Strengthening Fragile States: Evidence from Mobile Salary Payments in Afghanistan

Joshua Blumenstock Michael Callen Anastasiia Faikina
Stefano Fiorin Tarek Ghani*

December 11, 2025

Abstract

We conduct a randomized evaluation of a flagship government initiative to build administrative capacity in Afghanistan during a period of extreme fragility and uncertainty. The program aimed to improve employee management systems in the Ministry of Education by introducing biometric registration and digital payments for public-sector employees. The reform enhanced the state's ability to identify employees, boosted teacher retention, improved student learning outcomes, expanded financial inclusion, and strengthened employee support for scaling the initiative. Participation was higher in districts where citizens expressed greater confidence in — and consensus about — the government's prospects of defeating the Taliban. Together, the results demonstrate that digital technologies can improve state performance in fragile contexts, but also suggest that uncertainty about the state's survival may limit employees' willingness to engage in reforms.

Keywords: state capacity, fragile states, beliefs, institutions

JEL Classification: P00, O10, D70

*Blumenstock: UC Berkeley. Email: `jblumenstock@berkeley.edu`. Callen: LSE, CEPR, and CESifo. Email: `m.j.callen@lse.ac.uk`. 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, Saipremnath Muthukumaran, and Sami Safiullah for excellent research assistance. Oriana Bandiera, Eli Berman, Gharad Bryan, Ernesto Dal Bó, Jishnu Das, Asim Khwaja, Craig McIntosh, Karthik Muralidharan, Paul Niehaus, Rohini Pande, Imran Rasul, Jacob Shapiro, Charles Sprenger, Jonathan Weigel, Noam Yuchtman and 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 WUSTL. The authors declare they have no relevant or material financial interests that relate to the research described in this paper.

1 Introduction

Fragile states persistently fail to build the administrative foundations of effective governance. They are marked by contested authority, institutions at risk of collapse or radical change, and governments unable to implement even basic policies. Building administrative capacity is therefore a critical step in the transition from informal, personalized rule to more structured, enduring and effective institutions —- and a central part of how states exit fragility (Tilly, 1985; Olson, 1993; Besley and Persson, 2011; Acemoglu and Robinson, 2019).

This paper asks whether technology can help fragile governments build the capacity to perform basic administrative tasks, such as identifying and paying employees. Prior research has shown that technologies can strengthen state capacity in less fragile settings, such as India, Indonesia, Pakistan, Rwanda, and Uganda (Duflo, Hanna and Ryan, 2012; Callen et al., 2016; Muralidharan, Niehaus and Sukhtankar, 2016; Dhaliwal and Hanna, 2017; Callen et al., 2020; Muralidharan, Niehaus and Sukhtankar, 2020; Banerjee et al., 2020; Dodge et al., 2025; World Bank, 2024). Far less is known, however, about whether technology can help build foundational administrative functions at earlier stages of state development, when authority is contested and institutions are unstable. We address this gap through a field experiment conducted in partnership with the Afghan government to evaluate its flagship effort to modernize core administrative systems. The randomized trial was conducted in the lead-up to the government's collapse in 2021. To our knowledge, it represents the first large-scale experimental evaluation of whether new technologies advance foundational administrative functions in a highly fragile, low-capacity state.

We find that digital identification and mobile money helped a nascent government identify and pay employees more reliably amid extreme instability. The technologies rolled out through the reforms were popular with employees, reduced teacher turnover, and generated downstream gains in student test scores. We also provide suggestive evidence that political conditions closely tied to fragility—-including concerns that the national government might collapse—-influenced whether state employees actively participated in building administrative capacity. As a growing number of countries face periods of instability, understanding how fragile states can effectively initiate state-building is increasingly urgent (Besley, Collier and Khan, 2018; OECD, 2022; FCDO, 2023). Our findings reveal that technology can assist this process, even at an early and uncertain stage of state formation.

Afghanistan offers a valuable context to study these questions. Long marked by fragmented authority, weak institutions, and deep social divisions, it was among the most fragile states in the world during our study. Afghanistan's education system, last seriously developed in the 1960s, collapsed during decades of conflict and was only partially rebuilt after

2001. Moreover, despite unprecedented international support in the following years (Callen and Kabuli, 2022), the Afghan state lacked even the most basic administrative capabilities. There was no reliable system for verifying the identity, location, or status of public employees in much of the country. In our baseline audit, only 41% of employees were present in schools; among those present, just 30% held a government ID. Salary payments were also deeply unreliable: on average, employees had not been paid for 3.5 months, and 60% reported delays. Consistent with these findings, the Ministry of Education was under-performing: Afghanistan ranked 150th out of 176 countries in learning outcomes in 2020 (Angrist et al., 2021).[1] These conditions underscore the depth of administrative failure—even after two decades of state-building—and the critical need for more effective solutions.

Our study occurred during a window of opportunity, when the vulnerable government sought to strengthen core capacity and use modern research methods to evaluate impact. The randomized trial involved approximately 30,000 public employees (mainly teachers) working in more than 1,000 schools across three conflict-affected provinces (Kandahar, Nangarhar, and Parwan) that together account for about 15% of Afghanistan's population. Roughly one third of the geographic area in our study was contested by or controlled by the Taliban, the Islamic State, and other armed groups. The design separately randomized the timing of biometric registration and the switch to mobile-money salary payments. Our analysis combines payroll records from the Ministry of Education, transaction records from the mobile network operator, and original surveys of employees and students.

The experiment yields three main findings. First, biometric registration substantially improved the state's ability to track employees, creating, for the first time, an up-to-date single registry of employees. Using this registry and a complementary employee screening exercise, we estimate that between 8.3% and 14.1% of employees are ghosts.[2] Second, the new technology's popularity increased in areas where it was implemented: treatment increased employee support for scaling the reform by 25 percentage points after two years, and a clear majority of employees favored expansion. Yet the transition also imposed short-term costs: in the first year, employees experienced higher payment delays and greater travel burdens. Third, the reform improved state performance. Teacher retention rose sharply: employees hired in the year before the reform were 27 percentage points less likely to leave, nearly halving attrition relative to controls. Student learning outcomes also improved (standardized

[1] We calculate this using the 2020 harmonized test scores based on the methodology in Angrist et al. (2021).

[2] 2.8% of employees who were paid prior to the reform never registered in the biometric system. Our complementary screening exercise reveals that an additional 5.5% to 11.3% of payroll slots were filled by "stand-ins" —- individuals impersonating legitimate employees during registration to preserve salary lines. See section 4.1 for details.

test scores rose by 0.15σ), and financial inclusion increased as mobile money use expanded (treated employees were 27 percentage points more likely to make transactions, a fourfold increase over the control group). Together, these findings suggest that digital technologies can strengthen fragile governments' ability to manage personnel and deliver services, though not without — sometimes substantial — transitional frictions.

We next examine how employees' expectations about the durability of the Afghan government shaped the success of the reform.[3] Participation imposed real costs on teachers — coordinating enrollment drives, providing biometric data, enduring delays, and trusting that an unproven system would eventually improve salary delivery. These investments only made sense if the state endured. Using a contemporaneous, nationally representative survey that asked citizens who they believed would win the war, we find that employees were significantly more likely to register in areas where citizens expected the Taliban to lose — a relationship that holds when controlling for a variety of potentially confounding factors (e.g., general views of government performance, local security conditions).[4] A one percentage point increase in this belief is associated with a one percentage point increase in registration. The association is also strongest in districts where beliefs were more aligned, suggesting that both confidence in the state's survival and consensus about its future shaped employees' willingness to engage in state-building. This pattern suggests that fragility itself imposes constraints on state development: the absence of consensus regarding state durability reduces complementary investments that make technology effective in building state capacity.

Taken together, this paper contributes to three primary areas of research. First, it speaks to the literature on building state capacity, illustrating how technology can accelerate a natural first step in state-building: identifying and paying employees. A large literature examines how states initially form, build fiscal capacity, improve bureaucratic performance, and deliver services (e.g., Allen, Bertazzini and Heldring, 2023; Dal Bó, Finan and Rossi, 2013; Dincecco, 2015; Xu, 2018; Ashraf et al., 2020; Weigel, 2020; Finan, Olken and Pande, 2017; Balán et al., 2022; Okunogbe and Pouliquen, 2022; Mastrorocco and Teso, 2023; Xu et al., 2023; Aneja and Xu, 2024; Cantoni, Mohr and Weigand, 2024; Chiovelli et al., 2024; Bergeron et al., 2024; Bergeron, Tourek and Weigel, 2024; Besley et al., 2025). Much less is known about how fragile states can use new technologies — especially from experimental

[3]This exercise is similar to that in Callen, Weigel and Yuchtman (2024), where the authors find that Afghans are more likely to use the formal judicial system instead of informal courts when they believe the state will persist.

[4]Our measure of confidence that the Afghan government would prevail over the Taliban should also correlate positively with general regime support. While we cannot fully rule out this confound, we control for several related perceptions including whether citizens believe the government is effective, cares about their needs, is headed in the right direction, and is deserving of confidence, and find that our results remain robust.

work conducted at large scale — to build core administrative systems at early stages of state formation.[5] We intervene at an earlier stage of state building, by using new technologies to strengthen personnel-management systems that link the central government to field-level employees amid political contestation. Such administrative capabilities make citizens legible to the state (Scott, 1998), helping to centralize authority and build fiscal and coercive capacity (North and Thomas, 1973). We also connect to related work on how state-like functions emerge in the absence of uncontested authority (Balán et al., 2022; Sanchez de la Sierra et al., 2022; Haim, Ravanilla and Sexton, 2021; Sanchez de la Sierra, 2020; Henn et al., 2023) and on the relationship between state capacity and nation-building (Rohner and Zhuravskaya, 2023; Bazzi et al., 2019; Cantoni et al., 2017; Bandiera et al., 2019).

Second, we contribute to research on how expectations and uncertainty shape state-building by showing that employees' willingness to engage in state reforms was associated with their beliefs about the government's durability. This finding connects to a literature that views state capacity as a forward-looking investment problem (Besley and Persson, 2010, 2011), which depends on political stability (Besley and Persson, 2009) and reciprocal citizen-state interactions (Besley, 2020). It also speaks to an earlier literature emphasizing that institutions matter not only as formal rules but also as coordination devices that align expectations about how political power will be exercised, constrained, and transferred (North, 1990, 1991; North, Wallis and Weingast, 2009). In this view, institutions coordinate expectations and reduce uncertainty, thereby enabling investment, compliance, and cooperation. Our results underscore that in fragile states, investments in state capacity might face strong headwinds because the state's very survival is in question.

Third, this paper also contributes to the literature on interventions aimed at improving student learning, as systematically reviewed in GEEAP (2024). Within the education literature, our study particularly relates to research on interventions that support teacher retention in high-turnover environments (Morgan et al., 2023; Leaver et al., 2021) and studies that assess the learning costs associated with high teacher turnover (Zeitlin, 2021; Akhtari, Moreira and Trucco, 2022; Ronfeldt, Loeb and Wyckoff, 2013; Gibbons, Scrutinio and Telhaj, 2021). Relatedly, the paper also contributes to the literature on the use of digital technology to enhance financial access, as reviewed in Suri et al. (2023).

The rest of the paper proceeds as follows. Section 2 describes the institutional environment in Afghanistan during our study. Section 3 describes the Afghan Government's flagship Mobile Salary Payments (MSP) reform, our research design, and experimental data sources.

[5]Appendix C provides a systematic and reproducible review of the literature on building state capacity, with a focus on studies using technology in low-capacity and fragile states. The review combines a structured Web of Science search protocol and a large language model–based identification of articles published in 18 leading economics journals and 3 political science journals over the period 2000 - 2025.

Section 4 presents results on ghost employees, salary experience, and downstream outcomes. Section 5 examines the relationship between beliefs about the durability of the state and implementation of the MSP reform. Section 6 concludes.

2 Background

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).[6]

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 troop 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 Appendix Figure A.1.[7] Still, significant uncertainty regarding who would win – and so the institutional trajectory of the country – remained: in May 2018, at the onset of the MSP reform, 47.7% of respondents in the country thought the National Army would defeat the Taliban, 19.5% believed they would lose, and 31.7% 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 the Ministry of Education (MoE) for the scale-up given its central importance to the country's development, large size, and the critical importance to international donors of ensuring their

[6] 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%. 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).

[7] 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.

support to public salaries was well-managed. At the time, the 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).[8]

In theory, digital identification and payment systems appear well-suited to address at least four challenges involved in processing public salaries in fragile states, all of which were relevant in Afghanistan. First, Afghanistan's Tazkira national identification system is highly incomplete and subject to duplicate identifiers.[9] In low-capacity settings, where developing a comprehensive national ID system using a more traditional number-based approach is challenging for a range of reasons, biometric systems could offer an effective alternative by uniquely identifying citizens directly using their biometric data.

Second and related, identifying and removing ghost workers 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.[10]

Third, public salary disbursements in Afghanistan are often delayed for two broad reasons. First, the delay may originate from slow-moving "first-mile" approvals and administrative steps that must be taken in the MoE, the Ministry of Finance, and then by Kabul Bank, all in Kabul. As government salaries provide 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 lack 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 trusted agents designated by the government).[11] 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

[8] 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).

[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] A 2017 independent assessment noted that the MoE's "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% 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" (MEC, 2017).

[11] 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.

to first mile actors and 22.9% to last mile actors.[12].

Fourth, reports of leakage in public salary disbursement are common in fragile countries, particularly when physical cash must be transported. Digital salaries may be harder to intercept during transmission and may also be easier to audit because they leave an electronic record.

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.[13]

3 Research Design and Data

This section describes the research design, the experimental sample, provides details on the reforms, discusses treatment assignment, and describes the administrative and survey data used to evaluate the reform and the importance of beliefs about state durability.

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.[14] In

[12] The last-mile issues, in particular, reflect the substantial travel costs and risks, which have been shown elsewhere to undermine the development of state capacity (Mastrorocco and Teso, 2023; Chiovelli et al., 2024; Besley et al., 2025)

[13] 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.

[14] The zones are constructed using a spatial clustering algorithm that minimizes the distance between schools within a zone while enforcing a minimum size (typically at least 50 employees, whenever feasible). Each school's weight in the clustering corresponds to its number of employees, meaning that larger schools exert more pull in determining cluster boundaries. Clustering is restricted to schools within the same district.

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 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, and as a result our final sample consists of 28,669 employees working in 1,189 schools divided into 345 experimental registration zones and spanning a total of 37 districts.[15] 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 both girls and boys.

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 the teachers national ID numbers and biometric measurements of ten fingers. The fingerprints uniquely identified each employee, preventing the same individual from registering multiple times; fingerprint authentication

Within each resulting cluster, the safest and largest school is designated as the registration center. When none of the schools in a cluster meet the security threshold to serve as a registration center, the entire cluster is merged with the nearest secure zone. Appendix Figure A.2 provides satellite photographs of a typical urban (Panel A) and rural (Panel B) registration zone; urban zones are generally smaller and include fewer schools because schools there are geographically closer together and tend to have more employees, allowing the minimum cluster size to be met within a smaller radius.

[15] 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.

was also required for cash-out services. 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 an 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.

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

3.4 Treatment Compliance and Estimation

Figure 2 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 and July 2018. Most of the registration in the control group took place between November 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 transitioned 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 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. In April 2019 (when our first endline survey occurred), 72% of employees in the EE group, 44% in the ED group, and 8% in the control group reported being paid by mobile money. In May 2020 (our second endline survey), these numbers increased to 81%, 49%, and 33%, respectively.

While compliance with the research design schedule was imperfect, the randomization generated variation in the number of employees impacted by the reform across treatment groups, as displayed in Figure 2 and Appendix Table A.2. 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:

$$\begin{aligned} Y_{izdt} &= \gamma_t + \beta_t^{MSP} MSP_{izdt} + \mu_d + \varepsilon_{izdt} \\ MSP_{izdt} &= \theta_t + \varphi_t^{EE} EE_z + \varphi_t^{ED} ED_z + \eta_d t + \epsilon_{izdt} \end{aligned} \tag{1}$$

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. Coefficients like

β_t^{MSP} carry a time subscript t to indicate that the estimation is done separately by survey wave rather than pooling periods. 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 Ministry of Finance, 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. These visits also involved reviewing the schools' administrative records to attempt to gain basic facts about all employees working in the school.[16]

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.[17] 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

[16] 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.

[17] 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.

of treatment impacts between wave, are likely not contaminated by sample selection bias.[18]

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 representative sample of employees present during audit visits who were confirmed to be employees, absent during those visits, and comprehensively sampled all employees who never registered (see Appendix A 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 ghost workers and stand-ins.

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

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.[19] As such, this question plausibly captures citizens' beliefs about future political institutions.[20] The institutional arrangement if the National Army prevailed would resemble that outlined in the constitution, including a commitment to regular

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

[19]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".

[20]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.

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.[21]

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 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 who had been paid at least once between March 2017 and April 2018 (i.e., in the months before the start of registration and for which we have data) registered during the initial in-person drive. Appendix Table A.5 shows that registration was highest in the EE and ED 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.[22] Appendix Figure A.3 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

[21] In Section 5, we discuss how this measure links to theories that rely on beliefs about future institutions to 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.

[22] Appendix Table A.6 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.7 also reports wildcluster bootstrap p-values; the only meaningful difference relative to Appendix Table A.5 is that the estimate on the "contested district" dummy is no longer statistically significant.

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 ED) zones – where the MNO was most successful 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 (possibly quite) conservative lower bound estimate of the number of ghosts on the payroll prior to the reforms.[23] Their salaries amount to about 3.3 million AFN (43,000 USD 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 the MoE continued to request salaries for employees who never registered.[24] 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% (1.3 million AFN or 17,000 USD) of the total monthly wage bill in the EE and ED zones. Even so, back of the envelope calculations indicate the MSP reform had a return on investment of 50% by year 2.[25]

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

[23]We include in this analysis only employees who were paid in each 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. Additional details on this calculation, as well as calculations where we relax this restriction, are presented in Appendix A.

[24]We cannot verify whether these salaries were actually withheld, as required by the reform, or whether employees who did not register continued to receive payments via bank accounts or the trusted agent system used before the reform. This is because we lack data on salary disbursements through those pre-reform channels and can observe actual disbursements (through mobile money transactions) only for employees who registered for a mobile wallet.

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

working in the school.

As summarized in Appendix Figure A.4, 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 35% 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 5.7 percentage points (s.e.= 2 pp) less likely to score a perfect seven points, 9.6 percentage points (s.e. = 3 pp) less likely to score six, and 19.2 percentage points (s.e. = 4 pp) more likely to score three or less.

In Appendix A, 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 5.7% to 11.7% of registered employees (5.5% to 11.3% of all employees), depending on how strictly we mark the litmus test. This, in turn, indicates that some registrants knew very little about the school where they worked, and so were impersonating legitimate employees. This highlights that while technology can increase the state's ability to track field-level employees, it is not a complete solution.

Adding these potential stand-ins to the 2.8% of unregistered employees, we estimate that between 8.3% and 14.1% of all paid employees might be ghosts. If they were all removed from the payroll, the government would save 18 million AFN (0.24 million USD) 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 11.8% (median = 9.8%) of teachers are ghosts (standard deviation = 15.0 pp).

This estimate is within the above estimated range of 8.3% to 14.1% of all employees.[26]

4.2 Mobile Salary Payment Impacts on Employees

The second component of the MSP reform shifted teacher salary payments from a hybrid system that relied on banks in cities and trusted agents in rural areas to an integrated mobile money platform.[27] Policymakers hoped that digital payments would reduce leakage and delays, while also expanding access to digital wallets and mobile transfers. In this subsection, we report impacts on these outcomes.

Leakage was a salient issue, particularly because of potential problems related to physically transporting cash (Special Inspector General for Afghanistan Reconstruction, 2022). In practice, however, our data reveal little evidence of leakage both before and after the reform. On average, control group employees report paying less than 19 AFN (0.23 USD) 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 1 column 1, panel B).

