	[0.00]		
Share Voting in Favor of MSP	0.02	0.04	0.01	0.02	0.24	849
	(0.15)	[0.02]	[0.01]	[0.01]		
Panel C. Endline 2 – May 2020						
Payment to Receive Salary (Afg)	0.00	0.00	0.00	0.00	0.32	645
	(0.04)	[0.00]	[0.00]	[0.00]		
Share Experienced Salary Delays	0.00	0.00	0.00	0.01	0.08*	645
	(0.07)	[0.00]	[0.00]	[0.01]		
Travel Time to Cash-Out (Min)	0.02	0.02	0.01	0.03	0.45	645
	(0.14)	[0.01]	[0.01]	[0.01]		
Share Very Satisfied with Pay System	0.00	0.00	0.00	0.01	0.16	645
	(0.06)	[0.00]	[0.00]	[0.01]		
Share Voting in Favor of MSP	0.03	0.06	0.02	0.03	0.17	645
	(0.18)	[0.02]	[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. DD is the Delayed Registration, Delayed Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and ED is the Early Registration, Delayed Payment (Treatment 2) group. Standard errors in parentheses and robust standard errors clustered at the registration zone (treatment unit) level in squared brackets.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.5: Predictors of Registration At Any Time

Dependent Variable	Registered at Any Time (=1)			
	(1)	(2)	(3)	(4)
Early Reg., Early Mobile Money	0.201***			0.201***
	(0.035)			(0.036)
	[0.000]			[0.000]
Late Reg., Early Mobile Money	0.207***			0.205***
	(0.034)			(0.036)
	[0.000]			[0.000]
Based at Registration School (=1)		0.041**		0.038*
		(0.018)		(0.020)
		[0.082]		[0.099]
Female		–0.004		–0.001
		(0.014)		(0.012)
		[0.809]		[0.946]
Has At Least A Bachelor's		0.059**		0.058**
		(0.022)		(0.026)
		[0.003]		[0.010]
Government is Effective (Overall, 2018, %)			–0.051	–0.124
			(0.099)	(0.097)
			[0.657]	[0.318]
Pashtun (2018, %)			–0.006	–0.029
			(0.032)	(0.026)
			[0.869]	[0.394]
Contested District (=1)			–0.043	–0.050
			(0.043)	(0.034)
			[0.368]	[0.229]
Urban (=1)			0.034	0.010
			(0.027)	(0.025)
			[0.258]	[0.734]
On Primary Road (=1)			0.048*	0.052**
			(0.026)	(0.020)
			[0.292]	[0.135]
Police Capable of Protection (2018, %)			0.078	0.059
			(0.137)	(0.127)
			[0.633]	[0.689]
Poor Cell Coverage (%)			0.044	0.017
			(0.060)	(0.045)
			[0.550]	[0.704]
Constant	0.702***	0.812***	0.771***	0.687***
	(0.030)	(0.027)	(0.082)	(0.079)
Observations	19,626	19,626	19,626	19,626
R squared	0.065	0.009	0.012	0.084
# Districts	31	31	31	31
Mean Dependent Var	0.848	0.848	0.848	0.848

Notes: This table reports on the predictors of whether employees appeared for in-person registration for the new salary system. The dependent variable is a dummy variable equal to one for employees who registered during the registration drive. Standard errors clustered at the district level are reported in parentheses. Covariates in column 2 are measured at the individual level and in column 3 are measured at the district level except for , which is measured at the school level. The sample includes all three treatment groups, but excludes unsafe schools and schools in registration zones where less than 10% of teachers registered by April 2019.

Levels of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table A.6: Predictors of Registration During On-site Drive

Dependent Variable	Registered During Drive (=1)			
	(1)	(2)	(3)	(4)
Early Reg., Early Mobile Money	0.203***			0.202***
	(0.038)			(0.036)
	[.001]			[0]
Late Reg., Early Mobile Money	0.206***			0.203***
	(0.037)			(0.038)
	[.001]			[0]
Based at Registration School (=1)		0.048**		0.046**
		(0.022)		(0.021)
		[.053]		[.069]
Female		0.008		0.012
		(0.021)		(0.022)
		[.696]		[.615]
Has At Least A Bachelor's		0.066**		0.062*
		(0.026)		(0.031)
		[.013]		[.023]
Government is Effective (Overall, 2018, %)			–0.125	–0.218
			(0.135)	(0.131)
			[.414]	[.143]
Pashtun (2018, %)			0.020	–0.002
			(0.039)	(0.033)
			[.66]	[.965]
Contested District (=1)			–0.122*	–0.128**
			(0.069)	(0.061)
			[.253]	[.199]
Urban (=1)			0.043	0.016
			(0.033)	(0.031)
			[.238]	[.637]
On Primary Road (=1)			0.061	0.065*
			(0.041)	(0.033)
			[.454]	[.283]
Police Capable of Protection (2018, %)			0.050	0.036
			(0.158)	(0.149)
			[.768]	[.821]
Poor Cell Coverage (%)			0.096	0.073
			(0.076)	(0.062)
			[.305]	[.331]
Constant	0.662***	0.765***	0.752***	0.669***
	(0.037)	(0.037)	(0.127)	(0.116)
Observations	19,626	19,626	19,626	19,626
R squared	0.055	0.010	0.026	0.087
# Districts	31	31	31	31
Mean Dependent Var	0.809	0.809	0.809	0.809

Notes: This table reports on the predictors of whether employees appeared for in-person registration for the new salary system. The dependent variable is a dummy variable equal to one for employees who registered during the registration drive. Standard errors clustered at the district level are reported in parentheses. Covariates in column 2 are measured at the individual level and in column 3 are measured at the district level except for poor cell coverage, which is measured at the school level. The sample includes all three treatment groups.
Levels of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table A.7: Registration Effects on the Main Outcomes

