Incentivizing Frontline Workers and their Supervisors: Experimental Evidence from Agricultural Extension in Pakistan

Garance Genicot*, Zahra Mansoor§, and Ghazala Mansuri †.

August 2024 (DRAFT - PLEASE DO NOT CIRCULATE)

Abstract

Public sector service delivery often relies on frontline staff operating within complex bureaucracies, yet most incentive schemes target frontline workers alone, overlooking the role of middle management. We provide experimental evidence on the effectiveness of integrating supervisor incentives with frontline performance pay in the context of agricultural extension services in Punjab, Pakistan. Partnering with the Punjab Agriculture Extension Department, we randomly assigned tehsils to one of three pay-for-performance schemes targeting both frontline extension agents and their direct supervisors: (i) an Objective scheme linking bonuses to digital performance metrics, (ii) a Subjective scheme based on supervisors' discretionary evaluations, and (iii) a Subjective Plus scheme combining supervisor discretion with light-touch monitoring of supervisory performance. We find that all schemes improve extension outreach, particularly staff attendance and farmer visits, with the Objective scheme showing the strongest impact on direct farmer outreach. However, improvements in outreach effort on the extensive margin, extension quality, and farmer access to extension are only consistently observed in the Subjective Plus scheme. This effect is driven by greater supervisory engagement, with supervisors in Subjective Plus assigning more group-based trainings (Farmer Training Programs) in more unique villages. Further analysis shows that these supervisors also allocate bonuses based on both observable performance metrics and privately known indicators of staff quality, while placing less weight on social ties. Our findings highlight the importance of complementing frontline incentives with mechanisms to engage and monitor middle managers, enhancing both managerial effort and the effectiveness of performance pay schemes.

JEL Classification Numbers: O12, O13, Q16

Keywords: Agricultural Extension, Incentive schemes, Objective performance scheme, Subjective performance scheme, Bonus.

Acknowledgements: We are indebted to Captain Mehmood (former Secretary Agriculture) and Rashid Langrial (former Secretary Agriculture) for championing the project. We are also grateful to Dr. Anjum

Ali (Director General Extension) for his support and numerous insights over the last four years, and to participants at the APSA conference (2020) and CSAE workshops (2019, 2020), Clare Leaver, Christopher Woodruff, Simon Quinn, Julien Labonne, Stefan Dercon, and Martin Williams. Ayesha Shahid provided invaluable research and implementation support. We are also thankful to Saad Imtiaz, Usman Zahid, Maria Qazi and Bhavya Srivastava for research assistance. This research was made possible through ongoing support and insights from various counterparts at the Extension Department, the Agriculture Delivery Unit, and the Agriculture Ministry during the last four years. We gratefully acknowledge funding from the World Bank i2i, World Bank Strategic Research Programme, Georgetown Center for Economic Research and the International Growth Center. AEA registry number: AEARCTR-0002430. Ethical approval was acquired through the University of Georgetown.

§ University of Bristol; ∗ Georgetown University; † The World Bank.

1 Introduction

A high-performing public sector workforce is essential for fostering growth and development, yet achieving this remains a persistent challenge across many settings (Chaudhury et al., 2006). Well-designed incentive schemes have shown promise in enhancing employee performance (Duflo et al., 2012; Khan et al., 2016; Mbiti et al., 2019; Leaver et al., 2021), but designing effective incentives within the public sector is inherently complex. Service delivery often depends on large, multilayered bureaucracies where performance relies on complementary efforts across hierarchical levels (Wilson, 1989; Garicano, 2000), raising important questions about how to structure incentives in such systems. Moreover, public sector roles are typically multidimensional, with some tasks easier to measure and incentivize than others (Holmstrom and Milgrom, 1991), highlighting the potential importance of supervisors in capturing and incentivizing harder-to-measure aspects of performance.

These challenges give rise to several competing theoretical predictions. Incentives based on objective, measurable metrics can improve performance in those specific areas (Bandiera et al., 2007; Khan et al., 2016), but may inadvertently divert effort away from less observable tasks (Holmstrom and Milgrom, 1991). Alternatively, subjective evaluations by supervisors can capture a broader range of performance dimensions, yet are susceptible to bias, favoritism, or collusion, potentially undermining their reliability (Gibbs et al., 2004; Prendergast, 2007; Neal, 2011). Introducing oversight of supervisors can help align their incentives with those of the principal, encouraging them to supply critical complementary inputs and improve the quality of performance assessments (Prendergast and Topel, 1996; Dal Bó et al., 2021). Despite the theoretical importance of these mechanisms, empirical evidence on the effectiveness of multilayered incentive structures in enhancing frontline service delivery remains limited, with only a few recent contributions (Behrman et al., 2015; Deserranno et al., 2022).