The reform was also intended to reduce disbursement delays. These result from a combination of first-mile delays in the capital due to bureaucratic approval processes at the MoE, Ministry of Finance, and Kabul Bank, as well as last-mile delays in getting the final payment to the employee.[28]

Over the two years of the reform, delays came down substantially in both treatment and control groups. Before the reform, 60% of employees in the control group and 63% of employees in the treatment group experienced delays (Appendix Table A.1). After one year, 44% of control employees experienced delays with treatment increasing delays by a substantial 39 percentage points (Table 1 column 2, panel A). After two years, delays fell further in the control group to 28% of employees, with treatment having a modest and insignificant reduction of 5 percentage points (Table 1 column 2, panel B).

Further analysis suggests that the reductions in delays – for both the treatment and control groups – are largely driven by improvements in centralized bureaucratic first-mile obstacles to salary delivery. First-mile delays fell more in the treatment group in the first year of the reform, suggesting that the Treasury Department initially focused on improving

[26] Interestingly, the estimates of senior government officials (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%).

[27] In Appendix Table A.8 we check whether registration *per se* impacted the outcomes discussed in this section, and do not find evidence that it did.

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

salary approval protocols for treated employees (see Appendix Figure A.5). While the reform surfaced the issue of first-mile delays, the fact that they also fell in the control group suggests that the experiment itself helped draw ministry attention to the problem, prompting broader process improvements that did not necessarily depend on the introduction of new payment technologies. By contrast, the payment reform initially *increased* last-mile delays in the treated group, 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.[29] However, by the end of the second year, delays were reduced substantially for all employees. 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 1, column 3). Thus, although the reform reduced first-mile delays more for treated employees than for controls, the broader pattern – including some immediate reduction in first-mile delays for control zones – suggests that the observed reductions in delays are not necessarily driven by the technology *per se*. We therefore emphasize these delay results as informative about implementation dynamics, but do not view them as central to our conclusions about how the technology affects state capacity.

For the reform to succeed, it needed the support of employees, particularly given the opposition from those benefiting from the existing system. The government was especially focused on whether employees would vote to expand the reform nationwide.[30] 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 (60% of employees experienced delays at baseline; see Table 1, column 2). However, as shown in Table 1 (Panel A, column 4), support for the reform dropped during the first year of the reform due to the implementation issues discussed earlier.[31] As these initial problems were resolved, support for the reform rebounded (Panel B, column 4). First-hand experience with mobile salary payments increased support, particularly in the second year as the mobile network operator increased capacity.

A final objective of the reform was to expand financial inclusion by increasing employees' access to mobile wallets and their use of mobile transfers. To measure usage, we asked

[29] Appendix Figure A.6 highlights that there were fewer active mobile money agents when the program launched, and many of those agents had very little prior experience.

[30] 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.

[31] As can be seen in Appendix Figure A.7, the most common reasons to support or oppose the reform were issues of accessibility and timeliness, highlighting the central importance of program implementation.

employees whether they had sent money to or received money from a relative or friend through a mobile phone in the previous year. The share doing so was 7% in control zones, and the treatment increased this by 27 percentage points after the first year (Panel A, column 5).[32] These survey results are consistent with patterns observed in administrative mobile money transaction data, which we discuss in Section 4.5 below.

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.[33] 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.[34]

In this regard, the results are clear: the reform ultimately proceeded much more effectively in cities. In Table 2, we find that in the first year, the reform had modest 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. Yet, support for expanding the reform increased in urban areas, where the reform was showing signs of improvement even by the end of the first year (Panel A, column 4), and mobile money transactions (excluding the salary transfers themselves) were up everywhere (column 5).[35]

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, there was substantial support to scale the reform once again throughout the sample (Panel B, column 4).

Heterogeneity by Government and Insurgent Control In the Appendix, we examine the effectiveness of the reform in government-controlled and insurgent-controlled areas. We

[32]This question was not included in the phone survey conducted after the second year of the reform.

[33]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.

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

[35]Appendix Figure A.8 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.

measure control using assessments provided by the Long War Journal. The results are presented in Appendix Table A.9. We find that support for the reform increases in areas under government control in year 1 and that employees in these areas increase their use of mobile money. In year 2, delays are reduced in areas under government control and elevated in contested areas, mirroring our result on urban versus rural heterogeneity just discussed.

4.3 Impacts on Employee Turnover

The MoE faced significant turnover, particularly among recently hired employees. Departure rates surged during periods when the Taliban made significant advances against the government, suggesting that some teachers feared retaliation in the event of a Taliban victory. A substantial body of research shows that such turnover is costly for organizations, especially in the context of educational institutions where it can severely impact learning outcomes (Hanushek, Rivkin and Schiman, 2016; Zeitlin, 2021; Akhtari, Moreira and Trucco, 2022; Ronfeldt, Loeb and Wyckoff, 2013; Gibbons, Scrutinio and Telhaj, 2021).

The MSP reform could have influenced turnover through several channels. First, in the baseline survey, conducted just before registration began, nearly all employees expressed enthusiasm for the reform and frustration with the existing payment system. Registration was a highly visible government effort, signaling its intent to improve a process that was vital to employees – countering the widespread perception that the government was both ineffective and indifferent to their needs. Second, the swift reduction in initial delays was immediately noticeable to employees, even though they continued to face challenges in finding agents to access their deposited funds. This visible progress likely led employees to reasonably (and correctly) expect further improvements in salary payments over time. Third, the reform indicated a commitment from the government to provide teachers with access to financial services, a benefit many would otherwise have lacked. We therefore consider turnover as a metric of state performance.

The reform reduced turnover among new employees, due both to higher retention rates for employees who joined the MoE in the months preceding the reform and to less hiring of new employees right after the reform.

In column 1 of Table 3, we classify employees as having left the MoE if they are not paid for at least three consecutive months and do not re-appear in the payroll thereafter and find evidence of higher retention rates. We look at employees who joined the MoE in the 12 months preceding the reform, and find that retention rates are the lowest for this category: a remarkable 55% of recently hired employees left the MoE in the first year of the reform in the control group. In Panel A, where as above we instrument being paid through mobile money with the randomly assigned treatment status, we find that the reform reduced the share of recently hired employees leaving by 27 percentage points ($p < 0.1$). In Panel B, where we report reduced form results, we also find higher retention of recently hired employees in treatment zones, especially in ED zones. Columns 2 and 3 show that both retention rates in the control group, and the effects of the reform, are relatively similar in urban and rural areas. In Appendix Table A.10 we show that retention rates are higher among employees with longer employment histories at the MoE than among recently hired employees, with the share of people leaving in the first year of the reform being on average 15%. Correspondingly, we do not identify significant effects of the reform on retention rates for employees who joined the MoE more than 12 months before the reform.

In column 4 of Table 3, we look at the number of employees who joined the MoE after the beginning of the reform as a share of all employees and find evidence of lower hiring rates of new employees in treatment zones. In the control group, 18% of all employees joined the MoE after the reform had started. Panel A shows that the reform reduced the share of newly hired employees by 4 percentage points ($p < 0.05$). In Panel B, we find that the share of newly joined employees is lower by 3 percentage points ($p < 0.05$) in EE zones and 4 percentage points ($p < 0.01$) in ED zones. Columns 5 and 6 show that the effects of the reform are more than twice as large in urban areas with respect to rural areas.

4.4 Impacts on Student Learning

The ultimate objective of the reform was to improve student learning by modernizing the tracking and payment of employees. 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).[36] 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.[37] Across the sample, combined learning scores increase by 0.15σ ($p < 0.1$;

[36]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.11, 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.

[37]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

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 0.3σ ($p < 0.01$), and reading scores increase by 0.3σ ($p < 0.05$). 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 are realized in a single year? Estimates in Table 1 and 2 indicate that treated employees are more supportive of the reform. Estimates in Table 3 show less turnover of employees in treatment zones, due both to higher retention rates among employees who joined the MoE in the months preceding the reform and to less hiring of new employees right after the reform. This, in turn, is reflected in employees in treatment zones having worked for about 0.5 years longer relative to a control mean of about 10 years. While not by any means dispositive, this pattern is consistent with greater enthusiasm among recently hired employees, hopeful the reform would eventually lead to improvements, who decided to stay at the MoE. They then accumulated more experience with respect to the new employees who would have been hired otherwise after the reform, which could have led to better learning outcomes. 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 and A.13).

These results link to the literature on improving education in developing countries. The focus on measurement and credible evaluation in development economics has transformed our understanding of the severe learning gaps, now recognized as a learning crisis, and has generated a wealth of evidence on effective ways to improve educational outcomes (Angrist et al., 2020; Evans and Yuan, 2022; Banerjee et al., 2023*b*). Key studies include evaluations of teacher incentives (Loyalka et al., 2019; Mbiti et al., 2019; Behrman et al., 2015), increased funding (De Ree et al., 2018; Andrabi et al., 2024), contract hiring and facilitators (Duflo, Dupas and Kremer, 2015; Muralidharan and Sundararaman, 2011), and standardizing educational processes (Gray-Lobe et al., 2022). Relevant to our findings, teacher turnover is particularly costly as teachers tend to improve significantly during their initial years (Bau and Das, 2020). Additionally, psychosocial interventions that enhance teachers' sense of control can boost effort and learning outcomes (Kaur, 2023). Our results provide evidence that, at early stages of state development, systems which improve the efficiency of salary

mobile money prior to the learning assessment, and we instrument this with the school's randomly assigned treatment status.

payments can both save the government money and simultaneously reduce turnover and improve learning outcomes.

4.5 Impacts on Employee Financial Inclusion

The MSP reform also 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 percentage points from a base of 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:[38]

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

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 variables specification (Panel B) in which we instrument for $Months_{izd}$ using the randomly assigned treatment dummies.[39]

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

[38] 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.

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

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 seven 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).[40]

Taken together, these patterns – including increases in airtime transfers – are consistent with the idea that once government salary payments were delivered digitally, broader use of mobile-based financial services could begin to spread through routine interactions, as anticipated by the network-effects logic motivating the reform.

5 Does State Fragility Deter State Capacity?

5.1 Fragility, Capacity, and the Role of Beliefs

The reform we study was implemented by a government facing extreme risk that wanted to increase its capacity.[41] However, building state capacity is an investment with delayed benefits.

As such, the decision to make such investments may depend on beliefs about the state's durability.[42] If the state faces a risk of radical transformation, this uncertainty will manifest in 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 stakeholders in the state, such as MoE employees, trust that the state will endure and are confident that others share this belief, coordinated investments – like the major reform we

[40]Appendix Table A.14 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.

[41]Definitions of state fragility often highlight systemic risks such as potential state failure, violent displacement of the ruling regime (Besley and Persson, 2010, 2011), 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).

[42]This logic applies both to the government and to state employees (and citizens more broadly). The government must decide whether to undertake the effort to launch a reform that requires time, resources, and political capital; employees must then decide whether to invest in that reform by participating and bearing the associated coordination and adjustment costs. Such uncertainty can therefore shape both the government's decision to initiate a reform and individuals' decisions to invest once it is underway. Our analysis focuses on the latter: the investments employees make after the government has committed to the reform.

study – are more likely to succeed. Conversely, disagreement about the state's future could deter investments in state capabilities.

Thus, the coordination of investments in state capacity by state actors 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 beliefs. First, whether teachers register for the system depends on their own current perceptions of the Afghan government's durability. Second, this decision is also 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 Appendix Table A.5. 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' beliefs about the durability of the Afghan state.[43]

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

5.3 Measuring Beliefs

To proxy for beliefs about state durability, we use a survey question about who would win the war, discussed in Section 3.6.[44] We proxy for employees' beliefs with 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.[45] Furthermore, we proxy for teachers' beliefs about the dispersion of views among teachers in their district by using the standard deviation of beliefs across respondents in their district. There are a substantial number of respondents in each district-year.[46]

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 Appendix Figure A.1), consistent with when these measures were taken. Second, these beliefs respond to news shocks (Appendix Figure A.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.10). Fourth, these measures predict the likelihood of citizens using formal state structures instead of informal institutions for dispute resolution (Callen et al., 2023).[47]

5.4 Analysis

We use our data to conduct a large-scale empirical test of whether state actors' beliefs about institutional durability predict participation in a state-building reform. Specifically, we examine whether beliefs about who would win the war predict employee enrollment by

[44] An ideal design would directly and incentive-compatibly elicit these beliefs.

[45] Appendix Figure A.9 shows that the belief distribution for government employees in our experimental region in 2018 is comparable to that of the general population.

[46] 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 their own uncertainty. The standard deviation reflects disagreement between citizens. In Appendix Table A.15, we analyze the average response, the implied average individual-level confidence, and the standard deviation among respondents, finding the reform most effective where citizens were confident, certain, and agreed on the state's durability.

[47] 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.

estimating the following 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 (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 Appendix Table A.5.

In practice, reforms like biometric registration require that employees physically present themselves at specific locations and times – an act that entails risk, cost, and uncertainty. If employees do not expect others to participate, or doubt the state's survival, they may choose to abstain. But if they believe the reform is likely to succeed and is backed by a durable regime, they are more likely to act. To test the importance of consensus of beliefs about durability, we interact the mean belief with the within-district standard deviation of responses to the five-point Likert scale question.[48]

Table 6 presents the corresponding estimates. A higher share of respondents expressing confidence that the Taliban would lose is positively associated with employee appearance at the registration drive, even after adjusting for covariates such as controls for regime support (columns 1, 3, and 4). Similar measures of beliefs from prior years do not evidently 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 on confidence that the state will endure, but also relative consensus about that fact. We test this in columns 5, 6, and 7 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 3. This test should be interpreted with caution: if beliefs are measured with more error where there is greater within-district variation, then this same pattern could arise from attenuation bias.[49]

[48] 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.

[49] 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 nec-

Estimates reported in Table 6 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 rents). In Appendix Table A.16 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 we obtain a similar set of results.[50]

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 in Table 6 are not causally identified, as we lack exogenous variation in beliefs. For example, confidence that the Taliban would lose the war could reflect regime support, as well as expectations about state survival. That said, we observe other patterns that are consistent with a causal interpretation – our preferred, though necessarily qualified, reading. First, beliefs measured contemporaneously with the reform predict registration, whereas beliefs from prior years do so much less. Second, the relationship is robust to controlling for a broad range of potential confounds, including beliefs about state effectiveness, proxies of support for the regime, the presence of security forces, and other covariates that predict registration.

The war unfolded very differently across provinces. Kandahar was – and remains – the logistical and military heart of the Taliban movement: while its rural areas were consistently under Taliban control, the provincial capital hosted one of NATO's largest airbases. Nangarhar followed a similar pattern, with the government holding Jalalabad city while the surrounding countryside remained contested. In contrast, Parwan – located adjacent to the major U.S. airbase at Bagram and Kabul city – faced less Taliban pressure. These differences are reflected in the distribution of beliefs shown in Figure A.11. While 93% of respondents across our three study provinces believed the Taliban would lose, belief levels varied sub-

essarily extend across different employees. In Appendix Table A.15, 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

[50]In Appendix Tables A.17, A.18, and A.19, for completeness, we check for heterogeneous treatment impacts by splitting the sample at the median of beliefs that the Afghan government would win the war, examining impacts on payment experience, student learning, and mobile money use. These outcomes depend on multiple contextual factors – such as baseline service quality, Taliban presence, and mobile infrastructure – which vary with beliefs and affect reform feasibility. As such, we do not interpret these as tests of the hypothesis that beliefs shape investment in state capacity.

stantially between provinces: 88% of respondents in Kandahar believed the Taliban would lose, compared to 94% in Parwan, and 96% in Nangarhar. A variance decomposition shows that more than half of the total variation in beliefs arises between provinces rather than within them. Consistent with this, the positive relationship between beliefs and registration disappears once province fixed effects are included (Appendix Table A.20).

Taken together, these patterns suggest that fragile states may face a dual constraint in state-building: not only must governments implement reforms, but they must do so in environments where belief in the state's durability is often contested. Building capacity under such conditions requires both technical systems and shared expectations of survival.

6 Conclusion

The world's most fragile states are also its least capable — unable to deliver basic services, provide security, or perform core administrative functions. In part because of this lack of capacity, they are also among the world's poorest. Accelerating the development of state capacity in fragile states is therefore critical to sustaining progress on poverty reduction and other global development goals.

New technologies — particularly those that have been successfully deployed in more stable developing countries — may offer a promising path forward (Madon et al., 2023). A growing body of evidence shows how technology can improve administrative capacity and service delivery by enabling faster information flows, improving identification, and reducing transaction costs (e.g., Muralidharan, Niehaus and Sukhtankar, 2016; Dhaliwal and Hanna, 2017; Banerjee et al., 2020; Dodge et al., 2025). Much less is known, however, about whether similar technologies can work effectively in fragile states.

We provide experimental evidence from Afghanistan that biometric registration and mobile salary payments strengthened the government's core administrative ability to identify and pay employees. The reform also reduced ghost workers, improved teacher retention, modestly raised student learning, and expanded access to formal financial services. The program was far from perfect: some employees impersonated legitimate staff to avoid being taken off the payroll; payment delays initially worsened before improving; and identifying ghost workers did not always lead to their removal. Despite these frictions, the reform increased one of the world's least effective state's ability to monitor and manage its workforce under extraordinarily challenging conditions. The fiscal savings, roughly one percent of the payroll, were sufficient to cover the cost of implementing the program.

We also find that civil servants were more likely to participate in the reforms in areas where citizens expressed greater confidence in, and agreement about, the government's prospects of survival. When doubts about the state's endurance were stronger, participation was lower. This suggests that fragility constrains capacity in part through expectations: when the state's survival is uncertain, individuals have weaker incentives to invest time and effort in building it. In our setting, this investment involved up-front coordination and adoption of new systems whose benefits would materialize only if the government endured. In this sense, fragility can undermine the very purpose of institutions which, as North (1991) argued, are "devised by human beings to create order and reduce uncertainty in exchange."

More broadly, our study shows that new digital technologies can accelerate efforts to build state capacity, even under extreme instability. The context was unusual — marked by an ongoing insurgency and extensive international assistance. Yet precisely because its foundational institutions were in flux, it offers a rare window into the dynamics of state formation. Our results suggest that fragility generates headwinds not only through material constraints but also by eroding confidence in the state's survival. This supports theories that view state-building as an intertemporal investment problem, in which expectations about institutional durability shape behavior. In this view, beliefs about regime survival are not merely symptoms of fragility; they are themselves determinants of institutional development. For institutions to foster capacity, they must anchor the expectations of citizens and public servants, enabling them to invest in the state's future.

References

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

Akhtari, Mitra, Diana Moreira, and Laura Trucco. 2022. “Political Turnover, Bureaucratic Turnover, and the Quality of Public Services.” *American Economic Review*, 112(2): 442–493.

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

Allie, Feyaad. 2023. “Facial recognition technology and voter turnout.” *The Journal of Politics*, 85(1): 328–333.

Andrabi, Tahir, Natalie Bau, Jishnu Das, Naureen Karachiwalla, and Asim Ijaz Khwaja. 2024. “Crowding in Private Quality: The Equilibrium Effects of Public Spending in Education.” *The Quarterly Journal of Economics*, qjae014.

Aneja, Abhay, and Guo Xu. 2024. “Strengthening state capacity: Civil service reform and public sector performance during the gilded age.” *American Economic Review*, 114(8): 2352–2387.

Angrist, Noam, David K. Evans, Deon Filmer, Rachel Glennerster, F. Halsey Rogers, and Shwetlena Sabarwal. 2020. “How to Improve Education Outcomes Most Efficiently?: A Comparison of 150 Interventions Using the New Learning-Adjusted Years of Schooling Metric.” Center for Global Development.

Angrist, Noam, Simeon Djankov, Pinelopi K Goldberg, and Harry A Patrinos. 2021. “Measuring human capital using global learning data.” *Nature*, 592(7854): 403–408.