	Unregistered Mean (1)	Registration Effect (2)	N Obs. (3)
Panel A. Endline 1 – April 2019			
Payment to Receive Salary (Afg)	0.47	0.02	219
	[0.11]	[0.1]	
Experienced Delay	32.12	–7.03	222
	[9.19]	[7.72]	
Travel Time to Cash-Out (Min)	0.56	0.11	228
	[0.1]	[0.11]	
Very Satisfied with Pay System	0.51	–0.04	230
	[0.11]	[0.07]	
Vote in Favor of MSP	30.29	–5.93	228
	[9.92]	[12.66]	
Panel B. Endline 2 – May 2020			
Payment to Receive Salary (Afg)	0.61	–0.04	106
	[0.1]	[0.16]	
Experienced Delay	44.77	–6.41	115
	[6.24]	[7.5]	
Travel Time to Cash-Out (Min)	0.5	0.07	116
	[0.11]	[0.15]	
Very Satisfied with Pay System	0.55	–0.09	116
	[0.12]	[0.16]	
Vote in Favor of MSP	23.18	–7.61	116
	[6.39]	[6.95]	

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

Table A.8: Heterogeneity Treatment Effects – Security

	Payment to Receive Transfer (1)	Experienced Delay (2)	Travel Time to Cash-Out Salary (Min) (3)	Very Satisfied with the Pay System (4)	Vote in Favor of MSP (5)	Conducted Mobile Money Transaction (6)
Panel A. Year 1 Outcomes						
β_1: MSP × Controlled	–7.947	0.383***	42.176***	–0.365***	0.114	0.285***
	(5.354)	(0.079)	(10.171)	(0.083)	(0.074)	(0.052)
β_2: MSP × Contested	–7.888	0.551	17.650	–0.478	–0.011	0.060
	(4.797)	(0.341)	(14.328)	(0.303)	(0.081)	(0.087)
p-value $\beta_1 = \beta_2$	0.993	0.631	0.164	0.717	0.254	0.026
Control Mean	19.094	0.441	31.219	0.598	0.461	0.067
Observations	838	848	810	845	829	844
R squared	0.008	0.071	0.052	0.007	0.015	0.057
# Reg. Zones	310	310	303	310	308	310
Panel B. Year 2 Outcomes						
β_1: MSP × Controlled	–19.720***	–0.095	2.580	0.014	0.235**	
	(4.272)	(0.091)	(5.954)	(0.087)	(0.093)	
β_2: MSP × Contested	–10.358	0.402	–6.923	–0.313	0.416**	
	(7.233)	(0.251)	(30.580)	(0.283)	(0.199)	
p-value $\beta_1 = \beta_2$	0.266	0.064	0.761	0.271	0.411	
Control Mean	10.178	0.284	36.629	0.761	0.710	
Observations	644	642	632	643	623	
R squared	0.058	–0.012	–0.009	–0.003	0.102	
# Reg. Zones	286	287	283	287	282	

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

Table A.9: Heterogeneity Treatment Effects – Security

	Payment to Receive Transfer (1)	Experienced Delay (2)	Travel Time to Cash-Out Salary (Min) (3)	Very Satisfied with the Pay System (4)	Vote in Favor of MSP (5)	Conducted Mobile Money Transaction (6)
Panel A. Year 1 Outcomes						
β_1: MSP × Controlled	–5.532	0.340***	38.476***	–0.349***	0.162**	0.318***
	(5.393)	(0.080)	(9.180)	(0.084)	(0.075)	(0.053)
β_2: MSP × Contested	–24.009	0.660**	44.166	–0.528*	–0.192	–0.048
	(14.836)	(0.266)	(40.023)	(0.272)	(0.188)	(0.129)
p-value $\beta_1 = \beta_2$	0.243	0.250	0.890	0.529	0.081	0.009
Control Mean	19.094	0.441	31.219	0.598	0.461	0.067
Observations	838	848	810	845	829	844
R squared	0.001	0.060	0.054	0.001	0.004	0.051
# Reg. Zones	310	310	303	310	308	310
Panel B. Year 2 Outcomes						
β_1: MSP × Controlled	–15.523***	–0.172*	2.513	0.094	0.240**	
	(3.761)	(0.101)	(6.508)	(0.093)	(0.098)	
β_2: MSP × Contested	–30.342**	0.352**	1.274	–0.397**	0.289	
	(12.907)	(0.164)	(15.926)	(0.183)	(0.178)	
p-value $\beta_1 = \beta_2$	0.271	0.007	0.943	0.018	0.813	
Control Mean	10.178	0.284	36.629	0.761	0.710	
Observations	644	642	632	643	623	
R squared	0.068	–0.044	–0.001	–0.031	0.102	
# Reg. Zones	286	287	283	287	282	

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

Table A.10: Validating Learning Assessment Scores

	Math Score		Reading Score		Combined Score	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.300***	0.301***	0.251***	0.254***	0.275***	0.277***
	(0.044)	(0.045)	(0.045)	(0.046)	(0.042)	(0.043)
Female (=1)		–0.223*		–0.033		–0.128
		(0.126)		(0.131)		(0.118)
High SES (=1)		0.369***		0.230		0.300**
		(0.139)		(0.140)		(0.129)
# Students	242	242	242	242	242	242
R^2	0.172	0.363	0.121	0.324	0.165	0.357
Fixed Effects	None	District	None	District	None	District
Mean Outcome (0 Yrs Ed.)	–0.619	–0.619	–0.456	–0.456	–0.538	–0.538
Mean Outcome (4 Yrs Ed.)	0.558	0.558	0.485	0.485	0.522	0.522