This paper addresses this gap by asking how an incentive design that targets both frontline extension workers and their supervisors improves service delivery. Specifically, we evaluate (i) the impact of incentives based on objective metrics, (ii) the role of subjective supervisor

evaluations, and (iii) whether light-touch oversight of supervisors enhances the effectiveness of subjective incentives.

Our empirical setting is the Agriculture Extension Department in Punjab, Pakistan, where frontline extension agents are tasked with delivering timely and actionable advice to millions of smallholder farmers. Extension agents - AOs and FAs - core responsibilities include conducting Farmer Training programs (FTPs) to disseminate core information on sowing and harvesting, providing individual farmer advice on agricultural inputs, setting up demonstration plots, and handling departmental tasks such as reporting and office work. The direct supervisors of AOs/FAs in each tehsil are the assistant directors (ADs). They manage a variety of tasks that are essential for frontline extension service delivery. The most critical of these is to set the monthly FTP schedule, and assign these group trainings to staff members based on their location. As such, this hierarchical setting is ideal to study our research questions.

We partnered with the Extension Department to design and assess the impact of a province-wide incentive program targeted at frontline extension staff and their direct supervisors on frontline extension outreach and quality. Prior to the roll-out of the incentive program, the Extension Department, in collaboration with our research team, introduced a smartphone-based monitoring platform (*AgriSmart*) as a way to enhance staff supervision. Agrismart consists of smartphone-based app for daily activity reporting by all field staff and a digital dashboard for supervisors to conduct their managerial tasks. Leveraging the AgriSmart system, a total of 131 tehsils across the province of Punjab were randomly allocated to three pay-for-performance schemes, or a control groups.[1].

In the *Objective* arm, bonuses were allocated based solely on AgriSmart metrics. In the *Subjective* arm, supervisors were free to use the Agrismart portal to assess staff performance but had full discretion over bonus allocation. In the *Subjective Plus* arm, supervisor discretion was retained, but top-down monitoring of supervisors was introduced to align their incentives with departmental goals. To accomplish this, the top leadership of the department

[1]Tehsils are the fourth administrative unit below districts and divisions in Punjab

received a monthly report on the performance of the supervisor's tehsil, featuring AgriSmart based summary statistics on staff and supervisor performance.

We show treatment effects across three main categories of outcomes - extension outreach and quality, farmer outcomes, and supervisor effort. First, we focus on *extension outreach* and show positive and significant treatment effects on staff attendance (days and hours worked) and our measures extension outreach (such as villages visited and farmers visited) across all arms. For other outreach indicators such as village visits and individual farmer visits, the Objective arm significantly outperforms both the Subjective and Subjective Plus arms, intuitive given the incentives in this arm were explicitly linked to the AgriSmart outreach measures. Outreach through group meetings, called Farmer Training Programs (FTPs), is only significant in the Subjective Plus arm. We rule out concerns that these effects could be driven by staff reporting behavior, or gaming of thresholds in the case of the Objective arm.

Despite concerns that incentivizing quantifiable outreach metrics might compromise service quality, the findings show no evidence of such trade-offs. Across all treatment arms—including the Objective scheme—meeting length, used as a proxy for extension quality, increased relative to the control group. Extension staff in the Objective, Subjective, and Subjective Plus arms spent approximately 13–14 additional minutes in group meetings, representing a 20–23% improvement over baseline levels. These results suggest that higher outreach efforts did not reduce—and may have even enhanced—the quality of interactions between extension workers and farmers.

Our results on farmer outcomes focus on farmer access to extension and agricultural outcomes such as farmer knowledge and yield. We find that in comparison to the control group, the Subjective arms increase our farmer access index (comprised of whether farmers know the frontline staff, are informed about extension, and attend extension meetings) by about 0.06 standard deviations. Our treatment effects on farmer knowledge are statistically insignificant and close to zero. While the coefficients for wheat yield are positive in the Subjective and Subjective Plus arms, they are also insignificant. These null effects highlight the limits of performance pay schemes when not accompanied by improvements in the information

content of the services delivered.