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

Assunção, Juliano, Clarissa Gandour, and Romero Rocha. 2023. “DETER-ing deforestation in the Amazon: environmental monitoring and law enforcement.” *American Economic Journal: Applied Economics*, 15(2): 125–156.

Axbard, Sebastian, and Zichen Deng. 2024. “Informed enforcement: Lessons from pollution monitoring in china.” *American Economic Journal: Applied Economics*, 16(1): 213–252.

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

Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo. 2019. “Nation-building through compulsory schooling during the age of mass migration.” *The Economic Journal*, 129(617): 62–109.

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.

Banerjee, Abhijit, Rema Hanna, Benjamin A Olken, Elan Satriawan, and Sudarno Sumarto. 2023*a*. "Electronic food vouchers: Evidence from an at-scale experiment in Indonesia." *American Economic Review*, 113(2): 514–547.

Banerjee, Abhijit, Tahir Andrabi, Rukmini Banerji, Susan Dynarski, Rachel Glennerster, Sally Grantham-McGregor, Karthik Muralidharan, Benjamin Piper, Jaime Saavedra, Hirokazu Yoshikawa, Kwame Akyeampong, Sara Ruto, and Sylvia Schmelkes. 2023*b*. "Cost-effective Approaches to Improve Global Learning: What Does Recent Evidence Tell Us are "Smart Buys" for Improving Learning in Low-and Middle-income Countries?" Global Education Evidence Advisory Panel.

Barnwal, Prabhat. 2024. "Curbing leakage in public programs: Evidence from India's direct benefit transfer policy." *American Economic Review*, 114(12): 3812–3846.

Bau, Natalie, and Jishnu Das. 2020. "Teacher Value Added in a Low-Income Country." *American Economic Journal: Economic Policy*, 12(1): 62–96.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, and Maisy Wong. 2019. "Unity in diversity? How intergroup contact can foster nation building." *American Economic Review*, 109(11): 3978–4025.

Behrman, Jere R, Susan W. Parker, Petra E. Todd, and Kenneth I. Wolpin. 2015. "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools." *Journal of Political Economy*, 123(2): 325–364.

Bellon, Matthieu, Era Dabla-Norris, Salma Khalid, and Frederico Lima. 2022. "Digitalization to improve tax compliance: Evidence from VAT e-Invoicing in Peru." *Journal of Public Economics*, 210: 104661.

Beraja, Martin, Andrew Kao, David Y Yang, and Noam Yuchtman. 2023. "AI-tocracy." *The Quarterly Journal of Economics*, 138(3): 1349–1402.

Bergeron, Augustin, Elie Kabue Ngindu, Gabriel Tourek, and Jonathan L Weigel. 2024. "Does Collecting Taxes Erode the Accountability of Informal Leaders? Evidence from the DRC."

Bergeron, Augustin, Gabriel Tourek, and Jonathan L Weigel. 2024. "The state capacity ceiling on tax rates: Evidence from randomized tax abatements in the drc." *Econometrica*, 92(4): 1163–1193.

Besley, Timothy. 2020. "State capacity, reciprocity, and the social contract." *Econometrica*, 88(4): 1307–1335.

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. *Pillars of prosperity.* Princeton University Press.

Besley, Timothy, Dan Bogart, Jonathan Chapman, and Nuno Palma. 2025. "Justices of the Peace: Legal Foundations of the Industrial Revolution." Arthur Lewis Lab, The University of Manchester.

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).

Borcan, Oana, Mikael Lindahl, and Andreea Mitrut. 2017. "Fighting corruption in education: What works and who benefits?" *American Economic Journal: Economic Policy*, 9(1): 180–209.

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

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

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

Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan, and Arman Rezaee. 2020. “Data and policy decisions: Experimental evidence from Pakistan.” *Journal of Development Economics*, 146: 102523.

Callen, M, J Weigel, and N Yuchtman. 2024. “Experiments about institutions.” *Annual Review of Economics.*

Cantoni, Davide, Cathrin Mohr, and Matthias Weigand. 2024. “The rise of fiscal capacity: Administration and state consolidation in the Holy Roman Empire.” *Econometrica*, 92(5): 1439–1472.

Cantoni, Davide, Yuyu Chen, David Y Yang, Noam Yuchtman, and Y Jane Zhang. 2017. “Curriculum and ideology.” *Journal of political economy*, 125(2): 338–392.

Carney, Kevin, Leah R Rosenzweig, Wendy N Wong, Florence Akech, James Otieno, and Elisa M Maffioli. 2025. “Can SMS interventions increase vaccination? Evidence from the COVID-19 pandemic in Kenya.” *Journal of Development Economics*, 174: 103469.

Chiovelli, Giorgio, Leopoldo Fergusson, Luis R Martínez, Juan David Torres, and Felipe Valencia Caicedo. 2024. “Bourbon reforms and state capacity in the Spanish Empire.” Universidad de los Andes, Facultad de Economía, CEDE.

Cohen, Isabelle. 2024. “Technology and the state: Building capacity to tax via text.” *Journal of Public Economics*, 236: 105154.

Conover, Emily, Daniel Kraynak, and Prakarsh Singh. 2023. “The effect of traffic cameras on police effort: Evidence from India.” *Journal of Development Economics*, 160: 102953.

Dal Bó, Ernesto, Frederico Finan, Nicholas Y Li, and Laura Schechter. 2021. “Information technology and government decentralization: Experimental evidence from paraguay.” *Econometrica*, 89(2): 677–701.

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

Debnath, Sisir, Mrithyunjayan Nilayamgode, and Sheetal Sekhri. 2023. “Information Bypass: Using Low-cost technological innovations to curb leakages in welfare programs.” *Journal of Development Economics*, 164: 103137.

Deininger, Klaus, and Aparajita Goyal. 2012. “Going digital: Credit effects of land registry computerization in India.” *Journal of development Economics*, 99(2): 236–243.

De Ree, Joppe, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers. 2018. “Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia.” *The Quarterly Journal of Economics*, 133(2): 993–1039.

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.

Dodge, Eric, Yusuf Neggers, Rohini Pande, and Charity M Troyer Moore. 2025. "From Delay to Payday: Easing Bureaucrat Access to Implementation Information Strengthens Social Protection Delivery." National Bureau of Economic Research.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. "School Governance, Teacher Incentives, and Pupil–Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." *Journal of Public Economics*, 123: 92–110.

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.

Evans, David K, and Fei Yuan. 2022. "How Big Are Effect Sizes in International Education Studies?" *Educational Evaluation and Policy Analysis*, 44(3): 532–540.

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

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

Fujiwara, Thomas. 2015. "Voting technology, political responsiveness, and infant health: Evidence from Brazil." *Econometrica*, 83(2): 423–464.

Gallego, Jorge, and Federico Ortega. 2022. "Can Facebook ads and email messages increase fiscal capacity? Experimental evidence from Venezuela." *Economic Development and Cultural Change*, 70(4): 1531–1563.

Gallegos, Sebastian, Benjamin Roseth, Ana Cuesta, and Mario Sánchez. 2023. "Increasing the take-up of public health services: An at-scale experiment on digital government." *Journal of Public Economics*, 227: 104975.

Gibbons, Stephen, Vincenzo Scrutinio, and Shqiponja Telhaj. 2021. "Teacher Turnover: Effects, Mechanisms and Organisational Responses." *Labour Economics*, 73: 102079.

Gray-Lobe, Guthrie, Anthony Keats, Michael Kremer, Isaac Mbiti, and Owen W. Ozier. 2022. "Can Education Be Standardized? Evidence from Kenya." University of Chicago, Becker Friedman Institute for Economics Working Paper 2022-68.

Greenstone, Michael, Guojun He, Ruixue Jia, and Tong Liu. 2022. "Can technology solve the principal-agent problem? Evidence from China's war on air pollution." *American Economic Review: Insights*, 4(1): 54–70.

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.

Hanushek, Eric A., Steven G. Rivkin, and Jeffrey C. Schiman. 2016. "Dynamic effects of teacher turnover on the quality of instruction." *Economics of Education Review*, 55: 132–148.

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*.

Independent Joint Anti-Corruption Monitoring and Evaluation Committee (MEC). 2017. *Ministry-wide Vulnerability to Corruption Assessment of the Ministry of Education.* Kabul, Afghanistan.

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

Kaur, Jalnidh. 2023. "How Much Do I Matter? Teacher Self-Beliefs, Effort, and Student Learning." Working Paper.

Kochar, Anjini. 2018. "Branchless banking: Evaluating the doorstep delivery of financial services in rural India." *Journal of Development Economics*, 135: 160–175.

Kotsogiannis, Christos, Luca Salvadori, John Karangwa, and Innocente Murasi. 2025. "E-invoicing, tax audits and VAT compliance." *Journal of development economics*, 172: 103403.

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

Leaver, Clare, Owen Ozier, Pieter Serneels, and Andrew Zeitlin. 2021. "Recruitment, Effort, and Retention Effects of Performance Contracts for Civil Servants: Experimental Evidence from Rwandan Primary Schools." *American Economic Review*, 111(7): 2213–2246.

Lewis-Faupel, Sean, Yusuf Neggers, Benjamin A Olken, and Rohini Pande. 2016. "Can electronic procurement improve infrastructure provision? Evidence from public works in India and Indonesia." *American Economic Journal: Economic Policy*, 8(3): 258–283.

Loyalka, Prashant, Sean Sylvia, Chengfang Liu, James Chu, and Yaojiang Shi. 2019. "Pay by Design: Teacher Performance Pay Design and the Distribution of Student Achievement." *Journal of Labor Economics*, 37(3): 621–662.

Madon, Temina, Ashok J Gadgil, Richard Anderson, Lorenzo Casaburi, Kenneth Lee, and Arman Rezaee. 2023. *Introduction to development engineering: A framework with applications from the field.* Springer Nature.

Mastrorocco, Nicola, and Edoardo Teso. 2023. "State capacity as an organizational problem. Evidence from the growth of the US state over 100 years." National Bureau of Economic Research.

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

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

Morgan, Andrew J, Minh Nguyen, Eric A. Hanushek, Ben Ost, and Steven G. Rivkin. 2023. "Attracting and Retaining Highly Effective Educators in Hard-to-Staff Schools." National Bureau of Economic Research.

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

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

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2020. "Identity Verification Standards in Welfare Programs: Experimental Evidence from India." *NBER Working Paper*.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2023. "General equilibrium effects of (improving) public employment programs: Experimental evidence from India." *Econometrica*, 91(4): 1261–1295.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2025. "Identity verification standards in welfare programs: Experimental evidence from India." *Review of Economics and Statistics*, 107(2): 372–392.

Muralidharan, Karthik, Paul Niehaus, Sandip Sukhtankar, and Jeffrey Weaver. 2021. "Improving last-mile service delivery using phone-based monitoring." *American Economic Journal: Applied Economics*, 13(2): 52–82.

North, Douglass C. 1990. *Institutions, institutional change and economic performance.* Cambridge university press.

North, Douglass C. 1991. "Institutions." *Journal of Economic Perspectives*, 5(1): 97–112.

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

North, Douglass Cecil, John Joseph Wallis, and Barry R Weingast. 2009. *Violence and social orders: A conceptual framework for interpreting recorded human history.* Cambridge University Press.

Okunogbe, Oyebola, and Victor Pouliquen. 2022. "Technology, taxation, and corruption: evidence from the introduction of electronic tax filing." *American Economic Journal: Economic Policy*, 14(1): 341–372.

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.*

Rohner, Dominic, and Ekaterina Zhuravskaya. 2023. *Nation building: Big lessons from successes and failures.* CEPR Press.

Ronfeldt, Matthew, Susanna Loeb, and James Wyckoff. 2013. "How Teacher Turnover Harms Student Achievement." *American Educational Research Journal*, 50(1): 4–36.

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."

Suri, Tavneet, Jenny Aker, Catia Batista, Michael Callen, Tarek Ghani, William Jack, Leora Klapper, Emma Riley, Simone Schaner, and Sandip Sukhtankar. 2023. "Mobile money." *VoxDevLit*, 2(2): 3.

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.

Tohari, Achmad, Christopher Parsons, and Anu Rammohan. 2019. "Targeting poverty under complementarities: Evidence from Indonesia's unified targeting system." *Journal of Development Economics*, 140: 127–144.

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." *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.

Xu, Guo. 2018. "The costs of patronage: Evidence from the british empire." *American Economic Review*, 108(11): 3170–3198.

Xu, Guo, Erika Deserranno, Diana Moreira, and Edoardo Teso. 2023. "Bureaucracy." *VoxDevLit*, 8: 3.

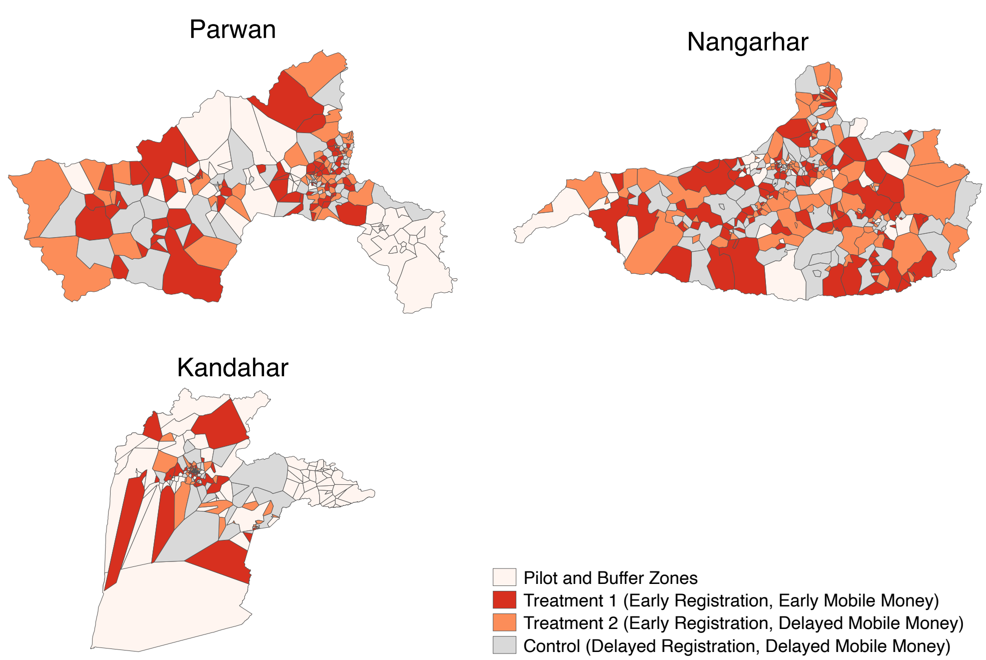
Xu, Xu. 2021. "To repress or to co-opt? Authoritarian control in the age of digital surveillance." *American Journal of Political Science*, 65(2): 309–325.

Yang, Lin, Yatang Lin, Jin Wang, and Fangyuan Peng. 2024. "Achieving Air Pollution Control Targets with Technology-Aided Monitoring: Better Enforcement or Localized Efforts?" *American Economic Journal: Economic Policy*, 16(4): 280–315.

Zeitlin, Andrew. 2021. "Teacher Turnover in Rwanda." *Journal of African Economies*, 30(1): 81–102.

Figures and Tables

Panel A: Studied Provinces of Afghanistan



Panel B: Registration Zones and Treatment Status

Figure 1: Map of Provinces and Registration Zones

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

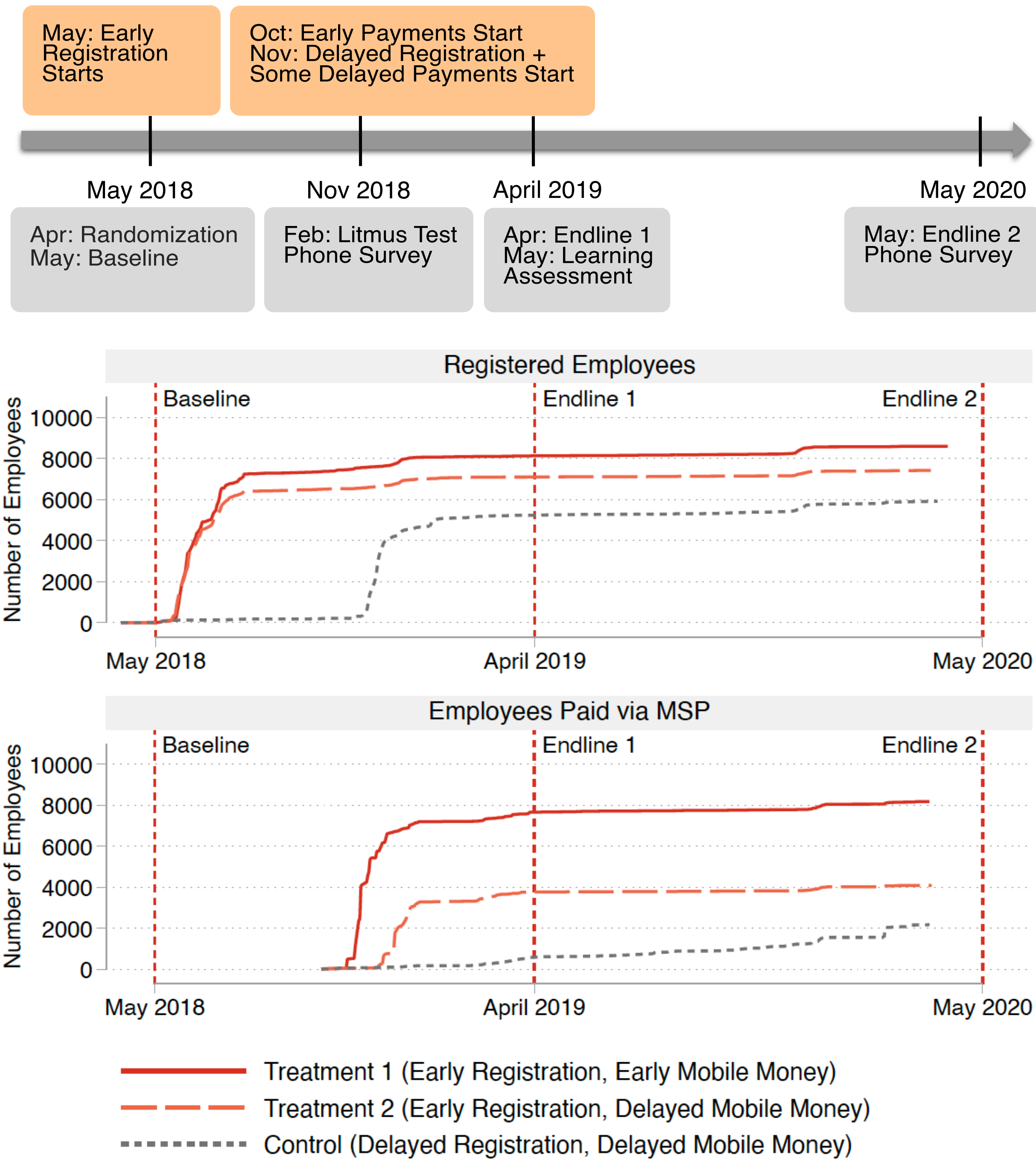


Figure 2: Project Implementation Timeline and Treatment Compliance

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

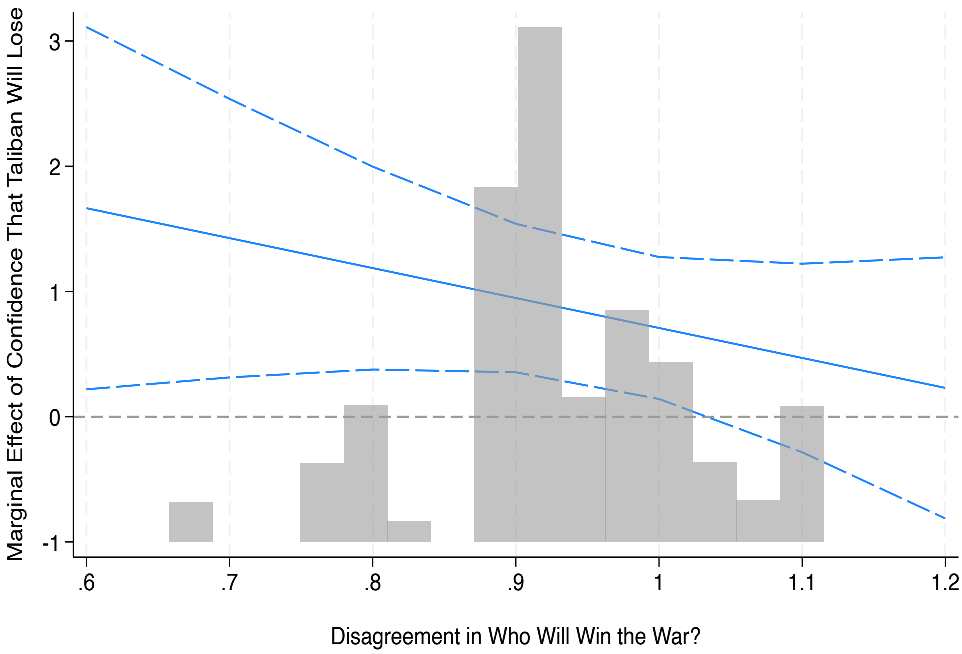


Figure 3: The Relationship Between Confidence and Registration by the Degree of Consensus