Notes: This table reports on the relationship between students' years of education and their standardized scores on the learning assessment conducted by our team. The learning assessment is described in Section 3.5. Scores are standardized using the control group mean and standard deviation. The sample comprises students aged between 6 and 10 years who attended a public school eligible for the Mobile Salary Payments reform within the last three months for whom we have demographic data in control registration zones. High SES is a variable for students living in High Socio-Economic Status households, which are on a maintained road with access to water and to electricity. Robust standard errors are reported in parentheses.
Levels of significance: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.11: Learning Assessment Mechanisms

	Joined After Registration			Tenure of Teachers		
	All	Urban	Rural	All	Urban	Rural
	(1)	(2)	(3)	(4)	(5)	(6)
EE (=1)	–0.021**	–0.027*	–0.012	0.551*	0.236	1.087**
	(0.010)	(0.014)	(0.012)	(0.302)	(0.402)	(0.427)
ED (=1)	–0.019**	–0.023*	–0.011	0.476*	0.432	0.557
	(0.009)	(0.013)	(0.009)	(0.284)	(0.375)	(0.415)
Observations	28,099	16,903	11,196	27,805	16,736	11,069
R^2	0.046	0.037	0.060	0.127	0.127	0.126
Control Mean	0.140	0.138	0.143	10.725	10.969	10.367
# Reg. Zones	345	195	150	345	195	150

Notes: This table presents the effects of the MSP reform on employee retention. Joined After Registration indicates if the employee joins in the first year of the experiment (after registration and before the learning assessment). Tenure of teachers is measured in years. EE is the Early Registration, Early Payment (Treatment 1) group; and ED is the Early Registration, Delayed Payment (Treatment 2) group.The sample is restricted to all employees paid at any point before the reform or during the first year of the reform. All regressions include district fixed effects. Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.12: Wage Bill Changes Post-Registration At Registration Zone Level

	May Treatment				November Treatment			
	Employees	Net Total	Net	Salary	Employees	Net Total	Net	Salary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Whole Sample								
Post × Treatment	–0.865	–731.021	122.423	–1615.307	–1.302**	–7509.652	–1457.747	–2179.397
	(0.623)	(7102.755)	(6069.442)	(4431.998)	(0.659)	(7577.718)	(6391.744)	(4747.110)
Observations	12,420	12,420	12,420	12,420	12,420	12,420	12,420	12,420
R squared	0.971	0.898	0.963	0.988	0.971	0.898	0.963	0.988
Control Mean (Pre)	62.090	548791.375	502521.094	453957.844	61.490	540936.062	499737.031	452287.844
# Reg. Zones	345	345	345	345	345	345	345	345
Panel B. Restricted Sample								
Post × Treatment	–0.221	5112.452	469.371	75.331	–0.173	4460.139	790.093	488.600
	(0.402)	(4475.010)	(3614.157)	(2711.572)	(0.478)	(4875.927)	(4216.192)	(3266.067)
Observations	12,384	12,384	12,384	12,384	12,384	12,384	12,384	12,384
R squared	0.994	0.947	0.970	0.992	0.994	0.947	0.970	0.992
Control Mean (Pre)	56.100	493261.188	468244.375	426134.938	56.100	490774.812	468480.625	426592.125
# Reg. Zones	344	344	344	344	344	344	344	344

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

Table A.13: Treatment Effects on Attendance

	DD Mean (1)	EE Effect (2)	ED Effect (3)	ToT Effect (4)	N Obs (5)
Share Present at Baseline	0.382	0.009	0.007		14,047
	[0.025]	[0.030]	[0.027]		
Share Present at Endline 1	0.356	–0.039*	–0.018	–0.052	13,793
	[0.020]	[0.023]	[0.022]	[0.046]	

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

Table A.14: Treatment Effects on Attendance – By Urban & Rural

	DD Mean (1)	EE Effect (2)	ED Effect (3)	ToT Effect (4)	N Obs (5)
Panel A. Urban					
Share Present at Baseline	0.330	0.024	0.023		10,790
	[0.026]	[0.037]	[0.033]		
Share Present at Endline 1	0.324	–0.035	–0.012	–0.058	10,682
	[0.021]	[0.027]	[0.026]	[0.049]	
Panel B. Rural					
Share Present at Baseline	0.569	–0.049	–0.056		3,257
	[0.042]	[0.039]	[0.035]		
Share Present at Endline 1	0.470	–0.058	–0.040	–0.068	3,111
	[0.034]	[0.046]	[0.036]	[0.083]	

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

Table A.15: Heterogeneity Impacts on Mobile Money Use from Administrative Data - Urban & Rural

	Deposit (1)	Transfer to Cust. Wallet (2)	Own Airtime Top-up (3)	Other Airtime Top-up (4)	Balance (5)
Panel A. OLS					
β_1: MSP Months × Urban	0.004***	0.003	0.001***	0.002***	0.009***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.002)
β_2: MSP Months × Rural	0.003**	0.003	0.001***	0.001***	0.025***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.004)
Observations	14,753	14,753	14,753	14,753	14,753
R^2	0.005	0.015	0.022	0.042	0.128
1-Month Mean	0.002	0	0.000	0.000	0.137
p-value $\beta_1 = \beta_2$	0.439	0.982	0.604	0.047	0.000
# Reg. Zones	284	284	284	284	284
Panel B. 2SLS					
β_1: MSP Months × Urban	0.004***	0.004	0.001***	0.002***	0.006*
	(0.001)	(0.003)	(0.000)	(0.001)	(0.003)
β_2: MSP Months × Rural	0.006**	0.002**	0.001	0.002**	0.003
	(0.003)	(0.001)	(0.001)	(0.001)	(0.009)
Observations	14,753	14,753	14,753	14,753	14,753
R^2	0.002	0.001	0.011	0.018	0.004
p-value $\beta_1 = \beta_2$	0.578	0.658	0.976	0.495	0.789
# Reg. Zones	284	284	284	284	284