While we observe treatment effects on extension outreach across all arms, farmer access to extension only shows improvement in the Subjective arms. We show that these results can be explained by the fact that different treatments encourage different types of effort. We find that the Objective arm encourages more effort on the intensive margin in comparison to the control group, and other treatment arms. Over the course of the program, frontline staff in the Objective arm visited the same farmer at least 3 more times in comparison to the control group, whereas both the Subjective arms visited the same farmer only 1 more time. Frontline staff could have expended effort on the extensive margin by expanding their outreach to new farmers (through individual farmer interactions or group-based FTPs) or villages. In comparison to the control group, all treatment arms reach out to more farmers. However, the Subjective arms engaged more farmers through group activities such as FTPs while the Objective arm mainly relied on individual visits. In the Subjective Plus arm, this effect was particularly pronounced, with more FTPs conducted in unique villages compared to both the control group and other treatment arms.

Next, we investigate why staff in the Subjective Plus arm conducted more FTPs? This brings our attention to the supervisors who are responsible for scheduling FTPs against staff. We show that supervisors in the Subjective Plus arm schedule and assign 6.7% points more Farmer Training Programs (FTPs) to extension agents as compared to the control group (about a 16% increase over the control group). These effects are also significantly different from the Subjective and Objective arms at the 10% level.

In our final analysis, we examine how bonuses are allocated across treatment arms and which staff characteristics predict these allocations, particularly in the subjective schemes. We find that monthly bonus amounts are significantly higher in the Subjective and Subjective Plus arms than in the Objective arm, consistent with the possibility of leniency bias. In the Objective arm, bonuses are primarily based on observable AgriSmart performance metrics such as days and hours worked. While these metrics remain relevant under subjective schemes, we find that Subjective Plus supervisors place greater weight on FTP completion and staff

knowledge, and rely less on social ties such as shared caste compared to the Subjective ADs. This suggests that light-touch monitoring in Subjective Plus helps redirect supervisor discretion toward performance-related factors, while limiting favoritism.

In summary, our results show that all treatment arms improve extension outreach with the Objective arm as the most effective in driving extension outreach with no negative effects on extension quality. However, improvements in outreach effort on the extensive margin, extension quality, and farmer access to extension are consistently observed in the Subjective Plus scheme. This effect is driven by two possible mechanisms. First, supervisors making important supervisory inputs into what extension frontline staff are supposed to do. In particular, they schedule more Farmer Training Programmes which results in frontline staff completing more trainings in more unique villages. Second, we find suggestive evidence that supervisors in Subjective Plus use their discretion to award incentives to staff based on measurable metrics (observable on Agrismart) as well as privately observed metrics of performance quality.

Our research contributes to several themes of literature. It contributes to the literature on performance pay in the public sector, where services are delivered through hierarchical organizations and tasks have observable and non-observable dimensions(see Burgess and Ratto, 2003; Manning et al., 2012;Deserranno et al.,). While existing evidence has considered the impact of Objective and Subjective schemes (Khan et al., 2016; Andrabi and Brown,2024), this paper moves beyond the dichotomy between objective and subjective incentives by unpacking how supervisor engagement and monitoring mediate the effectiveness of hybrid incentive systems. The challenges of discretion in subjective evaluation systems are well documented (Baker et al.,1994; Prendergast, 1999; Gibbs et al., 2004) with theoretical prediction that aligning supervisor incentives with the objectives of the principal could address these challenges (Prendergast and Topel, 1996). We experimentally test this in the field, and show the incentivizing supervisors can enhance the impact of performance incentives for the frontline. By analyzing frontline and supervisor effort, as well as supervisor bonus allocation behavior, we provide insight into how incentivising supervisors through an element of top-down monitoring effects their own effort on complementary inputs and discretion in

bonus assignments.

It also contributes to the literature on agricultural extension. While existing evidence has considered alternate ways of disseminating extension information to farmers to improve technology adoption (Fernando, 2016; Beaman et al., 2018), these studies do not focus on incentivising extension workers.

Finally, Our results also relate to the wider literature on how practices and performance of middle-management relates to frontline service delivery (Rasul and Rogger, 2018; Williams et al., 2020, Leaver et al, 2024). Existing evidence shows how management practices at the national and sub-national (for example, district) level is linked to task completion at these levels. We show direct evidence that task completion at the sub-national level has a direct impact on frontline service delivery.