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: Estimates of Impact on Employees' Payment Experience

	Payment to Receive Transfer (1)	Experienced Delay (2)	Travel Time to Cash-Out Salary (Min) (3)	Vote in Favor of MSP (4)	Conducted Mobile Money Transfer (5)
Panel A. Year 1 Outcomes					
Mobile Salary Payment (=1)	−7.925	0.386***	39.839***	0.114*	0.272***
	(4.921)	(0.076)	(9.400)	(0.068)	(0.048)
Observations	838	848	810	829	844
R squared	0.249	0.251	0.284	0.343	0.161
Control Mean (Year 1)	19.094	0.441	31.219	0.461	0.067
Control Mean (Baseline)	23	0.596	37.449	0.980	
# Reg. Zones	310	310	303	308	310
Panel B. Year 2 Outcomes					
Mobile Salary Payment (=1)	−18.620***	−0.054	2.169	0.251***	
	(3.827)	(0.083)	(5.934)	(0.083)	
Observations	644	642	632	623	
R squared	0.248	0.134	0.118	0.287	
Control Mean (Year 2)	10.178	0.284	36.629	0.710	
# Reg. Zones	286	287	283	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.24. Payment to Receive measures what respondents report paying to receive their salary in Afghanis. For comparison, the average (net) salary in the sample is 8,560 AFN 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. Vote in Favor of MSP is a dummy variable equal to one if the respondent indicates they would support scaling the reform across the Ministry of Education. Conducted Mobile Money Transfer is a dummy variable equal to one if the respondent indicates making a mobile money transfer to someone else in the previous month. This outcome is not recorded in year 2 because the survey was abbreviated due to the pandemic. Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 2: Estimates of 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)	Vote in Favor of MSP (4)	Conducted Mobile Money Transfer (5)
Panel A. Year 1 Outcomes					
β_1: MSP × Urban	–8.923	0.223**	32.790***	0.223**	0.304***
	(7.485)	(0.104)	(8.515)	(0.090)	(0.065)
β_2: MSP × Rural	–6.533	0.639***	48.990**	–0.064	0.223***
	(5.582)	(0.110)	(20.079)	(0.107)	(0.076)
Observations	838	848	810	829	844
R squared	0.007	0.072	0.063	0.013	0.062
Control Mean (Urban)	17.020	0.503	26.243	0.490	0.073
Control Mean (Rural)	22.136	0.350	38.175	0.415	0.058
p-value $\beta_1 = \beta_2$	0.798	0.006	0.458	0.041	0.416
# Reg. Zones	310	310	303	308	310
Panel B. Year 2 Outcomes					
β_1: MSP × Urban	–13.112***	–0.249*	–0.746	0.224*	
	(4.976)	(0.129)	(7.451)	(0.129)	
β_2: MSP × Rural	–25.817***	0.195**	6.208	0.284***	
	(6.280)	(0.097)	(10.343)	(0.109)	
Observations	644	642	632	623	
R squared	0.067	–0.031	0.002	0.100	
Control Mean (Urban)	4.914	0.345	37.035	0.776	
Control Mean (Rural)	17.716	0.198	36.038	0.620	
p-value $\beta_1 = \beta_2$	0.114	0.006	0.586	0.721	
# Reg. Zones	286	287	283	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. 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 AFN. Conducted Mobile 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 Impact on Employee Retention

	Left MoE			Joined MoE		
	All	Urban	Rural	All	Urban	Rural
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. 2SLS Estimates						
MSP (=1)	–0.275*	–0.255	–0.361*	–0.040**	–0.055**	–0.019
	(0.157)	(0.195)	(0.210)	(0.018)	(0.028)	(0.018)
R^2	0.323	0.304	0.345	0.077	0.076	0.074
Panel B. Reduced Form (=1)						
EE (=1)	–0.060	–0.061	–0.061	–0.027**	–0.033**	–0.015
	(0.054)	(0.065)	(0.088)	(0.011)	(0.015)	(0.012)
ED (=1)	–0.115**	–0.099	–0.165**	–0.035***	–0.042***	–0.020*
	(0.052)	(0.062)	(0.075)	(0.010)	(0.015)	(0.010)
Observations	1,295	907	388	31,079	18,915	12,164
R^2	0.202	0.165	0.284	0.084	0.090	0.076
p-value EE + ED = 0	0.067	0.167	0.095	0.001	0.006	0.085
Control Mean	0.550	0.550	0.530	0.180	0.190	0.170
# Reg. Zones	254	152	102	345	195	150

Notes: This table reports the ToT and Reduced Form estimates of impacts of mobile salary payments on employee turnover. Left MoE is a dummy variable that equals one when an employee has not been paid for at least three consecutive months. In columns 1 to 3 we restrict the sample to employees who had a tenure of less than one year at the beginning of the reform. Results for the impact on leaving MoE for all employees and for other tenure bands are reported in Appendix Table A.10. Joined MoE is a dummy variable that equals to one if an employee was never paid before the reform and was at least paid once after 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 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 six 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	Transfer to Cust. Wallet	Own Airtime Top-up	Other Airtime Top-up	Balance
	(1)	(2)	(3)	(4)	(5)
Panel A. OLS					
MSP Months	0.004***	0.003**	0.001***	0.002***	0.012***
	(0.001)	(0.002)	(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. 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 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)	(6)	(7)
Taliban Will Lose at Baseline (2018, %)	0.994***		0.918**	0.804*	6.488***	6.868***	5.641*
	(0.343)		(0.402)	(0.429)	(2.212)	(2.179)	(3.274)
SD of Beliefs					5.397**	5.826***	4.681
					(2.181)	(2.010)	(3.019)
Levels of Beliefs × SD of Beliefs					–5.510**	–6.207***	–4.896
					(2.224)	(2.135)	(3.316)
Taliban Will Lose Pre-Reform (2017, %)		0.413	–0.044	–0.173		–0.041	–0.090
		(0.300)	(0.290)	(0.304)		(0.265)	(0.281)
Taliban Will Lose Pre-Reform (2016, %)		0.036	–0.579*	–0.476		–0.610**	–0.557*
		(0.348)	(0.326)	(0.294)		(0.241)	(0.286)
Taliban Will Lose Pre-Reform (2015, %)		0.444	0.463	0.267		0.491*	0.249
		(0.366)	(0.299)	(0.274)		(0.268)	(0.222)
Police Capable of Protection (2018, %)			0.048	–0.132		0.056	–0.016
			(0.142)	(0.119)		(0.126)	(0.136)
Pashtun (2018, %)			–0.014	–0.155**		–0.022	–0.088
			(0.039)	(0.059)		(0.040)	(0.065)
On Primary Road (=1)			0.081**	0.077**		0.055**	0.046
			(0.030)	(0.031)		(0.024)	(0.028)
Urban (=1)			0.018	0.024		0.052	0.049
			(0.029)	(0.033)		(0.036)	(0.039)
Contested District (=1)			–0.122**	–0.048		–0.098**	–0.052
			(0.045)	(0.051)		(0.043)	(0.040)
Bad Cell Coverage (%)			–0.001	–0.046		0.023	–0.039
			(0.067)	(0.064)		(0.076)	(0.064)
Based at Registration School (=1)			0.046**	0.050**		0.048**	0.048**
			(0.021)	(0.021)		(0.021)	(0.021)
Female			0.011	0.014		0.016	0.016
			(0.022)	(0.021)		(0.021)	(0.021)
Has At Least A Bachelor's			0.059*	0.058*		0.058*	0.058*
			(0.031)	(0.031)		(0.031)	(0.031)
Early Reg., Early Mobile Money			0.198***	0.198***		0.196***	0.196***
			(0.035)	(0.035)		(0.035)	(0.035)
Late Reg., Early Mobile Money			0.195***	0.191***		0.193***	0.189***
			(0.036)	(0.036)		(0.036)	(0.036)
Government is Effective (Overall, 2018, %)			–0.389***	–0.432***		–0.418***	–0.420***
			(0.129)	(0.140)		(0.127)	(0.128)
Government is Going in the Right Direction (2018, %)				0.005			–0.092
				(0.208)			(0.210)
Confident in the Government (2018, %)				0.765***			0.606**
				(0.213)			(0.263)
Government Cares About Your Needs/Community (2018, %)				0.119			–0.017
				(0.148)			(0.178)
Constant	–0.111	0.009	0.099	–0.024	–5.486**	–5.493***	–4.531
	(0.328)	(0.349)	(0.308)	(0.275)	(2.191)	(1.934)	(2.760)
Observations	19,626	19,626	19,626	19,626	19,626	19,626	19,626
R squared	0.013	0.011	0.095	0.103	0.020	0.100	0.105
# Districts	31	31	31	31	31	31	31
WC Bootstrap p-value	0.031		0.128	0.218	0.028	0.049	0.300
Mean Dependent Var	0.809	0.809	0.809	0.809	0.809	0.809	0.809

Notes: This table reports on the relationship between citizens' 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. Support for government is proxied using four ANQAR questions - with exact question wording provided in Appendix Table A.26. The sample includes teachers from all three treatment arms who were paid at least once before the reform began. It excludes unsafe schools and any school where fewer than 10% of employees received a payment in the 12 months preceding the reform. Table A.20 provides 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, with province fixed effects. Standard errors clustered at the district level are reported in parentheses. Wild Cluster (WC) Bootstrap p-value calculated for Taliban Will Lose at Baseline (2018, %) coefficient.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Appendix - For Online Publication

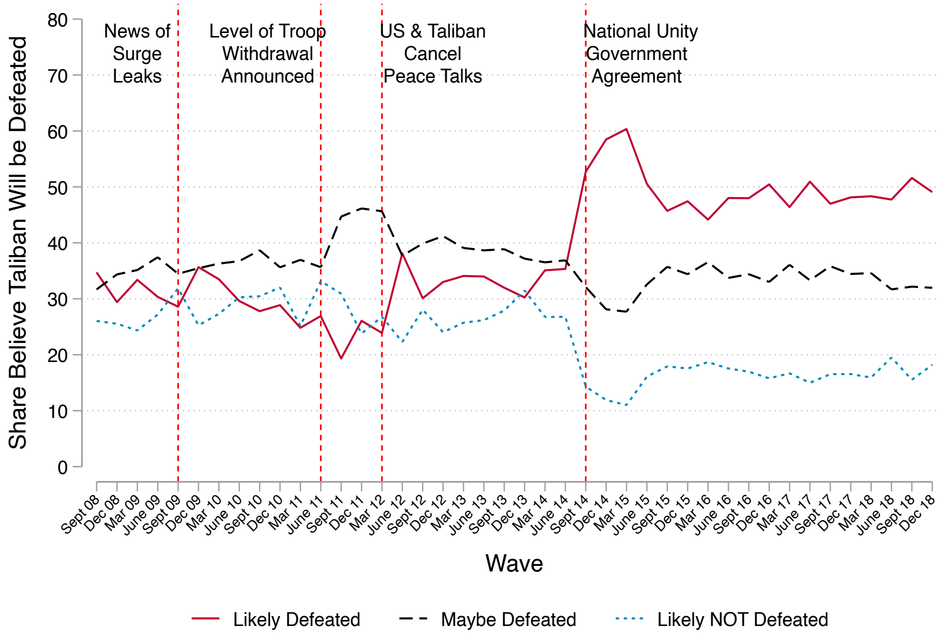
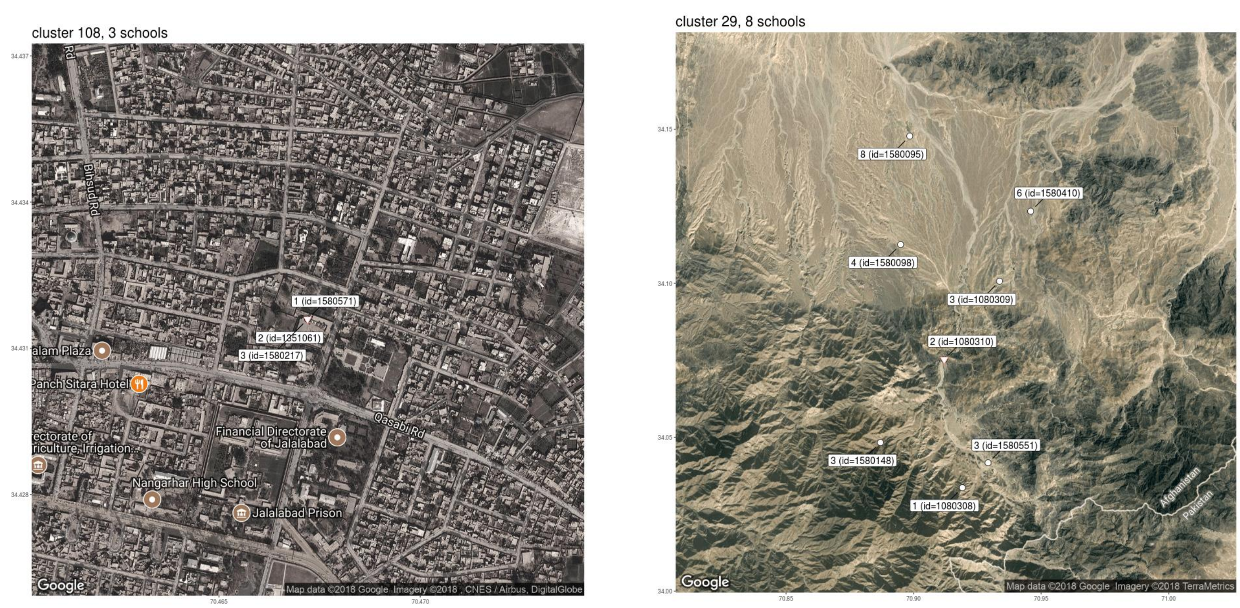


Figure A.1: Evolution of Beliefs Over Conflict's Outcome

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.

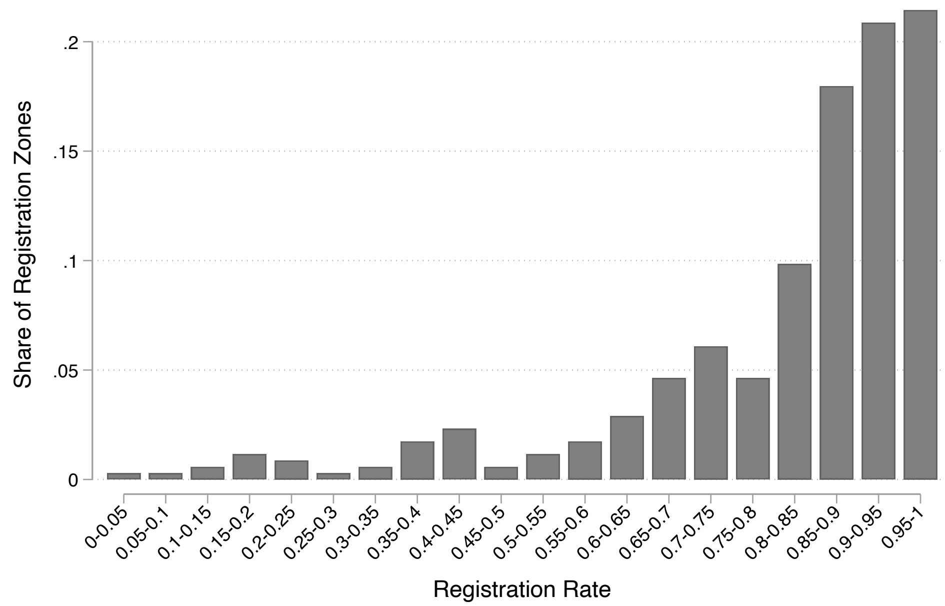


Panel A: Urban Registration Zone

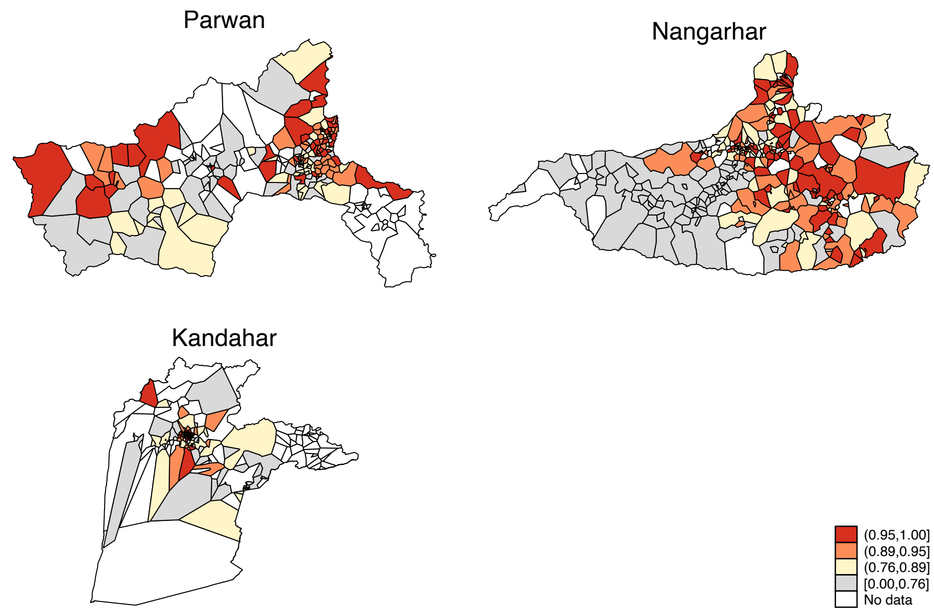
Panel B: Rural Registration Zone

Figure A.2: 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.



Panel A: Histogram



Panel B: Map of Registration Levels by Zone

Figure A.3: 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).