Notes: This table reports estimates of the impact of receiving salary via mobile money on other dimensions of mobile money use. The MSP months variable counts the number of months that an employee has been paid by mobile money. 2SLS outcomes instrument the number of months with the treatment assignment. Outcome data reflect transactions conducted between October 2018 and December 2020. Deposits are money added to the mobile money wallet via agents or bank transfers. Transfers to customer wallet are peer-to-peer transfers to another mobile money user. Airtime top-ups are money added to the pre-paid mobile phone call plan. Pre-pay balance is the remaining balance the day before salary payment for each employee. All variables are normalized by monthly salary. Robust standard errors clustered at the registration zone (treatment unit) level are reported in parentheses. Urban zones are those that have a population density of more than 300 inhabitants per squared kilometer.
Levels of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table A.16: Beliefs About State Durability and Registration During On-site Drive

Dependent Variable	Registered During Drive (=1)				
	(1)	(2)	(3)	(4)	(5)
AM p × AM p(1-p) × BM SD of p	0.099*		0.001	0.099*	0.001
	(0.056)		(0.088)	(0.056)	(0.088)
AM p × BM p(1-p) × AM SD of p	0.058		0.043	0.058	0.043
	(0.060)		(0.061)	(0.060)	(0.061)
AM p × BM p(1-p) × BM SD of p	0.122**		0.159***	0.122**	0.159***
	(0.055)		(0.055)	(0.055)	(0.055)
BM p × AM p(1-p) × BM SD of p	–0.026		0.064	–0.026	0.064
	(0.057)		(0.073)	(0.057)	(0.073)
BM p × BM p(1-p) × AM SD of p	0.127**		0.083	0.127**	0.083
	(0.059)		(0.065)	(0.059)	(0.065)
Taliban Will Lose Pre-Reform (2017, %)		0.413	0.327		0.327
		(0.300)	(0.193)		(0.193)
Taliban Will Lose Pre-Reform (2016, %)		0.036	–0.611***		–0.611***
		(0.348)	(0.215)		(0.215)
Taliban Will Lose Pre-Reform (2015, %)		0.444	0.278		0.278
		(0.366)	(0.334)		(0.334)
Police Capable of Protection (2018, %)			0.052		0.052
			(0.110)		(0.110)
Government is Effective (Overall, 2018, %)			–0.575***		–0.575***
			(0.150)		(0.150)
Pashtun (2018, %)			–0.055		–0.055
			(0.047)		(0.047)
On Primary Road (=1)			0.085*		0.085*
			(0.047)		(0.047)
Urban (=1)			0.010		0.010
			(0.024)		(0.024)
Contested District (=1)			–0.148***		–0.148***
			(0.045)		(0.045)
Bad Cell Coverage (%)			0.057		0.057
			(0.067)		(0.067)
Based at Registration School (=1)			0.046**		0.046**
			(0.022)		(0.022)
Female			0.017		0.017
			(0.021)		(0.021)
Has At Least A Bachelor's			0.061*		0.061*
			(0.031)		(0.031)
Early Reg., Early Mobile Money			0.192***		0.192***
			(0.035)		(0.035)
Late Reg., Early Mobile Money			0.193***		0.193***
			(0.037)		(0.037)
Constant	0.757***	0.009	0.863**	0.757***	0.863**
	(0.050)	(0.349)	(0.323)	(0.050)	(0.323)
Observations	19,626	19,626	19,626	19,626	19,626
R squared	0.023	0.011	0.099	0.023	0.099
# Districts	31	31	31	31	31
Mean Dependent Var	0.809	0.809	0.809	0.809	0.809

Notes: This table reports on the relationship between citizen's beliefs about who will win the war and the likelihood that employees appeared for in-person registration for the new salary system. The dependent variable is a dummy variable equal to one for employees who registered during the registration drive. Data on beliefs are from The Afghanistan Nationwide Quarterly Assessment Research (ANQAR) survey. Data for a given year include all the survey waves in said year. Baseline measures are from the 2018 survey rounds and measures pre-reform are from the 2015-2017 rounds. The sample includes teachers from all three treatment arms who were paid at least once before the reform began. Standard errors clustered at the district level are reported in parentheses. AM stands for Above Median and BM for Below Median.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.17: Beliefs About State Durability and Registration at Any Time

Dependent Variable	Registered at Any Time (=1)				
	(1)	(2)	(3)	(4)	(5)
Taliban Will Lose at Baseline (2018, %)	0.676***		0.547*	3.100*	2.607
	(0.219)		(0.280)	(1.766)	(1.757)
SD of Beliefs				2.375	2.114
				(1.777)	(1.693)
Levels of Beliefs × SD of Beliefs				–2.392	–2.421
				(1.816)	(1.822)
Taliban Will Lose Pre-Reform (2017, %)		0.293	0.013		–0.025
		(0.230)	(0.225)		(0.228)
Taliban Will Lose Pre-Reform (2016, %)		0.054	–0.363*		–0.363*
		(0.215)	(0.214)		(0.204)
Taliban Will Lose Pre-Reform (2015, %)		0.288	0.174		0.252
		(0.230)	(0.234)		(0.215)
Police Capable of Protection (2018, %)			0.060		0.047
			(0.109)		(0.113)
Government is Effective (Overall, 2018, %)			–0.217**		–0.215**
			(0.103)		(0.099)
Pashtun (2018, %)			–0.042		–0.054
			(0.031)		(0.035)
On Primary Road (=1)			0.063***		0.057***
			(0.020)		(0.017)
Urban (=1)			0.013		0.029
			(0.024)		(0.030)
Contested District (=1)			–0.046		–0.036
			(0.029)		(0.032)
Bad Cell Coverage (%)			–0.022		–0.003
			(0.050)		(0.057)
Based at Registration School (=1)			0.039*		0.041**
			(0.020)		(0.020)
Female			–0.001		0.001
			(0.012)		(0.011)
Has At Least A Bachelor's			0.055**		0.055**
			(0.026)		(0.026)
Early Reg., Early Mobile Money			0.199***		0.198***
			(0.036)		(0.036)
Late Reg., Early Mobile Money			0.201***		0.200***
			(0.036)		(0.036)
Constant	0.223	0.280	0.414*	–2.178	–1.413
	(0.208)	(0.237)	(0.238)	(1.740)	(1.551)
Observations	19,626	19,626	19,626	19,626	19,626
R squared	0.007	0.006	0.087	0.009	0.089
# Districts	31	31	31	31	31
WC Bootstrap p-value	0.054		0.175	0.117	0.268
Mean Dependent Var	0.848	0.848	0.848	0.848	0.848