This paper is structured as follows. Section 2 outlines the contextual setting and Section 3 presents a simple conceptual framework to fix ideas. Section 4 presents the experimental design, randomization, and implementation, and Section 5 presents an outline of the main data sources and outcome measures. Section 6 presents the results, Section 7 presents an analysis of the mechanisms, and Section 8 concludes.

2 Context

Out of the total cultivated land in Pakistan, 57% is located in Punjab, making agriculture a crucial sector for policy reforms in the province.[2] However, agricultural productivity has been declining in recent years, with little or no growth in yields (Punjab Agriculture Policy, 2018; Pakistan Economic Survey Survey, 2019). This decline has been primarily attributed to poor farming practices, low technology adoption, and limited crop diversification (Punjab Agriculture Policy, 2018). Smallholder farmers with less than 5 acres of land comprise over 60% of all farm households and are particularly affected by low yields. Providing these

[2]http://www.cabi.org/projects/project/10880.

farmers with relevant and timely information could greatly enhance agricultural productivity and reduce poverty. Yet access to information remains a major challenge for these remote and small-scale farmers, who primarily rely on public extension services as their main source of advice.

Agriculture Extension Department in Punjab. In Punjab, the Provincial Agriculture Extension Directorate is the primary entity responsible for providing farmers with agricultural information. The department maintains a frontline workforce of about 2,500 workers, with a presence extending down to the village level. The hierarchy begins with Field Assistants (FAs) at the union council level, covering approximately 7–8 villages, and Agricultural Officers (AOs) at the markaz level, responsible for overseeing 7–8 union councils. Assistant Directors (ADs) supervise at the tehsil level, followed by Deputy Directors (DDs) at the district level, and Directors at the divisional level (refer to Figure 1 for a visual representation).

AOs and FAs are the key frontline extension workers. Their responsibilities include conducting group training events called Farmer Training Programs (FTPs), providing individual farmer consultations on agricultural practices and inputs, setting up demonstration plots, and completing departmental reporting and office tasks. AOs and FAs collaborate on FTPs, with AOs serving as technical leads and FAs providing logistical support. Their immediate supervisors are the ADs, who manage core managerial tasks such as setting monthly FTP schedules, assigning trainings based on location, approving staff leave requests, and conducting surprise visits to ensure procedural quality.

Figure 1 about here

Despite this well-established institutional structure, the department has invested little in upgrading its monitoring systems or improving the quality of information disseminated. Public extension services are widely perceived to have been captured by larger landowners, operating under the guise of "lead farmers," which serves the interests of field staff who often lack incentives to reach out to smallholder farmers. Even when small farmers are reached, the information provided is often outdated. As a result, the department faces a twin

challenge: increasing staff effort and improving the relevance of the information provided. Given that extension services also impose significant costs on the government, enhancing their effectiveness is critical for both budget efficiency and agricultural productivity.

The AgriSmart Program. To address these challenges, the department worked closely with our research team from 2015-2018 to design the AgriSmart program from scratch - the first-ever digital platform in the Extension Department involving a smartphone-based staff activity reporting system and a management dashboard for supervisors to conduct key managerial tasks. The smartphone-based component of Agrismart required field staff to use the AgriSmart app to log daily time use, while supervisors used the dashboard for tasks such as assigning FTPs, approving leaves, and making HR updates online.[3]

Deployed across Punjab in April 2018, AgriSmart required field staff to report geo-tagged, time-stamped daily activities, enabling automated calculations of time use, attendance, village visits, farmer visits, and FTP completions. While the system was extensively debugged to incorporate geo-tagged and time-stamped activities to discourage misreporting, a concurrent farmer callback audit was also established to deter misreporting. Figure 2a illustrates extension activity data points over three months in 2019, while Figure 2b shows the homepage of the AgriSmart application.

As part of the AgriSmart rollout, Standard Operating Procedures (SOPs) were also introduced for ADs, including regular FTP scheduling, timely approval of staff leave requests, and accurate updates to staff profiles at the tehsil level.

Figure 2 about here

Travelling and Daily Allowance (TADA). The department also manages a non-salaried budget line called the Travelling and Daily Allowance (TADA) budget, through which staff can claim reimbursement for travel. *De jure*, claims were meant only for out-of-station

[3]The program was supported by the Punjab Public Sector Management Reform Program (PPRMP) and implemented by the Punjab Information Technology Board (PITB), the technology counterpart for the PPRMP.

travel (beyond a 16 km radius from a staff member's assigned jurisdiction). However, *de facto*, TADA disbursements have been ad hoc and inconsistently administered.