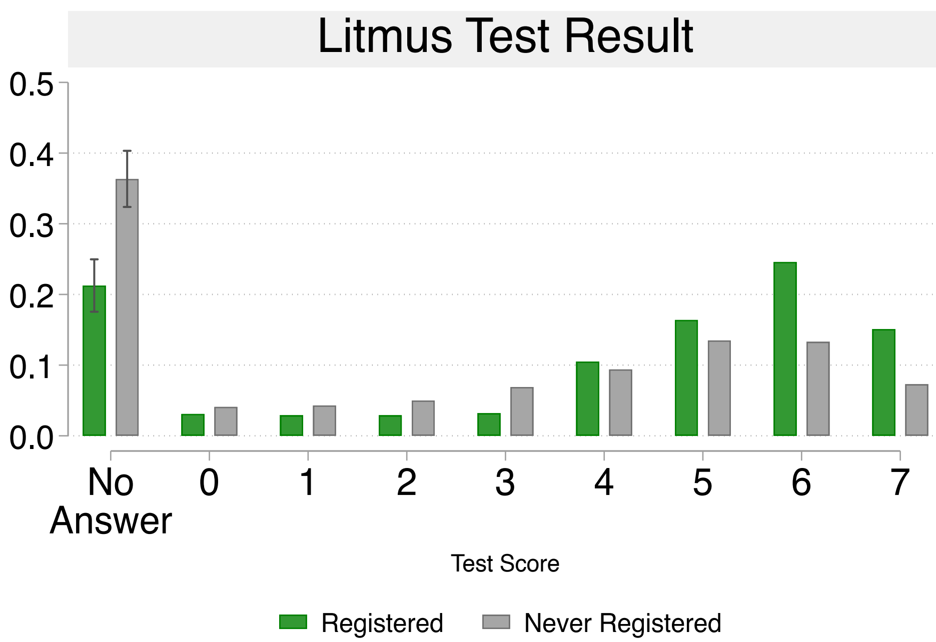


Figure A.4: Litmus Test Results

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: (i) school district, (ii) school name, (iii) employee's rank, (iv) employee's position, (v) principal's name, (vi) headmaster's name, and (vii) 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.

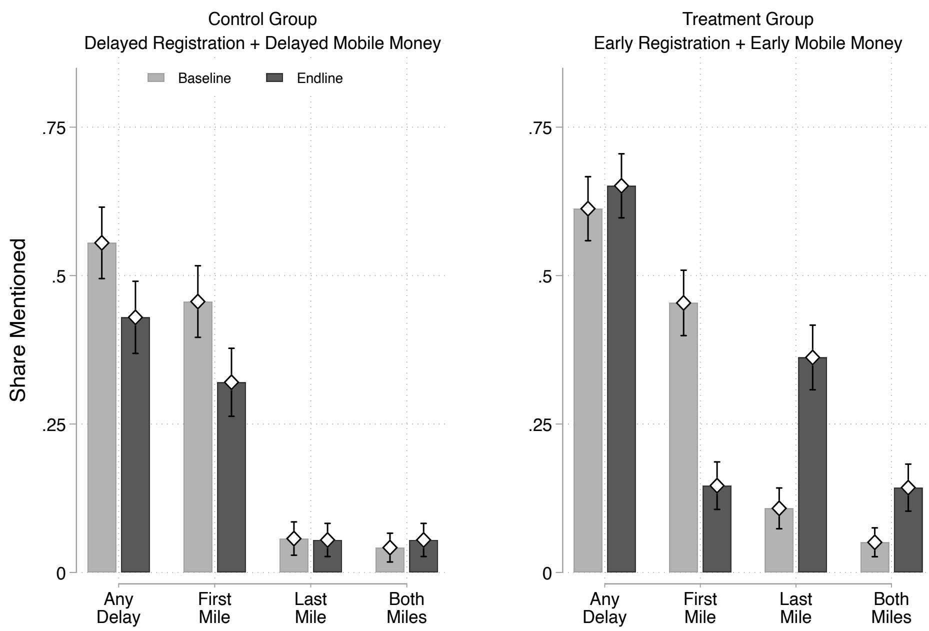


Figure A.5: Separating First Mile and Last Mile Delays

Notes: This figure depicts the causes of delays. Respondents who reported salary delays were asked who was responsible for the delay. First mile delays are those due to the Ministry of Education, the Ministry of Finance, or the Central Bank (Da Afghanistan Bank). Last mile delays are those due to New Kabul Bank, Motameds, the mobile network operator, or the mobile money agents. 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 end line 2 survey conducted by phone in May 2020 due to Covid-19.

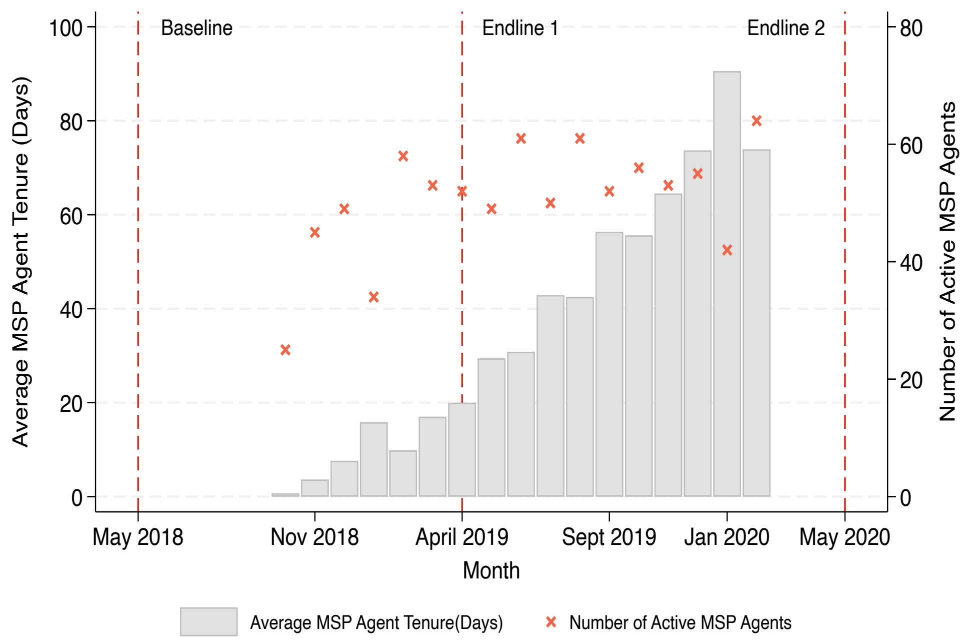
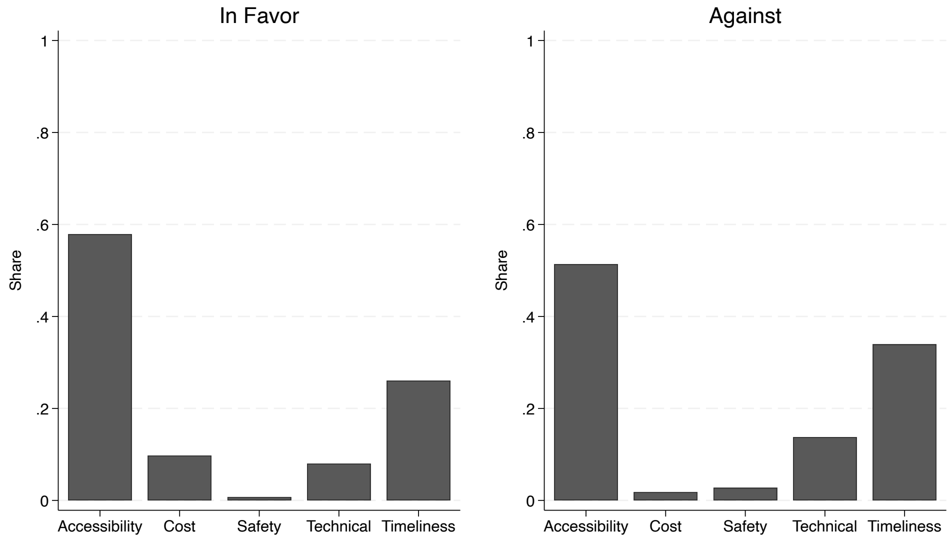
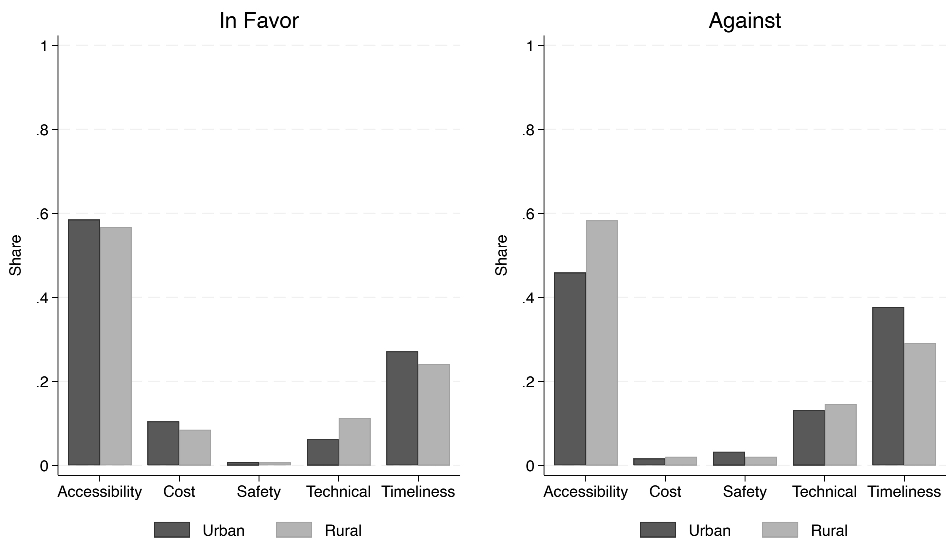


Figure A.6: Tenure of MSP Agents

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.7: 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 the 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).

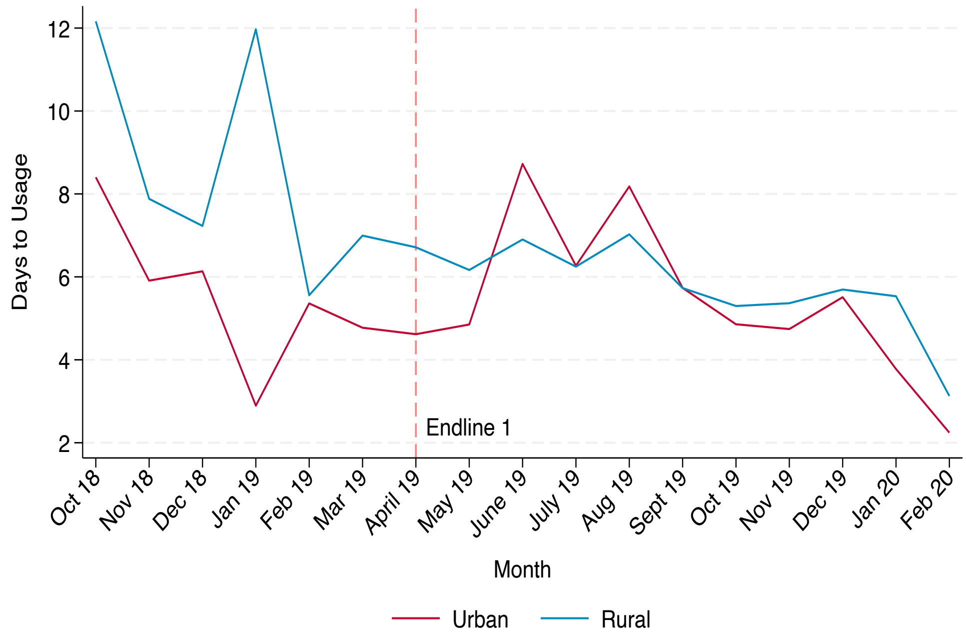


Figure A.8: Days to Mobile Account Usage

Notes: This figure shows the 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.

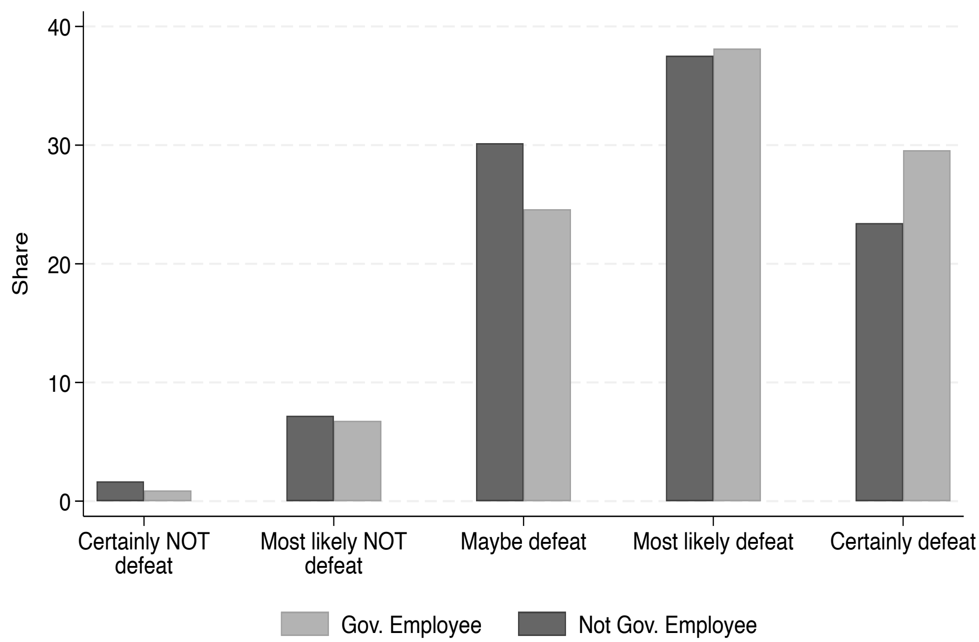
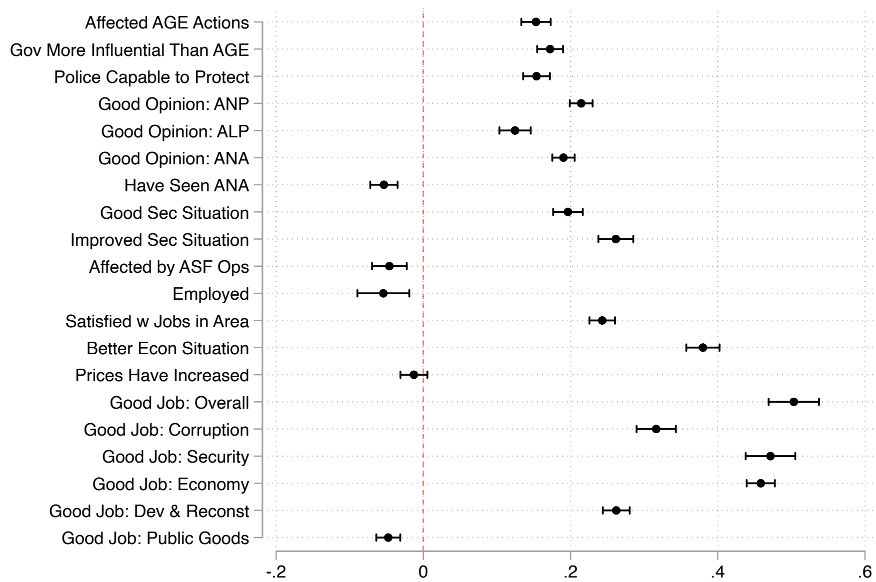
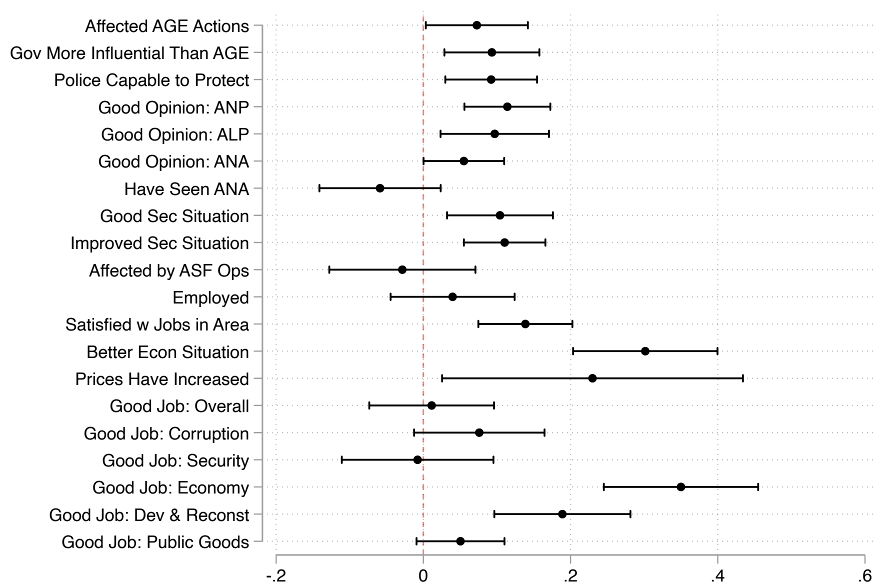


Figure A.9: Beliefs of Taliban Defeat by Occupation

Notes: This figure 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

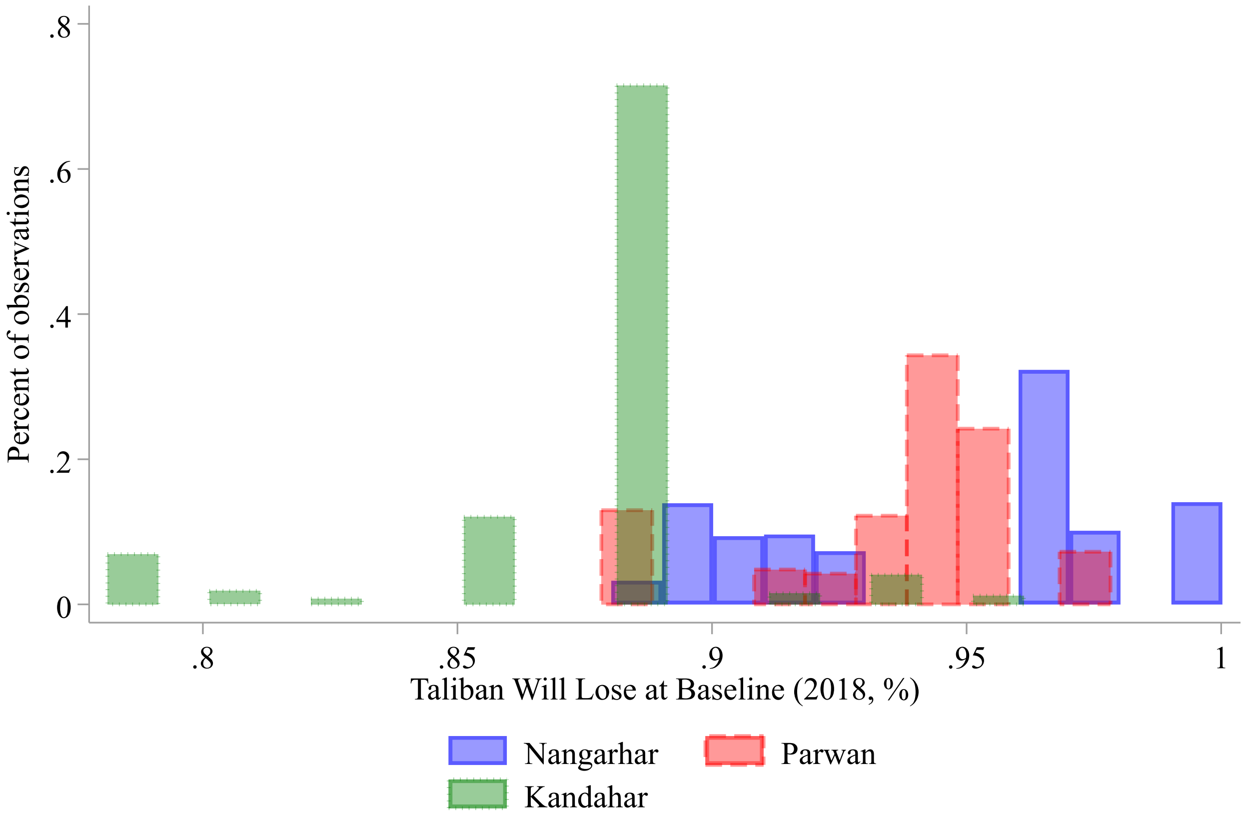


Panel B: 2015-2018, Intervention's Outcome

Figure A.10: 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.

Figure A.11: The Distribution of Beliefs about State Durability by Province



Notes: This figure displays the distribution of respondents' beliefs in 2018 as recorded by the ANQAR survey regarding whether the Taliban would be defeated shown separately for Nangarhar, Parwan, and Kandahar. Across the three provinces, 93% of respondents reported believing the Taliban would lose; the corresponding averages were 88% in Kandahar, 94% in Parwan, and 96% in Nangarhar. The standard deviation of beliefs was 0.0446 overall, with between-province and within-province components of 0.0381 and 0.036, respectively.