Notes: This table reports on the relationship between citizen's beliefs about who will win the war and the likelihood that employees appeared for in-person registration for the new salary system. The dependent variable is a dummy variable equal to one for employees who registered during the registration drive. Data on beliefs are from The Afghanistan Nationwide Quarterly Assessment Research (ANQAR) survey. Data for a given year include all the survey waves in said year. Baseline measures are from the 2018 survey rounds and measures pre-reform are from the 2015-2017 rounds. The sample includes teachers from all three treatment arms who were paid at least once before the reform began, but excludes unsafe schools and schools where less than 10% of employees paid in the 12 months prior to the reform registered. Standard errors clustered at the district level are reported in parentheses.
Levels of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table A.18: Beliefs About State Durability and Registration During On-site Drive – Controlling for Province Fixed Effects

Dependent Variable	Registered During Drive (=1)				
	(1)	(2)	(3)	(4)	(5)
Taliban Will Lose at Baseline (2018, %)	–0.159		–0.038	1.053	1.980
	(0.313)		(0.338)	(1.943)	(1.809)
SD of Beliefs				1.403	2.587*
				(1.795)	(1.497)
Levels of Beliefs × SD of Beliefs				–1.770	–3.447**
				(1.801)	(1.564)
Taliban Will Lose Pre-Reform (2017, %)		–0.202	–0.184		–0.314*
		(0.305)	(0.198)		(0.175)
Taliban Will Lose Pre-Reform (2016, %)		–0.414	–0.573**		–0.514***
		(0.248)	(0.232)		(0.179)
Taliban Will Lose Pre-Reform (2015, %)		0.268	0.139		0.337*
		(0.270)	(0.263)		(0.189)
Police Capable of Protection (2018, %)			0.012		–0.025
			(0.091)		(0.079)
Government is Effective (Overall, 2018, %)			–0.191		–0.133
			(0.113)		(0.098)
Pashtun (2018, %)			–0.107		–0.072
			(0.069)		(0.048)
On Primary Road (=1)			–0.001		–0.026
			(0.028)		(0.020)
Urban (=1)			–0.002		0.019
			(0.025)		(0.029)
Contested District (=1)			–0.106***		–0.105***
			(0.035)		(0.036)
Bad Cell Coverage (%)			–0.064		–0.049
			(0.049)		(0.042)
Based at Registration School (=1)			0.045**		0.051**
			(0.021)		(0.020)
Female			0.015		0.020
			(0.021)		(0.020)
Has At Least A Bachelor's			0.061*		0.060*
			(0.031)		(0.031)
Early Reg., Early Mobile Money			0.190***		0.189***
			(0.034)		(0.033)
Late Reg., Early Mobile Money			0.185***		0.183***
			(0.036)		(0.035)
Constant	0.956***	1.129***	1.457***	0.049	–0.007
	(0.292)	(0.308)	(0.401)	(1.961)	(1.694)
Observations	19,626	19,626	19,626	19,626	19,626
R squared	0.036	0.039	0.108	0.038	0.114
# Districts	31	31	31	31	31
WC Bootstrap p-value	0.714		0.936	0.571	0.391
Mean Dependent Var	0.809	0.809	0.809	0.809	0.809

Notes: This table reports on the relationship between citizen's beliefs about who will win the war and the likelihood that employees appeared for in-person registration for the new salary system. The dependent variable is a dummy variable equal to one for employees who registered during the registration drive. Data on beliefs are from The Afghanistan Nationwide Quarterly Assessment Research (ANQAR) survey. Data for a given year include all the survey waves in said year. Baseline measures are from the 2018 survey rounds and measures pre-reform are from the 2015-2017 rounds. The sample includes teachers from all three treatment arms who were paid at least once before the reform began. Standard errors clustered at the district level are reported in parentheses. Controls for province fixed effects.

Levels of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table A.19: Registration Outcomes

	Registered on Time (1)	Registered With Delay (2)	Unregistered (3)
Panel A. Share Registered			
Paid In Each Of The 6 Months Before Start Of Registration	94%	3.22%	2.78%
	13333	457	394
Paid At Any Point Before Start Of Registration	85.95%	4.82%	9.23%
	14837	832	1593
Paid At Any Point	73.68%	15.12%	11.2%
	15608	3203	2372
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 i) in each of the 6 months between December 2017 and April 2018 (i.e., the 6 months before the start of registration – first row of Panel A, and Panel B), ii) at any point between March 2017 and April 2018 (i.e., before the start of registration – second row of Panel A), or iii) at any point between March 2017 and March 2020 (i.e., the entire period for which we have data – third row of Panel A). 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.20: Spatial Treatment Externalities on Salary Payment Experience (within 2 km)