Given the improved monitoring enabled by AgriSmart, the department sought to link TADA payments more explicitly to field performance, using this pre-existing budget line to implement the incentive program.

3 Conceptual Framework

The design of incentive schemes in public service delivery is challenging and complex. The objective—social welfare—is an elusive concept that is difficult to measure. Even if regular measurement were feasible, incentivizing outcomes for the service's potential beneficiaries (e.g., farmers' welfare or agricultural output) would introduce significant noise, resulting in a low signal-to-noise ratio (Baker 1992; Gibbons 1998). Consequently, most incentive schemes focus on employee performance.

While Agrismart makes it easier to monitor attendance, activities, and location, qualitative aspects of performance remain harder to assess. Supervisors, at the middle management level, often have better information on these dimensions of performance of front line staff, although their objectives may not always be fully aligned with those of the principal. Moreover, supervisors also supply complementary inputs—such as planning, resource allocation and human resource management for front line staff —that are critical to worker performance, making their own incentives an essential part of overall system performance.

Faced with these challenges, a principal who seeks to use performance incentives to improve service delivery in a public sector context, must choose between an *objective* scheme that rewards easily measurable forms of effort and a *subjective* scheme that relies on supervisor discretion. In what follows, we present a simple conceptual framework that highlights the advantages and disadvantages of each approach. We also emphasize the importance of aligning middle management (in our case, AD) objectives with those of the principal under

subjective schemes. This framework provides the foundation for our experimental design.

A Model

For simplicity, assume that there are two dimensions of worker effort: one dimension (observable and easy to measure) labeled e, and another (harder to observe and measure) called *quality* and denoted by q. Both of these contribute to the creation of social benefit (i.e. farmers' welfare or productivity). An additional input is supervisor effort exerted for scheduling activities, denoted by s.

In each jurisdiction, a levels of outreach e and quality q by a field worker, together with input s from the supervisor, generate a social benefit of $f(e, q, s)$ that is not directly observable. The marginal product of each type of effort is strictly positive and decreasing.

The principal is concerned with maximizing the social benefit generated by the worker $f(e, q, s)$, taking into account the effort costs for the worker $c(e, q)$ and for the supervisor $k(s)$, as well as the costs $\kappa(B)$ of the total bonus disbursed. To do this, he/she chooses an incentive scheme knowing that supervisors and workers will choose their effort levels.

Effectively, an incentive scheme determines the probability $p_i(e_i, q_i, s_{j(i)})$ with which a worker with effort levels (e_i, q_i) gets a bonus b when their supervisor $j(i)$'s scheduling effort is $sj(i)$.

The principal aims to maximize net social surplus:

$$\sum_i f(e_i, q_i, j(i)) - c(e_i, q_i) - k(s_{j(i)}) - \kappa\left(B\right), \tag{1}$$

where $j(i)$ is the supervisor of worker i and $B = \sum_i p_i(e_i, q_i, s_{j(i)}))b$ is the total bonus costs.

Workers are assumed to balance their financial gain (*extrinsic motivation*) with a weight of α, and the social benefits produced by their activities (*intrinsic motivation*) with a weight of $1 - \alpha$. There is also a cost of effort, denoted as $c(e, q)$, which increases with effort. Given a bonus plan $p_i(e, q, s)$ and supervisor effort s, a worker i selects the effort levels (e, q) to

maximize their utility:

$$\alpha_i p_i(e,q,s)b + (1-\alpha_i)f(e,q,s) - c(e,q). \tag{2}$$

Under the regular assumptions (differentiability and concavity), the following first order conditions would characterize the worker's choice:

$$\begin{cases} \alpha_i \frac{\partial p_i}{\partial e} b + (1-\alpha_i)\frac{\partial f}{\partial e} - \frac{\partial c}{\partial e} = 0, \\ \alpha_i \frac{\partial p_i}{\partial q} b + (1-\alpha_i)\frac{\partial f}{\partial q} - \frac{\partial c}{\partial q} = 0. \end{cases}$$

Denote as $(\widehat{e}, \widehat{q})$ the choice that maximize their utility.

Clearly, if $\alpha_i = 0$ then the worker would choose the optimal effort and a zero bonus would be optimal. However, in the absence of a bonus, any worker with $\alpha_i > 0$ chooses a suboptimal level of effort.