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]	
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]	
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 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 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 or her 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 24,768 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 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 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 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 level in squared brackets.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A.5: 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 A.6: 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 poor cell coverage, 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.7: 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.8: 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 the time of the survey. Robust standard errors clustered at the registration zone level are reported in parentheses.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.9: Heterogeneous Treatment Effects on Payment Experience – Security

	Payment to Receive Transfer (1)	Experienced Delay (2)	Travel Time to Cash-Out Salary (Min) (3)	Vote in Favor of MSP (4)	Conducted Mobile Money Transfer (5)
Panel A. Year 1 Outcomes					
β_1: MSP $\times$ Controlled	–5.532	0.340***	38.476***	0.162**	0.318***
	(5.393)	(0.080)	(9.180)	(0.075)	(0.053)
β_2: MSP $\times$ Contested	–24.009	0.660**	44.166	–0.192	–0.048
	(14.836)	(0.266)	(40.023)	(0.188)	(0.129)
p-value $\beta_1 = \beta_2$	0.243	0.250	0.890	0.081	0.009
Control Mean	19.094	0.441	31.219	0.461	0.067
Observations	838	848	810	829	844
R squared	0.001	0.060	0.054	0.004	0.051
# Reg. Zones	310	310	303	308	310
Panel B. Year 2 Outcomes					
β_1: MSP $\times$ Controlled	–15.523***	–0.172*	2.513	0.240**	
	(3.761)	(0.101)	(6.508)	(0.098)	
β_2: MSP $\times$ Contested	–30.342**	0.352**	1.274	0.289	
	(12.907)	(0.164)	(15.926)	(0.178)	
p-value $\beta_1 = \beta_2$	0.271	0.007	0.943	0.813	
Control Mean	10.178	0.284	36.629	0.710	
Observations	644	642	632	623	
R squared	0.068	–0.044	–0.001	0.102	
# Reg. Zones	286	287	283	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. Government control is based on territorial control assessments provided by the Long War Journal. Favor MSP is a dummy variable equal to one if the respondent indicates they would support scaling the reform across the Ministry of Education. Experienced Delay is a dummy variable equal to one if the respondent reports their salary being delayed. Travel Time is the time to convert the mobile money payment to cash 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 8,560 AFN. Conducted Mobile 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 two 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: Estimates of Impact on Leaving MoE

	Left MoE				
	All	1-2 Years Years	2-5 Years Years	6-10 Years Years	> 10 Years
	(1)	(2)	(3)	(4)	(5)
Panel A. 2SLS Estimates					
MSP (=1)	–0.020	–0.049	–0.036	–0.001	–0.008
	(0.018)	(0.093)	(0.031)	(0.018)	(0.014)
R^2	0.104	0.196	0.095	0.072	0.020
Panel B. Reduced Form (=1)					
EE (=1)	–0.016	–0.013	–0.018	–0.006	–0.007
	(0.013)	(0.036)	(0.016)	(0.012)	(0.013)
ED (=1)	–0.020	–0.045	0.003	–0.015	–0.006
	(0.013)	(0.036)	(0.016)	(0.010)	(0.013)
Observations	24,570	1,642	4,029	7,767	9,836
R^2	0.095	0.168	0.083	0.072	0.017
p-value: EE + ED = 0	0.126	0.361	0.598	0.295	0.597
Control Mean	0.150	0.310	0.150	0.100	0.100
# Reg. Zones	345	286	341	345	345

Notes: This table reports the ToT and Reduced Form estimates of the impacts of mobile salary payments on leaving the Ministry of Education (MoE). Left MoE is a dummy variable that equals one when an employee has not been paid for at least three consecutive months. The heterogeneous impacts on "Left MoE" are presented for different tenure bands (in years). 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.11: 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 six 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.12: 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 level in squared brackets.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A.13: Treatment Effects on Attendance across Urban & Rural Districts

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

Table A.14: Heterogeneous Impacts on Mobile Money Use - 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 heterogenous impact of receiving salary via mobile money on other dimensions of mobile money use across urban and rural districts. 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. 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 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.15: 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	–0.046	–0.170**
	(0.056)		(0.088)	(0.092)	(0.068)
AM p × BM p(1-p) × AM SD of p	0.058		0.043	0.065	0.134**
	(0.060)		(0.061)	(0.041)	(0.050)
AM p × BM p(1-p) × BM SD of p	0.122**		0.159***	0.049	–0.003
	(0.055)		(0.055)	(0.066)	(0.048)
BM p × AM p(1-p) × BM SD of p	–0.026		0.064	–0.014	0.011
	(0.057)		(0.073)	(0.036)	(0.058)
BM p × BM p(1-p) × AM SD of p	0.127**		0.083	0.074	–0.004
	(0.059)		(0.065)	(0.048)	(0.073)
Taliban Will Lose Pre-Reform (2017, %)		0.413	0.327		0.154
		(0.300)	(0.193)		(0.127)
Taliban Will Lose Pre-Reform (2016, %)		0.036	–0.611***		–0.584**
		(0.348)	(0.215)		(0.228)
Taliban Will Lose Pre-Reform (2015, %)		0.444	0.278		–0.145
		(0.366)	(0.334)		(0.324)
Police Capable of Protection (2018, %)			0.052		–0.249**
			(0.110)		(0.106)
Pashtun (2018, %)			–0.055		–0.267***
			(0.047)		(0.046)
On Primary Road (=1)			0.085*		0.083**
			(0.047)		(0.039)
Urban (=1)			0.010		0.085***
			(0.024)		(0.024)
Contested District (=1)			–0.148***		0.041
			(0.045)		(0.038)
Bad Cell Coverage (%)			0.057		–0.029
			(0.067)		(0.048)
Based at Registration School (=1)			0.046**		0.050**
			(0.022)		(0.020)
Female			0.017		0.017
			(0.021)		(0.021)
Has At Least A Bachelor's			0.061*		0.063**
			(0.031)		(0.031)
Early Reg., Early Mobile Money			0.192***		0.194***
			(0.035)		(0.035)
Late Reg., Early Mobile Money			0.193***		0.192***
			(0.037)		(0.037)
Government is Effective (Overall, 2018, %)			–0.575***	–0.048	–0.305
			(0.150)	(0.159)	(0.251)
Government is Going in the Right Direction (2018, %)				–0.235	–0.534**
				(0.296)	(0.242)
Confident in the Government (2018, %)				0.860**	1.669***
				(0.382)	(0.312)
Government Cares About your Needs/Community (2018, %)				0.065	0.738***
				(0.241)	(0.139)
Constant	0.757***	0.009	0.863**	0.189	0.115
	(0.050)	(0.349)	(0.323)	(0.327)	(0.410)
Observations	19,626	19,626	19,626	19,626	19,626
R squared	0.023	0.011	0.099	0.034	0.109
# 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. Support for government is proxied using four ANQAR questions - with exact question wording provided in Appendix Table A.26. 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.16: Beliefs About State Durability and Registration at Any Time

Dependent Variable	Registered at Any Time (=1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Taliban Will Lose at Baseline (2018, %)	0.676***		0.547*	0.527*	3.100*	2.607	–0.145
	(0.219)		(0.280)	(0.290)	(1.766)	(1.757)	(2.565)
SD of Beliefs					2.375	2.114	–0.665
					(1.777)	(1.693)	(2.363)
Levels of Beliefs × SD of Beliefs					–2.392	–2.421	0.735
					(1.816)	(1.822)	(2.587)
Taliban Will Lose Pre-Reform (2017, %)		0.293	0.013	–0.121		–0.025	–0.125
		(0.230)	(0.225)	(0.243)		(0.228)	(0.258)
Taliban Will Lose Pre-Reform (2016, %)		0.054	–0.363*	–0.255		–0.363*	–0.245
		(0.215)	(0.214)	(0.188)		(0.204)	(0.206)
Taliban Will Lose Pre-Reform (2015, %)		0.288	0.174	0.043		0.252	0.027
		(0.230)	(0.234)	(0.213)		(0.215)	(0.175)
Police Capable of Protection (2018, %)			0.060	–0.103		0.047	–0.121
			(0.109)	(0.089)		(0.113)	(0.104)
Pashtun (2018, %)			–0.042	–0.157***		–0.054	–0.169***
			(0.031)	(0.051)		(0.035)	(0.054)
On Primary Road (=1)			0.063***	0.061***		0.057***	0.065***
			(0.020)	(0.020)		(0.017)	(0.023)
Urban (=1)			0.013	0.015		0.029	0.011
			(0.024)	(0.028)		(0.030)	(0.033)
Contested District (=1)			–0.046	0.011		–0.036	0.014
			(0.029)	(0.035)		(0.032)	(0.029)
Bad Cell Coverage (%)			–0.022	–0.057		–0.003	–0.061
			(0.050)	(0.048)		(0.057)	(0.047)
Based at Registration School (=1)			0.039*	0.042**		0.041**	0.042**
			(0.020)	(0.020)		(0.020)	(0.020)
Female			–0.001	0.001		0.001	0.001
			(0.012)	(0.011)		(0.011)	(0.011)
Has At Least A Bachelor's			0.055**	0.055**		0.055**	0.055**
			(0.026)	(0.026)		(0.026)	(0.026)
Early Reg., Early Mobile Money			0.199***	0.199***		0.198***	0.199***
			(0.036)	(0.036)		(0.036)	(0.036)
Late Reg., Early Mobile Money			0.201***	0.197***		0.200***	0.198***
			(0.036)	(0.036)		(0.036)	(0.036)
Government is Effective (Overall, 2018, %)			–0.217**	–0.300***		–0.215**	–0.302***
			(0.103)	(0.097)		(0.099)	(0.102)
Government is Going in the Right Direction (2018, %)				0.096			0.106
				(0.165)			(0.185)
Confident in the Government (2018, %)				0.605***			0.640***
				(0.158)			(0.195)
Government Cares About your Needs/Community (2018, %)				0.067			0.099
				(0.113)			(0.133)
Constant	0.223	0.280	0.414*	0.254	–2.178	–1.413	0.847
	(0.208)	(0.237)	(0.238)	(0.210)	(1.740)	(1.551)	(2.159)
Observations	19,626	19,626	19,626	19,626	19,626	19,626	19,626
R squared	0.007	0.006	0.087	0.094	0.009	0.089	0.094
# Districts	31	31	31	31	31	31	31
WC Bootstrap p-value	0.054		0.175	0.251	0.117	0.268	0.964
Mean Dependent Var	0.848	0.848	0.848	0.848	0.848	0.848	0.848

Notes: This table reports on the relationship between citizens' 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. Support for government is proxied using four ANQAR questions - with exact question wording provided in Appendix Table A.26. 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.17: Heterogeneous Impacts on Payment Experience by Beliefs

	Payment to Receive Transfer (1)	Experienced Delay (2)	Travel Time to Cash-Out Salary (Min) (3)	Vote in Favor of MSP (4)	Conducted Mobile Money Transfer (5)
Panel A. Year 1 Outcomes					
β_1: MSP × Above Median Taliban Will Lose at Baseline (2018, %)	–6.138	0.378***	46.065***	0.176*	0.272***
	(7.942)	(0.104)	(14.823)	(0.091)	(0.063)
β_2: MSP × Below Median Taliban Will Lose at Baseline (2018, %)	–11.504*	0.414***	42.082***	0.008	0.209**
	(6.432)	(0.146)	(15.365)	(0.127)	(0.087)
Observations	717	726	696	708	722
R squared	0.254	0.238	0.283	0.329	0.174
Control Mean	20.276	0.427	29.219	0.469	0.069
Control Mean - Below Median Taliban Will Lose at Baseline (2018, %)	17.273	0.364	25.827	0.635	0.092
Control Mean - Above Median Taliban Will Lose at Baseline (2018, %)	22.797	0.479	32.188	0.324	0.051
p-value $\beta_1 = \beta_2$	0.600	0.840	0.852	0.283	0.554
# Reg. Zones	267	267	261	265	267
Panel B. Year 2 Outcomes					
β_1: MSP × Above Median Taliban Will Lose at Baseline (2018, %)	–13.172**	–0.247*	1.926	0.193*	
	(6.080)	(0.130)	(9.974)	(0.111)	
β_2: MSP × Below Median Taliban Will Lose at Baseline (2018, %)	–24.916***	0.102	3.356	0.293*	
	(4.760)	(0.152)	(11.117)	(0.178)	
Observations	546	545	536	529	
R squared	0.244	0.130	0.120	0.286	
Control Mean	10.610	0.293	36.642	0.723	
Control Mean - Below Median Taliban Will Lose at Baseline (2018, %)	9.176	0.224	31.059	0.714	
Control Mean - Above Median Taliban Will Lose at Baseline (2018, %)	12.152	0.367	42.805	0.732	
p-value $\beta_1 = \beta_2$	0.128	0.081	0.924	0.632	
# Reg. Zones	244	245	241	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 A.24. Favor MSP is a dummy variable equal to one if the respondent indicates they would support scaling the reform across the Ministry of Education. Experienced Delay is a dummy variable equal to one if the respondent reports their salary being delayed. Travel Time is the time to convert the mobile money payment to cash 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. Conducted Mobile 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.18: Heterogeneous Impacts on Student Learning by Beliefs

	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 hereogenous 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 six 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. 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 ED groups by April 2019.

Table A.19: Heterogeneous Impacts on Mobile Money Use by 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	0.009
	(0.002)	(0.002)	(0.000)	(0.000)	(0.007)
β_2: MSP Months × Below Median Taliban Will Lose at Baseline (2018, %)	0.004***	0.004	0.001***	0.002***	0.006***
	(0.001)	(0.003)	(0.000)	(0.000)	(0.002)
Observations	9,773	9,773	9,773	9,773	9,773
R^2	0.004	0.014	0.026	0.055	0.149
1-Month Mean	0.002	0	0.000	0.001	0.163
1-Month Mean (Below Median Taliban Will Lose at Baseline (2018, %))	0.002	0	0.000	0.001	0.150
1-Month Mean (Above Median Taliban Will Lose at Baseline (2018, %))	0	0	0	0	0.813
p-value $\beta_1 = \beta_2$	0.145	0.256	0.000	0.000	0.708
# Reg. Zones	235	235	235	235	235
Panel B. 2SLS					
β_1: MSP Months × Above Median Taliban Will Lose at Baseline (2018, %)	0.013	0.008	0.000	0.000	0.006
	(0.014)	(0.018)	(0.000)	(0.000)	(0.040)
β_2: MSP Months × Below Median Taliban Will Lose at Baseline (2018, %)	0.004***	0.005	0.001***	0.003***	0.008***
	(0.001)	(0.003)	(0.000)	(0.001)	(0.003)
Observations	9,773	9,773	9,773	9,773	9,773
R^2	–0.002	–0.000	0.008	0.016	0.002
p-value $\beta_1 = \beta_2$	0.487	0.871	0.049	0.000	0.957
# Reg. Zones	235	235	235	235	235

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 level are reported in parentheses.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A.20: Beliefs About State Durability, Government and Registration During On-site Drive

Dependent Variable	Registered During Drive (=1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Taliban Will Lose at Baseline (2018, %)	–0.159		–0.038	–0.146	1.053	1.980	–1.206
	(0.313)		(0.338)	(0.296)	(1.943)	(1.809)	(2.585)
SD of Beliefs					1.403	2.587*	–0.242
					(1.795)	(1.497)	(2.348)
Levels of Beliefs × SD of Beliefs					–1.770	–3.447**	–0.614
					(1.801)	(1.564)	(2.566)
Taliban Will Lose Pre-Reform (2017, %)		–0.202	–0.184	–0.234		–0.314*	–0.441**
		(0.305)	(0.198)	(0.170)		(0.175)	(0.163)
Taliban Will Lose Pre-Reform (2016, %)		–0.414	–0.573**	–0.449**		–0.514***	–0.378
		(0.248)	(0.232)	(0.204)		(0.179)	(0.225)
Taliban Will Lose Pre-Reform (2015, %)		0.268	0.139	0.113		0.337*	0.441**
		(0.270)	(0.263)	(0.220)		(0.189)	(0.195)
Police Capable of Protection (2018, %)			0.012	–0.161		–0.025	–0.146
			(0.091)	(0.096)		(0.079)	(0.112)
Pashtun (2018, %)			–0.107	–0.135**		–0.072	–0.056
			(0.069)	(0.058)		(0.048)	(0.050)
On Primary Road (=1)			–0.001	–0.002		–0.026	–0.031
			(0.028)	(0.029)		(0.020)	(0.020)
Urban (=1)			–0.002	–0.009		0.019	–0.016
			(0.025)	(0.024)		(0.029)	(0.033)
Contested District (=1)			–0.106***	–0.078**		–0.105***	–0.136***
			(0.035)	(0.036)		(0.036)	(0.035)
Bad Cell Coverage (%)			–0.064	–0.086*		–0.049	–0.058
			(0.049)	(0.043)		(0.042)	(0.037)
Based at Registration School (=1)			0.045**	0.049**		0.051**	0.051**
			(0.021)	(0.021)		(0.020)	(0.020)
Female			0.015	0.020		0.020	0.022
			(0.021)	(0.020)		(0.020)	(0.020)
Has At Least A Bachelor's			0.061*	0.062*		0.060*	0.062*
			(0.031)	(0.031)		(0.031)	(0.031)
Early Reg., Early Mobile Money			0.190***	0.192***		0.189***	0.189***
			(0.034)	(0.034)		(0.033)	(0.034)
Late Reg., Early Mobile Money			0.185***	0.187***		0.183***	0.184***
			(0.036)	(0.036)		(0.035)	(0.036)
Government is Effective (Overall, 2018, %)			–0.191	–0.281**		–0.133	–0.201*
			(0.113)	(0.133)		(0.098)	(0.101)
Government is Going in the Right Direction (2018, %)				0.192			0.371*
				(0.183)			(0.195)
Confident in the Government (2018, %)				0.281			–0.220
				(0.267)			(0.275)
Government Cares About Your Needs/Community (2018, %)				0.304***			0.142
				(0.100)			(0.143)
Constant	0.956***	1.129***	1.457***	1.264***	0.049	–0.007	3.136
	(0.292)	(0.308)	(0.401)	(0.376)	(1.961)	(1.694)	(2.178)
Observations	19,626	19,626	19,626	19,626	19,626	19,626	19,626
R squared	0.036	0.039	0.108	0.112	0.038	0.114	0.115
# Districts	31	31	31	31	31	31	31
WC Bootstrap p-value	0.714		0.936	0.718	0.571	0.391	0.718
Mean Dependent Var	0.809	0.809	0.809	0.809	0.809	0.809	0.809

Notes: This table reports on the relationship between citizens' 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. Support for government is proxied using four ANQAR questions - with exact question wording provided in Appendix Table A.26. The sample includes teachers from all three treatment arms who were paid at least once before the reform began. All regressions control for province fixed effects. 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.21: 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 six months between December 2017 and April 2018 (i.e., the six 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 or she registered for a mobile money wallet during the main registration wave (before July 18, 2018). Alternatively, the employee is "Registered with Delay" if he or she registered after the main registration wave (after July 18, 2018).