	DD Mean (1)	EE Effect (2)	ED Effect (3)	# EE Neigh Schools 2 km (4)	# ED Neigh Schools 2 km (5)	# Total Neighb Schools 2 km (6)	N Obs. (7)
Panel A. Endline 1 – April 2019							
Payment to Receive Salary (Afg)	0.44	0.24***	0.16***	0.01	–0.01	0.01	848
	[0.04]	[0.06]	[0.06]	[0.02]	[0.02]	[0.01]	
Experienced Delay	0.6	–0.24***	–0.14**	–0.01	–0.01	0.00	845
	[0.04]	[0.06]	[0.06]	[0.02]	[0.02]	[0.01]	
Travel Time to Cash-Out (Min)	19.09	–8.04**	–3.68	1.94	–0.74	–0.35	838
	[2.96]	[3.32]	[3.31]	[1.46]	[1.04]	[0.57]	
Vote in Favor of MSP	0.46	0.04	0.06	0.02	–0.02	0.00	829
	[0.05]	[0.05]	[0.05]	[0.02]	[0.02]	[0.01]	
Conducted Mobile Money Transfer	31.22	27.27***	22.31***	0.55	0.2	–1.01	810
	[3.95]	[7.41]	[7.4]	[1.84]	[1.82]	[0.83]	
Panel B. Endline 2 – May 2020							
Payment to Receive Salary (Afg)	0.28	–0.04	–0.02	0.01	0.01	–0.01	642
	[0.03]	[0.04]	[0.05]	[0.01]	[0.02]	[0.01]	
Experienced Delay	0.76	–0.01	–0.02	0.00	0.01	0.00	643
	[0.04]	[0.05]	[0.05]	[0.01]	[0.02]	[0.01]	
Travel Time to Cash-Out (Min)	10.18	–9.8***	1.74	1.56	–0.36	–0.25	644
	[2.05]	[2.62]	[2.54]	[1.08]	[0.95]	[0.4]	
Vote in Favor of MSP	0.71	0.09*	–0.01	0.02**	0.00	0.00	623
	[0.04]	[0.05]	[0.04]	[0.01]	[0.02]	[0.01]	
Conducted Mobile Money Transfer	36.63	0.54	–3.26	–0.18	0.38	–0.27	632
	[2.44]	[3.47]	[3.05]	[1.03]	[0.98]	[0.58]	

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 as dicussed in Appendix D. The Endline 1 (2) sample consists of 970 (739) MoE employees who participated in the full survey and self-reported the payment system. DD is the Delayed Registration, Delayed Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and ED is the Early Registration, Delayed Payment (Treatment 2) group. # EE (ED) Neigh Schools 2 km is the number of neighboring schools within a two-kilometer radius located in EE (ED) treatment zone, # Total Neigh Schools 2 km is the total number of neighboring schools within the same radius. Robust standard errors clustered at the registration zone (treatment unit) level in squared brackets. *Levels of significance*: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.21: Effects on Salary Payment Experience – With Stratum FEs

	DD Mean (1)	EE Effect (2)	EL Effect (3)	ToT Effect (4)	N Obs (5)
Panel A. Endline 1					
Payment to Receive Salary (Afg)	19.094	–5.427	–3.655	–3.683	838
	(2.957)	(3.556)	(3.305)	(6.487)	
		[0.056]	[0.118]		
Experienced Delay	0.441	0.262***	0.160***	0.376***	848
	(0.042)	(0.054)	(0.056)	(0.084)	
		[0.001]	[0.001]		
Travel Time to Cash-Out (Min)	31.219	27.233***	22.131***	47.841***	810
	(3.947)	(6.865)	(7.055)	(10.190)	
		[0.001]	[0.001]		
Very Satisfied with Pay System	0.598	–0.252***	–0.147***	–0.359***	845
	(0.042)	(0.054)	(0.054)	(0.089)	
		[0.001]	[0.001]		
Vote in Favor of MSP	0.461	0.076	0.059	0.100	829
	(0.046)	(0.048)	(0.049)	(0.094)	
		[0.056]	[0.118]		
Conducted Mobile Money Transfer	0.067	0.181***	0.170***	0.262***	844
	(0.017)	(0.035)	(0.036)	(0.051)	
		[0.001]	[0.001]		
Panel B. Endline 2					
Payment to Receive Salary (Afg)	10.178	–7.884***	2.076	–20.190***	644
	(2.054)	(2.149)	(2.767)	(3.924)	
		[0.001]	[0.893]		
Experienced Delay	0.284	–0.030	–0.016	0.010	642
	(0.034)	(0.045)	(0.044)	(0.099)	
		[0.913]	[0.981]		
Travel Time to Cash-Out (Min)	36.629	–0.081	–3.341	11.278	632
	(2.444)	(3.280)	(3.023)	(7.504)	
		[0.981]	[0.619]		
Very Satisfied with Pay System	0.761	–0.013	–0.020	–0.093	643
	(0.035)	(0.045)	(0.045)	(0.104)	
		[0.981]	[0.981]		
Vote in Favor of MSP	0.710	0.110**	–0.007	0.170	623
	(0.040)	(0.047)	(0.045)	(0.103)	
		[0.015]	[0.981]		

Notes: This table reports treatment effects on salary payment experience, controlling for the district (strata) fixed effects. The Endline 1 (2) sample consists of 970 (739) MoE employees who participated in the full survey and self-reported their payment system. DD is the Delayed Registration, Delayed Payment (Control) group; EE is the Early Registration, Early Payment (Treatment 1) group; and ED is the Early Registration, Delayed 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. FWER-adjusted p-values within each panel in squared brackets (following Romano & Wolf, 2005, using 1000 repetitions). Appendix D discusses spatial externalities of the MSP reform on salary payment experience across studied schools. The results are presented in Appendix Table A.20.
Levels of significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table A.22: Estimates of Impact on Employees' Payment Experience By Beliefs (Sample: Median Based on Survey Data, All Districts)

	Payment to Receive Transfer (1)	Experienced Delay (2)	Travel Time to Cash-Out Salary (Min) (3)	Very Satisfied with the Pay System (4)	Vote in Favor of MSP (5)	Conducted Mobile Money Transaction (6)
Panel A. Year 1 Outcomes						
β_1: MSP × Above Median Taliban Will Lose at Baseline (2018, %)	–6.138	0.378***	46.065***	–0.374***	0.176*	0.272***
	(7.942)	(0.104)	(14.823)	(0.107)	(0.091)	(0.063)
β_2: MSP × Below Median Taliban Will Lose at Baseline (2018, %)	–11.504*	0.414***	42.082***	–0.545***	0.008	0.209**
	(6.432)	(0.146)	(15.365)	(0.163)	(0.127)	(0.087)
Observations	717	726	696	724	708	722
R squared	0.254	0.238	0.283	0.140	0.329	0.174
Control Mean	20.276	0.427	29.219	0.606	0.469	0.069
p-value $\beta_1 = \beta_2$	0.600	0.840	0.852	0.381	0.283	0.554
# Reg. Zones	267	267	261	267	265	267
Panel B. Year 2 Outcomes						
β_1: MSP × Above Median Taliban Will Lose at Baseline (2018, %)	–13.172**	–0.247*	1.926	0.064	0.193*	
	(6.080)	(0.130)	(9.974)	(0.138)	(0.111)	
β_2: MSP × Below Median Taliban Will Lose at Baseline (2018, %)	–24.916***	0.102	3.356	0.046	0.293*	
	(4.760)	(0.152)	(11.117)	(0.118)	(0.178)	
Observations	546	545	536	546	529	
R squared	0.244	0.130	0.120	0.166	0.286	
Control Mean	10.610	0.293	36.642	0.750	0.723	
p-value $\beta_1 = \beta_2$	0.128	0.081	0.924	0.918	0.632	
# Reg. Zones	244	245	241	245	241	

Notes: This table reports ToT estimates of impacts of mobile salary payments on payment quality and mobile money use. Corresponding ITT estimates are reported in Appendix Tables **??** and A.21. Favor MSP is a dummy variable equal to one if the respondent indicates they would support scaling the reform across the Ministry of Education. Satisfied with Pay System is a dummy equal to one of the respondent indicates a very high level of satisfaction. Experienced Delay is a dummy variable equal to one if the respondent reports their salary being delayed. Travel Time is the time to convert the mobile money payment to cash in minutes. Payment to Receive measures what respondents report paying to receive their salary in Afghanis. For comparison, the average (net) salary in the sample is 8560 Afg. Money Transfer is a dummy variable equal to one if the respondent indicates making a mobile money transfer to someone else in the previous month. This outcome is not recorded in year 2 because the survey was abbreviated due to the pandemic. Robust standard errors clustered at the registration zone level are reported in parentheses. Data for a given year include all the survey waves in said year. Excluding registration zones with less than 10% of teachers in the EE and EL groups by April 2019.

Table A.23: Estimates of Impact on Student Learning (Sample: Median Based on Survey Data, All Districts)

	Math Score			Reading Score			Combined Score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β_1: Treated School × Above Median Taliban Will Lose at Baseline (2018, %)	0.046	0.472**	–0.277	0.059	0.219	–0.041	0.053	0.346	–0.159
	(0.158)	(0.230)	(0.194)	(0.145)	(0.224)	(0.179)	(0.138)	(0.211)	(0.166)
β_2: Treated School × Below Median Taliban Will Lose at Baseline (2018, %)	0.219	0.300*	0.047	0.251*	0.366**	–0.162	0.235*	0.333**	–0.058
	(0.133)	(0.157)	(0.233)	(0.150)	(0.181)	(0.227)	(0.129)	(0.157)	(0.194)
# Students	719	395	324	719	395	324	719	395	324
R squared	0.340	0.301	0.396	0.253	0.230	0.302	0.329	0.297	0.385
Sample	all	urban	rural	all	urban	rural	all	urban	rural
p-value $\beta_1 = \beta_2$	0.408	0.542	0.294	0.360	0.612	0.680	0.340	0.961	0.697
Untreated School Mean	0.050	0.020	0.090	0.040	0.050	0	0.050	0.040	0.040
# Reg. Zones	318	189	129	318	189	129	318	189	129
FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports the impacts of the Mobile Salary Payment Reform on student learning outcomes measured in the assessment described in Section 3.5. Scores are standardized using the control group mean and standard deviation. The sample comprises students aged between 6 and 10 years who attended a public school eligible for the Mobile Salary Payments reform within the last three months. Treated School is a dummy variable equal to one for students who attend a school where at least 50% of teachers were paid using mobile money by April 2019, the month before the learning assessment. Data on beliefs about who will win the war are from ANQAR survey. Data on beliefs for 2018 include all the survey waves in 2018. Covariates include years of education, years of age, a dummy variable equal to one if a student lives in a high socio-economic status household, and gender. High SES households are those that are on a maintained road with access to water and to electricity. Standard errors clustered at the registration zone level are reported in parentheses. Data for a given year include all the survey waves in said year. Excluding registration zones with less than 10% of teachers in the EE and EL groups by April 2019.

Table A.24: Heterogeneity Analysis Mobile Money Use, Standardised by Salary – Beliefs

	Deposit (1)	Transfer to Cust. Wallet (2)	Own Airtime Top-up (3)	Other Airtime Top-up (4)	Balance (5)
Panel A. OLS					
β_1: MSP Months × Above Median Taliban Will Lose at Baseline (2018, %)	0.001	0.000	0.000	0.000*	0.016***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.005)
β_2: MSP Months × Below Median Taliban Will Lose at Baseline (2018, %)	0.004***	0.003*	0.001***	0.002***	0.012***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.002)
Observations	13,324	13,324	13,324	13,324	13,324
R^2	0.005	0.013	0.022	0.047	0.128
1-Month Mean	0.002	0	0.000	0.000	0.138
p-value $\beta_1 = \beta_2$	0.102	0.207	0.000	0.000	0.396
# Reg. Zones	251	251	251	251	251
Panel B. 2SLS					
β_1: MSP Months × Above Median Taliban Will Lose at Baseline (2018, %)	0.009	0.004	–0.001	–0.000	–0.007
	(0.011)	(0.014)	(0.001)	(0.001)	(0.036)
β_2: MSP Months × Below Median Taliban Will Lose at Baseline (2018, %)	0.004***	0.003	0.001***	0.002***	0.006**
	(0.001)	(0.003)	(0.000)	(0.000)	(0.003)
Observations	13,324	13,324	13,324	13,324	13,324
R^2	0.000	0.001	0.011	0.024	0.002
p-value $\beta_1 = \beta_2$	0.649	0.932	0.019	0.002	0.717
# Reg. Zones	251	251	251	251	251