Table A.22: 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. Employees Paid At Any Point								
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)
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
# Months	36	36	36	36	36	36	36	36
Observations	12,420	12,420	12,420	12,420	12,420	12,420	12,420	12,420
Panel B. Employees Paid In Each Of The 6 Months Before Start Of Registration								
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)
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
# Months	36	36	36	36	36	36	36	36
Observations	12,384	12,384	12,384	12,384	12,384	12,384	12,384	12,384

Notes: This table estimates the impact of the reform on the wage bill. The analysis is aggregated at the registration zone and month level. Treatment is a dummy variable equal to one for registration zones in the EE and ED treatment groups. Post is a dummy variable equal to one starting from either May 2018 (columns 1 to 4) or November 2018 (column 5 to 8). Outcome variables are number of employed teachers (columns 1 and 5), net total monthly wage bill (columns 2 and 6, includes for example overtime payments), net monthly wage bill (columns 3 and 7), and gross monthly salary (columns 4 and 8). The sample used to calculate the variables, outcome variables at the registration zone, and month level consists of employees who were paid at any point between March 2017 and March 2020 (i.e., the entire period for which we have data - Panel A), or in each of the 6 months between December 2017 and April 2018 (i.e., the 6 months before the start of registration – Panel B). 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.23: Spatial Treatment Externalities on Salary Payment Experience

	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)	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]	
Experienced Delay	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]	
Travel Time to Cash-Out (Min)	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]	
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]	
Panel B. Endline 2 – May 2020							
Payment to Receive Salary (Afg)	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]	
Experienced Delay	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]	
Travel Time to Cash-Out (Min)	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]	
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]	

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 B. 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 level in squared brackets.
Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.24: Effects on Salary Payment Experience – With Stratum FEs

	DD Mean (1)	EE Effect (2)	ED Effect (3)	ToT Effect (4)	N Obs (5)
Panel A. Endline 1					
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]		
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]		
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. All regressions include stratum fixed effects. Robust standard errors clustered at the registration zone level. FWER-adjusted p-values within each panel in squared brackets (following Romano & Wolf, 2005, using 1,000 repetitions). Appendix B discusses spatial externalities of the MSP reform on salary payment experience across studied schools. The results are presented in Appendix Table A.23.
Levels of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table A.25: District-Level Correlations Between Beliefs About the Government and Taliban Defeat

	Taliban Will Lose at Baseline (2018, %)	Government is Effective (Overall) (2018, %)	Government is Going in the Right Direction (2018, %)	Confident in the Government (2018, %)	Government Cares About Needs/ Community (2018, %)
Taliban Will Lose at Baseline (2018, %)	1.0000				
Government is Effective (Overall, 2018, %)	0.1801	1.0000			
	(0.3238)				
Government is going in the Right Direction (2018, %)	-0.1592	0.6362*	1.0000		
	(0.3843)	(0.0001)			
Confident in the Government (2018, %)	0.2487	0.5188*	0.4444*	1.0000	
	(0.1698)	(0.0023)	(0.0108)		
Government Cares about Needs/Community (2018, %)	0.0494	0.3027	0.3873*	-0.0678	1.0000
	(0.7884)	(0.0921)	(0.0285)	(0.7124)	

Notes: This table reports the pairwise correlation coefficients computed at the district level. Asterisk indicates significance at the 5% level (*$p < 0.05$). Standard errors (p-values) are in parentheses.

Table A.26: Survey Measures Proxying Support for the Government of Afghanistan

Variable Name	**Survey Question**
Government is Effective (Overall, 2018, %)	How well does the Government of Afghanistan do its job? Is it doing its job very well, a little well, neither well nor poorly, a little poorly, or very poorly?
Government is going in the right direction (2018, %)	Generally speaking, do you believe the Government of Afghanistan is going in the right direction, the wrong direction, or staying in the same place?
Confident in the government (2018, %)	How confident are you in the Government of Afghanistan? Are you very confident, somewhat confident, not confident, or not at all confident?
Government cares about your needs/community (2018, %)	Does the Government of Afghanistan care about your needs or the needs of your community?

Table A.27: Studies of Technology and State Capacity in Low-, Lower-Middle-, and Upper-Middle-Income Countries

Study Title	Description	Country	Technology Evaluated	State-Capacity Type	Fragility *Values: Extremely fragile; Fragile; Non-fragile*
Allie (2023)	Facial recognition technology at polling stations reduces voter turnout in Indian local elections.	India	Facial recognition in polling stations	Coercive capacity	Non-Fragile
Assunção, Gandour and Rocha (2023)	Satellite-based DETER system issues alerts to enforce deforestation laws in the Amazon.	Brazil	Satellite deforestation monitoring	Coercive capacity	Non-Fragile
Axbard and Deng (2024)	Real-time pollution monitoring improves enforcement of environmental regulations in China.	China	Real-time SMS pollution/compliance monitoring	Coercive / Administrative capacity	Non-Fragile
Banerjee et al. (2020)	E-governance reform with e-invoicing in NREGS; reduces leakage and corruption, improves transparency.	India	E-governance / e-invoicing	Administrative capacity	Non-Fragile
Banerjee et al. (2023*a*)	Electronic food vouchers in Indonesia increase quantity and quality of food received by the poor.	Indonesia	Electronic food vouchers	Service delivery / Administrative capacity	Non-Fragile
Barnwal (2024)	Direct Benefit Transfer for cooking fuel subsidies; reduces leakage by sending money directly to bank accounts.	India	Direct Benefit Transfer (DBT)	Service delivery / Administrative capacity	Non-Fragile
Bellon et al. (2022)	VAT e-invoicing reform in Peru; increases reported sales, purchases and tax liabilities by moving to electronic invoicing.	Peru	VAT e-invoicing	Fiscal / Administrative capacity	Non-Fragile
Beraja et al. (2023)	Facial-recognition AI in China: unrest leads to AI adoption; AI suppresses unrest and stimulates AI innovation, reinforcing autocratic control.	China	Facial-recognition AI for social control	Coercive capacity	Non-Fragile
Borcan, Lindahl and Mitrut (2017)	CCTV and punishment threats reduce corruption in high-stakes exams in Romania.	Romania	CCTV monitoring	Coercive / Administrative capacity	Non-Fragile
Callen et al. (2020)	Health inspectors in Pakistan use a smartphone app; inspections increase initially but effects decline; supervisory dashboards improve doctor attendance.	Pakistan	Health inspection smartphone app	Administrative capacity	Fragile
Carney et al. (2025)	SMS reminders and incentives for COVID-19 vaccination in Kenya; no significant increase in second-dose completion.	Kenya	SMS vaccination reminders	Service delivery capacity	Fragile
Cohen (2024)	SMS tax reminders in Uganda increase compliance, especially in areas with weak state presence — technology extends fiscal extraction capacity.	Uganda	SMS tax reminders	Fiscal capacity	Fragile

Study Title	Description	Country	Technology Evaluated	State-Capacity Type	Fragility
Conover, Kraynak and Singh (2023)	Traffic CCTV increases on-the-ground ticketing and reduces passive corruption in policing.	India	Traffic cameras (CCTV)	Coercive capacity	Non-Fragile
Dal Bó et al. (2021)	Digital monitoring system strengthened administrative capacity, but its effectiveness depended on supervisors' information and decentralized allocation.	Paraguay	Digital monitoring and performance management system for agricultural extension agents	Administrative Capacity	Non-Fragile
Debnath, Nilayamgode and Sekhri (2023)	IVR-based monitoring system for welfare delivery; reduces leakage by frequent reporting from last-mile agents.	India	Interactive Voice Response (IVR) system	Administrative capacity	Non-Fragile
Deininger and Goyal (2012)	Computerization of land registry offices in Andhra Pradesh improves urban credit access by reducing transaction frictions and improving land record access.	India	Land registry computerization	Fiscal / Administrative capacity	Non-Fragile
Dhaliwal and Hanna (2017)	Biometric attendance for health workers increases presence; limited enforcement uptake constrains impact.	India	Biometric attendance monitoring	Administrative / Coercive capacity	Non-Fragile
Duflo, Hanna and Ryan (2012)	Camera-based teacher attendance monitoring + salary incentives reduce absenteeism and increase test scores.	India	Camera-based attendance monitoring	Administrative / Coercive capacity	Non-Fragile
Fujiwara (2015)	Electronic voting reduces residual votes, enfranchises poorer voters, shifts spending toward health, improves infant health.	Brazil	Electronic voting	Administrative capacity	Non-Fragile
Gallego and Ortega (2022)	Tax compliance experiment using Facebook ads + email reminders in Caracas; evaluates whether digital communication increases tax revenue.	Venezuela	Facebook ads / email tax reminders	Fiscal capacity	Fragile
Gallegos et al. (2023)	Digital health appointment system; increases uptake of preventive health services among women in Uruguay.	Uruguay	Digital appointment system	Service delivery capacity	Non-Fragile
Greenstone et al. (2022)	Automated pollution monitors increased accurate reporting of PM10 levels and reduced data tampering.	China	Automated air pollution monitoring	Administrative capacity	Non-Fragile
Kochar (2018)	Branchless banking/mobile agents expand doorstep financial access; increases savings in rural India.	India	Branchless banking / mobile agents	Service delivery / Fiscal capacity	Non-Fragile
Kotsogiannis et al. (2025)	E-invoicing in Rwanda increases VAT compliance and improves efficiency of tax audits; especially impacts audited firms.	Rwanda	E-invoicing	Fiscal capacity	Fragile

Study Title	Description	Country	Technology Evaluated	State-Capacity Type	Fragility
Lewis-Faupel et al. (2016)	E-procurement in India and Indonesia improves road quality and reduces delays in public works.	India / Indonesia	E-procurement	Administrative capacity	Non-Fragile
Muralidharan, Niehaus and Sukhtankar (2016)	Biometric smartcards for NREGS and pensions reduce leakage and improve payment delivery.	India	Biometric smartcards	Administrative / Service delivery capacity	Non-Fragile
Muralidharan et al. (2021)	Phone-based monitoring system for NREGS; reduces failed transfers and improves service delivery.	India	Phone-based monitoring	Service delivery / Administrative capacity	Non-Fragile
Muralidharan, Niehaus and Sukhtankar (2023)	Improved implementation of India's NREGS public employment program; increases earnings and reduces poverty.	India	Public employment program reform	Service delivery capacity	Non-Fragile
Muralidharan, Niehaus and Sukhtankar (2025)	Biometric ID and stringent verification in India's welfare programs reduce corruption but exclude legitimate beneficiaries due to implementation challenges.	India	Biometric ID standards in welfare	Administrative / Coercive capacity	Non-Fragile
Okunogbe and Pouliquen (2022)	Electronic tax filing in Tajikistan reduces compliance time and increases tax paid by high-evasion firms; reduces bribe demands.	Tajikistan	Electronic tax filing (e-filing)	Fiscal / Administrative capacity	Fragile
Tohari, Parsons and Rammohan (2019)	Unified Targeting System increases receipt of complementary programs; highlights program complementarities.	Indonesia	Unified Social Registry / targeting system	Administrative capacity	Non-Fragile
Xu (2021)	Digital surveillance in China enables selective repression and reduces public goods provision.	China	Digital surveillance systems	Coercive capacity	Non-Fragile
Yang et al. (2024)	Automated pollution monitoring improves regulatory enforcement and reduces pollution near monitors.	China	Pollution monitoring systems	Administrative / Coercive capacity	Non-Fragile

A 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.

A.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.21, 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.[51,52] Their salaries amount to about 3.3 million AFN (43,000 USD using the 2018/19 exchange rate) or 2.6% of the total monthly wage bill in these schools.

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

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.[53] Panel B of Appendix Table A.21 (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.[54] However,

[51]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%.

[52]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.

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

[54]The incomplete removal of non-registered employees limits the effect of biometric registration on the payroll. In Appendix Table A.22, we test for post-registration reductions in the total wage bill aggregating 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

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% (1.3 million AFN or 17,000 USD) of the total monthly wage bill in the EE and ED registration zones.

A.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 Appendix Figure A.4). Recognizing these limitations, we use litmus scores to estimate bounds on the share of stand-ins. The idea is simple: the lower the scores of registered employees are with respect to the scores of known true employees (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 $R = T + Z$.

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

group are equal to the standard turnover rates observed in the control group.

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 not 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 exists 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 11.7% of registered employees ($1-\frac{15.51}{17.56}$, 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 8.2% ($1 - \frac{66.37}{72.26}$), and simply answering the phone in a bound of 5.7% ($1 - \frac{80.93}{85.78}$). That is, we find that estimates of stand-ins range from 5.7% to 11.7% of registered employees. Adjusting these numbers for the share of registered employees (97.2%), the bounds range from 5.5% to 11.3% of all employees. Adding these potential stand-ins to the 2.8% of unregistered employees, we estimate that between 8.3% and 14.1% of all paid employees might be ghosts. If they were all removed from the payroll, the government would save 18 million AFN (0.24 million USD) per month in the early registration zones (calculated as 14.1% of a monthly wage bill of about 126 million AFN).

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:[55]

[55] 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.

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 seven 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 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.

B 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:

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

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

Appendix Table A.23 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.24. 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.

C Extended Literature Review

Methodology for Literature Review

This paper reports on an experiment evaluating whether modern technologies can strengthen core administrative functions in an extremely fragile state. To situate our contribution, we search 18 of the top general interest and field-relevant economics journals and the top 3 political science journals to identify studies in three main categories of research:

1. Studies conducted in or focused on fragile states.

2. Studies that centrally examine state capacity.

3. Studies that evaluate technologies or technological systems.

To ensure that no relevant studies are missed, our search strategy combines both Boolean keyword searches and large language model (LLM)–based classification. First, we conduct a boolean, key-word based search in the Web of Science (WoS) database. In parallel, we assemble a separate corpus of articles from leading journals and use an LLM to classify each article based on its abstract, title, and JEL codes, identifying whether it studies state capacity, whether it evaluates a technology *per se*, and which country it focuses on. We then take the union of articles found through both approaches. Because each method generates some false positives, we manually review the combined list to produce the final list of studies.

We then further narrow our review by identifying studies that cover at least two of these three categories, yielding the following groups: (i) Fragility & State Capacity, (ii) Fragility & Technology, and (iii) State Capacity & Technology.

Defining Fragility

To identify whether a study was conducted in a fragile context, we use country classifications from the OECD States of Fragility reports for the available years - 2018, 2020, 2022, and 2025. These reports provide annual lists of countries categorized as either *fragile* or *extremely fragile*. Across all editions in this period, Afghanistan is consistently classified as extremely fragile.

We compile a cumulative list of countries appearing in either category. If a country was classified as *fragile* [56] and later as *extremely fragile* [57] over this period, we retain the more severe label. This is a more lenient approach that allows us to capture whether a study took place in a context that was at any point during this window considered fragile, without making assumptions about the timing of the study itself. This allows us to include a wider body of relevant work while maintaining a consistent external criterion for fragility.

[56] Fragile: Angola, Bangladesh, Burkina Faso, Cameroon, Comoros, Congo, Côte d'Ivoire, Djibouti, Egypt, Equatorial Guinea, Eswatini (Swaziland), Gambia, Guatemala, Guinea, Guinea-Bissau, Honduras, Iran, Kenya, Lao PDR, Liberia, Libya, Madagascar, Malawi, Mauritania, Mozambique, Myanmar, Nepal, Niger, Nigeria, Pakistan, Papua New Guinea, Rwanda, Sierra Leone, Solomon Islands, Tajikistan, Tanzania, Timor-Leste, Togo, Uganda, Venezuela, West Bank and Gaza, Zambia, Zimbabwe.

[57] Extremely Fragile: Afghanistan, Central African Republic, Democratic Republic of the Congo, Somalia, South Sudan, Yemen, Syria, Iraq, Haiti, Mali, Chad, Burundi, Eritrea, Ethiopia, Sudan.

Defining State Capacity

We define state capacity as a state's ability to implement the policy it chooses effectively. This includes administrative ability, coercive power, fiscal effectiveness, and service delivery.

A study is included under the state capacity category if it meets at least one of the following criteria:

- Theorizes about state capacity.
- Measures or quantifies capacity empirically.
- Identifies causal effects involving variation in state capacity or capacity itself as an outcome.

We exclude studies that reference state capacity only rhetorically or in passing.

Defining Technology Evaluation

We include studies that evaluate a specific technology or technological systems, such as biometric tools, digital ID systems, algorithmic management, or e-governance platforms.

To qualify, the technology must be a central object of empirical analysis, and the study must provide evidence (descriptive, correlational, or causal) about the effects of this technology. We include all types of technology where the state would lead implementation (i.e., those that would fall to finance ministries to fund). So, for example, the election monitoring technology evaluated in Afghanistan in Callen and Long (2015) would not qualify, as independent election monitoring is not typically the responsibility of the state.

Search Strategy

To systematically identify relevant studies, we follow a three-step strategy. First, we run a structured Boolean keyword search in the Web of Science database. Second, we independently apply a large language model (LLM)–based classification to the full metadata of the corpus of articles in our target journals. These two methods serve as cross-validation tools to ensure that studies are not missed due to keyword choice or model error. Third, we take the union of all studies identified by either approach and manually review each to confirm inclusion.

We impose three constraints across both methods:

- Journal Scope: We restrict our search to 18 of the top and field-relevant economics journals and the top 3 political science journals. These journals are: American Economic Review, American Economic Review: Insights, Quarterly Journal of Economics, Journal of Political Economy, Journal of Political Economy: Microeconomics, Journal of Political Economy: Macroeconomics, Econometrica, Review of Economic Studies, Review of Economics and Statistics, Journal of the European Economic Association, American Economic Journal: Applied Economics, American Economic Journal: Economic Policy, American Economic Journal: Macroeconomics, American Economic Journal: Microeconomics, Journal of Development Economics, Journal of Public Economics, Economic Development and Cultural Change, Economic Journal, American Political Science Review, American Journal of Political Science, and Journal of Politics.

- Time Frame: We limit our review to publications from 2000 to 2025.
- Country Constraint: We restrict our sample to countries whose modal income classification between 2010 and 2023 – based on the World Bank's income groupings in the "Country Analytical History" sheet of the World Development Indicators – was Low Income, Lower-Middle Income, or Upper-Middle Income.

Boolean Search on Web of Science

We conduct a structured keyword search using the Web of Science (WoS) database. Using Boolean operators, we construct queries based on thematic keywords associated with fragility, state capacity, and technology. For example:

- **Fragility terms:** "fragile state" OR "post-conflict" OR country names tagged as fragile by OECD
- **Capacity terms:** "state capacity" OR "bureaucratic quality" OR "fiscal extraction" OR "public service delivery"
- **Technology terms:** "biometric" OR "digital ID" OR "e-governance" OR "information systems" OR "digital technology"

We provide the full and exact search queries below to enable replication of our search.

LLM-Based Classification

To complement the Boolean approach, we develop a large language model (LLM)-based classification pipeline.

We begin by constructing a metadata corpus of all articles published in the target journals between 2000 and 2025, extracting titles, abstracts, author names, and keywords. We then use a fine-tuned LLM to classify each study along three dimensions:

1. The main country or countries studied.
2. Whether the study is about state capacity.
3. Whether the study evaluates a technology.

This approach adds relevant studies to our corpus that may not use standard keywords – for instance, studies analyzing "Rwandan administrative data" without mentioning "state capacity" or "fragility".

The LLM is validated on a hand-coded dataset of studies independently verified to fall into at least two of the three categories. This benchmark set was used to select the best performing prompts that improves classification accuracy and reduces false positives and negatives.

After initial classification, we filter the corpus to retain only studies at the intersection of at least two categories. We then re-run the model over this filtered subset to verify consistency. Remaining ambiguous cases are manually reviewed and coded.

Studies of Technology and State Capacity

Table A.27 provides a structured summary of all studies identified through this process described above that regard the use of technology to advance state capacity. For each paper, the table reports the study context, type of technology evaluated, the dimension of state capacity involved, and whether the country is classified as non-fragile, fragile, or extremely fragile. There are no studies of countries in the extremely fragile category.