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

C Biometric Registration and Employee Identification

This appendix: (i) reports the share of teachers on the payroll prior to the reform who were registered, (ii) documents the share of employees removed from the payroll following biometric registration, and (iii) explores efforts to "game" the registration requirement and attempts to quantify the number of total ghosts and stand-ins on the payroll.

C.1 Identifying Employees

We exploit the rollout of biometric registration to estimate the number of ghost MoE employees. As can be seen in the first row of Panel A of Appendix Table A.19, the majority of employees in early registration zones registered during the first registration wave (94%) or with some delays (3.2%). 2.8% of these employees never registered (despite official warnings that failing to do so would prevent them from being paid), providing a conservative, lower bound estimate for the number of ghosts.[47,48] Their salaries amount to about AFN 3.3 million (USD 43,000 using the 2018/19 exchange rate) or 2.6% of the total monthly wage bill in these schools.

C.2 How many employees were removed from the payroll?

However, not all of the people who failed to register were removed from the payroll. After the primary registration wave, the MoE formed an adjudication committee tasked with eliminating ghost employees. The committee compared payroll disbursement records to registration records and determined the identity of employees who did not register or whose records did not match. If an employee could not be confirmed, the committee was responsible for removing them from the payroll. Using payroll records, we can track whether MoE continued to request salaries for employees who never registered.[49] Panel B of Table A.19 (column 3) shows the share of never-registered employees who continued receiving their salaries after the registration process (among those who were paid in each of the six months before registration activities started). 89.6% of them remain on the payroll one month after the start of the reform. This share decreases to 56.9% within 12 months of the registration.[50]

[47] To generate conservative estimates, we exclude all employees who were paid only during some, but not all, of the six months before the registration, in order to conservatively estimate the share of employees who never registered among those who clearly appear as the most regular and stable employees according to the payroll records. In the other rows of Panel A we relax this restriction, and include either all employees paid at any point before the start of registration activities (row 2), or paid at any point in time (row 3). The share of employees who registered during the first registration wave in these two groups is respectively 86% and 73.7%, while the share who never registered is respectively 9.2% and 11.2%.

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

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

[50] The incomplete removal of non-registered employees limits the effect of biometric registration on the payroll. In Appendix Table A.12, we test for post-registration reductions in the total wage bill aggregating

However, even 20 months after the registration a salary was still being requested for 53.6% of never-registered employees. As a result, the incomplete removal of employees who never registered limited savings to only 1% (AFN 1.3 million or USD 17,000) of the total monthly wage bill in the EE and EL registration zones.

C.3 The Persistent Problem of Stand-in Employees

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

To better quantify the extent of this problem, we conducted a short litmus test phone survey with MoE employees between February and March 2019. We sampled a total of 2,663 employees who had a phone number listed in the payroll records. Importantly, these records are from a roster held by the government prior to the reform. From the Early Registration groups, we included all employees who did not register, all who registered after the in-person registration drive, and a random sample of those who registered on time for comparison (see below for further details on the sampling procedure). The litmus test then asked employees seven simple questions about their job, which any legitimate employee should be able to answer easily.

An ideal litmus test would perfectly discriminate between true employees and stand-ins. Such a test requires a set of questions that: i) would be easy enough for all true employees to answer correctly, ii) would be non-trivial, so that stand-ins could not answer correctly, and iii) could be graded using official records. While our litmus test was not perfect (true employees can make mistakes and stand-ins can provide correct answers), it is nevertheless informative (as shown in Figure 5). Recognizing these limitations, we use litmus scores to estimate bounds on the share of stand-ins. The idea is simple: the lower are the scores of registered employees with respect to the scores of known true employees (who were certified to be present during a separate round of unannounced audit visits), the higher the number of stand-ins must be.

Bounding Exercise 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 $R = T + Z$.

at registration zone level. While the results are consistently negative, they are generally not significant at conventional levels. Thus, we cannot reject the null hypothesis that payroll reductions in the treatment group are equal to the standard turnover rates observed in the control group.

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.

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

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 $P(Z = 1)$.

Stand-ins Estimates 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% ($1 - \frac{66.69}{73.14}$), and simply answering the phone in a bound of 5.9% ($1 - \frac{81.04}{86.08}$). That is, we find that estimates of stand-ins range from 5.9% to 18.1% of registered employees. Adjusting these numbers for the share of registered employees (97.2%), the bounds range from 5.7% to 17.6% of all employees. Adding these potential stand-ins to the 2.8% of unregistered employees, we estimate that between 8.5% and 20.4% of all paid employees might be ghosts. If they were all removed from the payroll, the government would save AFN 26 million (USD 0.34 million) per month in the early registration zones (calculated as 20.4% of a monthly wage bill of about AFN 126 million).

Sampling Details We sampled a total of 2,663 employees who had a phone number listed in the payroll records, with a breakdown as follows:[51]

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.

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

[51] 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.

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. We sample one employee from each of these school.

In our analysis, we use sampling weights to adjust for this sampling selection procedure and recover estimates that represent the whole population from which we sampled.

D 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.20 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.21. 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.