Web of Science Queries

*Fragile x State Capacity

```
TS = (
  "fragile state*" OR "failed state*" OR "weak state*" OR
  fragility OR "state fragility" OR "state failure" OR
  "institutional collapse" OR "state collapse" OR "institutional weakness" OR
  "state weakness" OR "conflict affected" OR "post conflict" OR
  "post-conflict" OR "civil war" OR insurgency OR
  "government breakdown" OR "absence of state" OR "lack of state capacity" OR
  "weak institutions" OR "ungoverned space*" OR "limited statehood" OR
  "limited governance" OR "government control" OR "state reconstruction" OR
  "institutional resilience" OR "fragility index" OR "state legitimacy" OR
  "fragile peace" OR peacebuilding OR "post-conflict reconstruction" OR
  "fragile context*" OR "fragile and conflict-affected" OR FCAS OR FCS OR
  "Afghanistan" OR "Burundi" OR "Central African Republic" OR
  "Chad" OR "Congo, Dem. Rep." OR "Eritrea" OR
  "Ethiopia" OR "Haiti" OR "Iraq" OR
  "Mali" OR "Somalia" OR "South Sudan" OR
  "Sudan" OR "Syrian Arab Republic" OR "Yemen, Rep." OR
  "Angola" OR "Bangladesh" OR "Burkina Faso" OR
  "Cameroon" OR "Comoros" OR "Congo, Rep." OR
  "Cte d'Ivoire" OR "Djibouti" OR "Egypt, Arab Rep." OR
  "Equatorial Guinea" OR "Eswatini" OR "Gambia, The" OR
  "Guatemala" OR "Guinea" OR "Guinea-Bissau" OR
  "Honduras" OR "Iran, Islamic Rep." OR "Kenya" OR
  "Korea, Dem. Rep." OR "Lao PDR" OR "Liberia" OR
  "Libya" OR "Madagascar" OR "Malawi" OR
  "Mauritania" OR "Mozambique" OR "Myanmar" OR
  "Nepal" OR "Niger" OR "Nigeria" OR
  "Pakistan" OR "Papua New Guinea" OR "Rwanda" OR
  "Sierra Leone" OR "Solomon Islands" OR "Tajikistan" OR
  "Tanzania" OR "Timor-Leste" OR "Uganda" OR
  "Venezuela, RB" OR "West Bank and Gaza" OR "Zambia" OR
  "Zimbabwe"
)

AND
```

```
TS = (
  "technology" OR "technologies" OR "digital" OR
  "digitization" OR "digitalization" OR "mobile" OR
  "ICT" OR "information and communication technology" OR
  "e-governance" OR "e-government" OR "digital governance" OR
  "tech-enabled" OR "computerized" OR "online" OR
  "digital transformation" OR "electronic systems" OR "electronic platform" OR
  "electronic services" OR "biometric" OR "biometrics" OR
  "smartcard" OR "smartcards" OR "digital ID" OR
  "national ID" OR "Aadhaar" OR "e-KYC" OR
  "identification system" OR "identity verification" OR "authentication" OR
  "electronic benefits transfer" OR "cash transfer platform" OR "mobile money" OR
  "mobile payments" OR "payment infrastructure" OR "government-to-person payments
     " OR
  "G2P payments" OR "welfare technology" OR "subsidy platform" OR
  "e-wallet" OR "digital cash transfer" OR "electronic filing" OR
  "e-filing" OR "e-invoicing" OR "e-tax" OR
  "digital tax system" OR "digital receipts" OR "point-of-sale data" OR
  "POS data" OR "online tax payment" OR "digital tax registry" OR
  "fiscal registry" OR "taxpayer database" OR "taxpayer identification" OR
  "property mapping" OR "tax automation" OR "e-audit" OR
  "audit trail" OR "blockchain" OR "tax technology" OR
  "remote sensing" OR "satellite imagery" OR "satellite monitoring" OR
  "geospatial data" OR "administrative data systems" OR "management information
     system" OR
  "MIS" OR "real-time dashboard" OR "digital audit" OR
  "e-monitoring" OR "mobile monitoring" OR "SMS tracking" OR
  "IVR monitoring" OR "crowdsourcing" OR "mobile survey" OR
  "field audit tool" OR "digital grievance redressal" OR "digital M&E" OR
  "performance tracking system" OR "electronic monitoring" OR "data dashboard" OR
  "automated supervision" OR "artificial intelligence" OR "AI" OR
  "machine learning" OR "predictive targeting" OR "predictive analytics" OR
  "algorithmic targeting" OR "algorithmic allocation" OR "automated decision-
     making" OR
  "digital tools for targeting" OR "data-driven targeting" OR "AI-driven
     governance" OR
  "ML models in service delivery" OR "automated detection" OR "algorithmic
     detection" OR
  "fraud detection algorithm" OR "mobile phone" OR "smartphone" OR
  "feature phone" OR "SMS" OR "WhatsApp" OR
  "tablet" OR "IVR" OR "voice-based system" OR
  "phone-based system" OR "call center technology" OR "push notification" OR
  "mobile app" OR "digital interface" OR "mobile application" OR
  "hotline system" OR "broadband" OR "internet" OR
  "telecom" OR "telecom infrastructure" OR "4G" OR
  "5G" OR "mobile network" OR "connectivity" OR
  "network coverage" OR "digital penetration" OR "digital divide" OR
```

```
  "mobile signal" OR "digital access" OR "rural connectivity" OR
  "internet expansion" OR "ICT infrastructure" OR "digital rollout" OR
  "telecom expansion"
)

AND

SO = (
  "American Economic Review" OR "American Economic Review: Insights" OR
  "Quarterly Journal of Economics" OR "Journal of Political Economy" OR
  "Journal of Political Economy: Microeconomics" OR "Journal of Political Economy
     : Macroeconomics" OR
  "Econometrica" OR "Review of Economic Studies" OR
  "Review of Economics and Statistics" OR "Journal of the European Economic
     Association" OR
  "American Economic Journal: Applied Economics" OR "American Economic Journal:
     Economic Policy" OR
  "American Economic Journal: Macroeconomics" OR "American Economic Journal:
     Microeconomics" OR
  "Journal of Development Economics" OR "Journal of Public Economics" OR
  "Economic Development and Cultural Change" OR "Economic Journal" OR
  "American Political Science Review" OR "American Journal of Political Science"
     OR
  "Journal of Politics"
)

AND

PY = (2000-2025)
```

*Fragile States x Technology

```
TS = (
  "fragile state*" OR "failed state*" OR "weak state*" OR
  fragility OR "state fragility" OR "state failure" OR
  "institutional collapse" OR "state collapse" OR "institutional weakness" OR
  "state weakness" OR "conflict affected" OR "post conflict" OR
  "post-conflict" OR "civil war" OR insurgency OR
  "government breakdown" OR "absence of state" OR "lack of state capacity" OR
  "weak institutions" OR "ungoverned space*" OR "limited statehood" OR
  "limited governance" OR "government control" OR "state reconstruction" OR
  "institutional resilience" OR "fragility index" OR "state legitimacy" OR
  "fragile peace" OR peacebuilding OR "post-conflict reconstruction" OR
  "fragile context*" OR "fragile and conflict-affected" OR FCAS OR FCS OR
  "Afghanistan" OR "Burundi" OR "Central African Republic" OR
  "Chad" OR "Congo, Dem. Rep." OR "Eritrea" OR
  "Ethiopia" OR "Haiti" OR "Iraq" OR
  "Mali" OR "Somalia" OR "South Sudan" OR
```

```
  "Sudan" OR "Syrian Arab Republic" OR "Yemen, Rep." OR
  "Angola" OR "Bangladesh" OR "Burkina Faso" OR
  "Cameroon" OR "Comoros" OR "Congo, Rep." OR
  "Cte d'Ivoire" OR "Djibouti" OR "Egypt, Arab Rep." OR
  "Equatorial Guinea" OR "Eswatini" OR "Gambia, The" OR
  "Guatemala" OR "Guinea" OR "Guinea-Bissau" OR
  "Honduras" OR "Iran, Islamic Rep." OR "Kenya" OR
  "Korea, Dem. Rep." OR "Lao PDR" OR "Liberia" OR
  "Libya" OR "Madagascar" OR "Malawi" OR
  "Mauritania" OR "Mozambique" OR "Myanmar" OR
  "Nepal" OR "Niger" OR "Nigeria" OR
  "Pakistan" OR "Papua New Guinea" OR "Rwanda" OR
  "Sierra Leone" OR "Solomon Islands" OR "Tajikistan" OR
  "Tanzania" OR "Timor-Leste" OR "Uganda" OR
  "Venezuela, RB" OR "West Bank and Gaza" OR "Zambia" OR
  "Zimbabwe"
)

AND

TS = (
  "state capacit*" OR "state capability" OR "state effectiveness" OR
  "bureaucratic capacity" OR "administrative capacity" OR "bureaucratic autonomy"
     OR
  "meritocratic bureaucracy" OR "fiscal capacity" OR "tax capacity" OR
  "tax administration" OR "revenue mobilization" OR "coercive capacity" OR
  "enforcement capacity" OR "monopoly of violence" OR "regulatory capacity" OR
  "monitoring and enforcement" OR "implementation capacity" OR "policy
     implementation" OR
  "public service delivery" OR "service delivery" OR "infrastructure delivery" OR
  "state building" OR "institutional capacity" OR "civil service reform" OR
  "capacity-building reform" OR "frontline worker*" OR "technical capacity" OR
  "public administration" OR "state authority" OR "state functionality" OR
  "stationary bandit*" OR "informal governance" OR "local administration" OR
  "traffic police"
)

AND

SO = (
  "American Economic Review" OR "American Economic Review: Insights" OR
  "Quarterly Journal of Economics" OR "Journal of Political Economy" OR
  "Journal of Political Economy: Microeconomics" OR "Journal of Political Economy
     : Macroeconomics" OR
  "Econometrica" OR "Review of Economic Studies" OR
  "Review of Economics and Statistics" OR "Journal of the European Economic
     Association" OR
```

```
  "American Economic Journal: Applied Economics" OR "American Economic Journal:
     Economic Policy" OR
  "American Economic Journal: Macroeconomics" OR "American Economic Journal:
     Microeconomics" OR
  "Journal of Development Economics" OR "Journal of Public Economics" OR
  "Economic Development and Cultural Change" OR "Economic Journal" OR
  "American Political Science Review" OR "American Journal of Political Science"
     OR
  "Journal of Politics"
)

AND

PY = (2000-2024)
```

State Capacity x Technology

```
TS = (
  "fragile state*" OR "failed state*" OR "weak state*" OR
  fragility OR "state fragility" OR "state failure" OR
  "institutional collapse" OR "state collapse" OR "institutional weakness" OR
  "state weakness" OR "conflict affected" OR "post conflict" OR
  "post-conflict" OR "civil war" OR insurgency OR
  "government breakdown" OR "absence of state" OR "lack of state capacity" OR
  "weak institutions" OR "ungoverned space*" OR "limited statehood" OR
  "limited governance" OR "government control" OR "state reconstruction" OR
  "institutional resilience" OR "fragility index" OR "state legitimacy" OR
  "fragile peace" OR peacebuilding OR "post-conflict reconstruction" OR
  "fragile context*" OR "fragile and conflict-affected" OR FCAS OR FCS OR
  "Afghanistan" OR "Burundi" OR "Central African Republic" OR
  "Chad" OR "Congo, Dem. Rep." OR "Eritrea" OR
  "Ethiopia" OR "Haiti" OR "Iraq" OR
  "Mali" OR "Somalia" OR "South Sudan" OR
  "Sudan" OR "Syrian Arab Republic" OR "Yemen, Rep." OR
  "Angola" OR "Bangladesh" OR "Burkina Faso" OR
  "Cameroon" OR "Comoros" OR "Congo, Rep." OR
  "Cte d'Ivoire" OR "Djibouti" OR "Egypt, Arab Rep." OR
  "Equatorial Guinea" OR "Eswatini" OR "Gambia, The" OR
  "Guatemala" OR "Guinea" OR "Guinea-Bissau" OR
  "Honduras" OR "Iran, Islamic Rep." OR "Kenya" OR
  "Korea, Dem. Rep." OR "Lao PDR" OR "Liberia" OR
  "Libya" OR "Madagascar" OR "Malawi" OR
  "Mauritania" OR "Mozambique" OR "Myanmar" OR
  "Nepal" OR "Niger" OR "Nigeria" OR
  "Pakistan" OR "Papua New Guinea" OR "Rwanda" OR
  "Sierra Leone" OR "Solomon Islands" OR "Tajikistan" OR
  "Tanzania" OR "Timor-Leste" OR "Uganda" OR
```

```
  "Venezuela, RB" OR "West Bank and Gaza" OR "Zambia" OR
  "Zimbabwe"
)

AND

TS = (
  "state capacity" OR "state capability" OR "state effectiveness" OR
  "bureaucratic capacity" OR "administrative capacity" OR "bureaucratic autonomy"
      OR
  "meritocratic bureaucracy" OR "fiscal capacity" OR "tax capacity" OR
  "tax administration" OR "revenue mobilization" OR "coercive capacity" OR
  "enforcement capacity" OR "monopoly of violence" OR "regulatory capacity" OR
  "monitoring and enforcement" OR "implementation capacity" OR "policy
     implementation" OR
  "public service delivery" OR "service delivery" OR "infrastructure delivery" OR
  "state building" OR "institutional capacity" OR "civil service reform" OR
  "capacity-building reform" OR "frontline worker*" OR "technical capacity" OR
  "public administration" OR "state authority" OR "state functionality" OR
  "stationary bandit*" OR "informal governance" OR "local administration" OR
  "traffic police" OR "health services"
)

AND

TS = (
  "technology" OR "technologies" OR "digital" OR
  "electronic" OR "digitization" OR "digitalization" OR
  "mobile" OR "ICT" OR "information and communication technology" OR
  "e-governance" OR "e-government" OR "digital governance" OR
  "tech-enabled" OR "computerized" OR "online" OR
  "digital transformation" OR "electronic systems" OR "electronic platform" OR
  "electronic services" OR "biometric" OR "biometrics" OR
  "smartcard" OR "smartcards" OR "digital ID" OR
  "national ID" OR "Aadhaar" OR "e-KYC" OR
  "identification system" OR "identity verification" OR "authentication" OR
  "electronic benefits transfer" OR "cash transfer platform" OR "mobile money" OR
  "mobile payments" OR "payment infrastructure" OR "government-to-person payments
     " OR
  "G2P payments" OR "welfare technology" OR "subsidy platform" OR
  "e-wallet" OR "digital cash transfer" OR "electronic filing" OR
  "e-filing" OR "e-invoicing" OR "e-tax" OR
  "digital tax system" OR "digital receipts" OR "point-of-sale data" OR
  "POS data" OR "online tax payment" OR "digital tax registry" OR
  "fiscal registry" OR "taxpayer database" OR "taxpayer identification" OR
  "property mapping" OR "tax automation" OR "e-audit" OR
  "audit trail" OR "blockchain" OR "tax technology" OR
```

```
  "remote sensing" OR "satellite imagery" OR "satellite monitoring" OR
  "geospatial data" OR "administrative data systems" OR "management information
     system" OR
  "MIS" OR "real-time dashboard" OR "digital audit" OR
  "e-monitoring" OR "mobile monitoring" OR "SMS tracking" OR
  "IVR monitoring" OR "crowdsourcing" OR "mobile survey" OR
  "field audit tool" OR "digital grievance redressal" OR "digital M&E" OR
  "performance tracking system" OR "electronic monitoring" OR "data dashboard" OR
  "automated supervision" OR "artificial intelligence" OR "AI" OR
  "machine learning" OR "predictive targeting" OR "predictive analytics" OR
  "algorithmic targeting" OR "algorithmic allocation" OR "automated decision-
     making" OR
  "digital tools for targeting" OR "data-driven targeting" OR "AI-driven
     governance" OR
  "ML models in service delivery" OR "automated detection" OR "algorithmic
     detection" OR
  "fraud detection algorithm" OR "mobile phone" OR "smartphone" OR
  "feature phone" OR "SMS" OR "WhatsApp" OR
  "tablet" OR "IVR" OR "voice-based system" OR
  "phone-based system" OR "call center technology" OR "push notification" OR
  "mobile app" OR "digital interface" OR "mobile application" OR
  "hotline system" OR "broadband" OR "internet" OR
  "telecom" OR "telecom infrastructure" OR "4G" OR
  "5G" OR "mobile network" OR "connectivity" OR
  "network coverage" OR "digital penetration" OR "digital divide" OR
  "mobile signal" OR "digital access" OR "rural connectivity" OR
  "internet expansion" OR "ICT infrastructure" OR "digital rollout" OR
  "telecom expansion"
)

AND

SO = (
  "American Economic Review" OR "American Economic Review: Insights" OR
  "Quarterly Journal of Economics" OR "Journal of Political Economy" OR
  "Journal of Political Economy: Microeconomics" OR "Journal of Political Economy
     : Macroeconomics" OR
  "Econometrica" OR "Review of Economic Studies" OR
  "Review of Economics and Statistics" OR "Journal of the European Economic
     Association" OR
  "American Economic Journal: Applied Economics" OR "American Economic Journal:
     Economic Policy" OR
  "American Economic Journal: Macroeconomics" OR "American Economic Journal:
     Microeconomics" OR
  "Journal of Development Economics" OR "Journal of Public Economics" OR
  "Economic Development and Cultural Change" OR "Economic Journal" OR
```

```
  "American Political Science Review" OR "American Journal of Political Science"
     OR
  "Journal of Politics"
)

AND

PY = (20002024)
```

*Prompt for the LLM Classifier

```
# ---- prompt ----
prompt <- glue::glue("

Given the **title** and **abstract** of an academic article, do four things:


1. Identify the **single main country** the article is about.
    Only choose from this list: {states_txt}
    Apply these principles:
      If the country name appears and the article studies that countrys
       policies, context or people, use it.
      If a demonym/adjective appears (\"Swedish\"  Sweden, \"US\"  United States),
       map it **only if** it is the subject of study.
      Infer the country from sub-national clues (e.g. Medicaid  United States,
       Addis Ababa  Ethiopia) **only if** it is clear.
      If several countries are mentioned, pick the one that is **analysed most
       deeply**.
      If you cannot be 90 % confident, return ‘None‘.

2. Decide whether the article is **centrally about state capacity**.
    Use this definition:

       > **State capacity** is the ability of the state to *implement the
       > policies it chooses*that is, to translate political decisions into
       > effective action through administration, coercion, fiscal extraction
       > and service delivery.

    The study counts as a state-capacity study **only if** it analyses,
     measures, theorises or causally identifies one or more of the following
     dimensions **as a primary object of inquiry** (not merely background):

        Administrative or bureaucratic effectiveness
        Fiscal capacity (tax collection, revenue mobilisation)
        Coercive or policing capacity (monopoly of violence, enforcement)
        Regulatory capacity (ability to enforce rules/standards)
        Service-delivery capacity (health, education, infrastructure, etc.)
        Capacity-building reforms or institutional strengthening programmes
```

```
    **Exclude** studies that merely:
        Mention a ministry, agency or government programme in passing
        Analyse political preferences, elections, regime type or corruption
         **without** evaluating a capacity dimension listed above
        Study policy *choice* (agenda-setting) rather than *implementation*

3. Determine whether the article evaluates a specific technology or technological
     system.

    Count **true** only if the study empirically evaluates the *use,
      implementation, or effects* of a real-world technologysuch as digital ID
      systems, biometric tools, mobile payments, e-government portals,
      algorithmic policy tools, or electronic surveillance systems.

    Focus on cases where the **technology is a central object of analysis**either
      descriptively or causallyand where the study generates evidence about how
      the technology affects behavior, outcomes, or institutions.

    Do **not** count studies where:
      Technology is discussed only in abstract terms (e.g., technological change,
        innovation, productivity in macro/growth models).
      The focus is on econometric or statistical methods (e.g., GARCH, IV, ML)
        that are **not themselves deployed in institutional or policy settings**.
      The technology (e.g., satellite data, algorithms, sensors) is **only used to
         collect data** or select subjectswithout being analyzed as a treatment
        or evaluated in terms of effectiveness.
      Algorithms or predictive tools are **simulated or modeled**, but not
        empirically implemented or studied in real-world settings.

4. Output the following exact JSON (no markdown, no extra keys):
{
  \"identified_country\": \"<country or None>\",
  \"is_state_capacity_study\": <true | false>,
  \"is_technology_study\": <true | false>
}

### Article Title: {title}
### Abstract: {abstract}
")
```