Objective Incentives. A bonus increase directly incentivizes outreach e but may increase or decrease q, depending on whether e and q are complements or substitutes. Supervisor effort s can also be indirectly incentivized, assuming complementarities between e and s.

In an objective scheme, p_i depends only on e (and not q), uniformly across workers ($p_i = p$). It is easy to see that if $\alpha_i > 0$, an increase in b increases i's outreach effort $\widehat{e}_i$. Whether it promotes or discourages quality depends on whether e and q are complements or substitutes. If, for instance, the social benefits are independent but $c_{eq} < 0$, then an increase in b decreases $\widehat{q}$. Such natural complementarities are common in public service delivery; e.g. attendance and outreach efforts are necessary for providing advice to farmers. In contrast, a time budget constraint can lead to substitutability between outreach e and quality q ($c_{eq} > 0$). In this case, an increase in b reduces $\widehat{q}$.

Even though supervisors are not directly incentivized under an objective scheme, their scheduling activities could nevertheless be affected. We assume that supervisors have their

own cost of effort $k(s)$ and a utility function with weights β on their private preference component and $1-\beta$ on the net benefit from the principal's perspective. We will delve deeper into supervisors' incentives when subjective schemes are in place, allowing them greater decision-making power. For now, it is sufficient to remark that we view supervisor outreach and activity scheduling as complementary factors for producing a surplus of $f_{es} > 0$. Hence, supervisors could be indirectly incentivized to put in more effort when an objective incentive system is established for workers.

Subjective Incentives. Under a subjective scheme, supervisors are tasked with allocating bonuses among field workers. Supervisors have the same information on e as the principal but are assumed to have access to some information on q.[4] Subject to a feasibility constraint (i.e. conditioning on what is observable), the supervisor assigns, or not, a bonus b to each worker. The mapping from effort to bonus that results from the supervisor's bonus allocation behavior determines $p_i(e, q, s)$ for each individual i. This and the supervisor's scheduling effort s in turn leads to worker effort choices $(\widehat{e}, \widehat{q})$ that maximize their utility.

A supervisor j's preferences are defined over their own effort, a private preference component, and the net benefit from the perspective of the principal. Specifically, let $\beta_j \in (0, 1)$ be the weight that supervisor j puts on their private preferences when allocating bonuses, so that $1-\beta_j$ captures the supervisor alignment with the principal's objective (1). We can think of β_j as either a measure of intrinsic motivation or a measure of career concerns for the supervisors. A reduction in β can be the result of a heightened top-down monitoring of the supervisors.

The private component of j's preference is $V_j(\mathbf{p}) \equiv \sum_i \omega_{ji} u(p_i(e_i, q_i)B)$, where ω_{ji} is the weight that supervisor j assigns to worker i, and $u(0) = 0$, $u' > 0$ and $u'' < 0$. This formulation allows us to capture a wide range of potential supervisors' biases (Gibbs et al., 2004; Prendergast, 2007; Neal, 2011), such as favoritism ($\omega_{ji} > 0$ for some select individuals) or centrality bias (where $\omega_{ji} = \omega_j > 0$ for all i).

[4]Alternatively we could assume that that information comes at some cost.

Supervisor j chooses a scheduling effort s_j and an bonus allocation to maximize:

$$\beta_j V_j(\mathbf{p}) + (1-\beta_j)\left(\sum_i \left(f(e_i, q_i, s_j) - c(e_i, q_i) - \kappa p_i(e_i, q_i) B_i)\right)\right) - k(s_j), \tag{3}$$

knowing that (e_i, q_i) maximizes (2).

Low β_j supervisors (more mission-aligned) are more likely to exert complementary effort s and allocate bonuses to maximize social surplus. Subjective schemes thus create opportunities and risks depending on the extent of supervisor bias.

In addition, supervisors does not internalize the cost of bonuses, and therefore has an incentive to be overly generous when distributing them (the leniency bias).

Testable Implications.

This simple framework highlights the tradeoff between subjective and objective incentive schemes: subjective schemes allow supervisors to condition bonuses on information about q, but also introduce the risk of supervisory bias.

From this framework, we derive several hypotheses:

- **Hypothesis I:** All incentive schemes will increase measurable outreach effort e.
- **Hypothesis II:** Outreach e will improve more under objective than under subjective or no incentive schemes.
- **Hypothesis III:** Harder-to-measure effort q will improve more under subjective schemes.
- **Hypothesis IV:** Supervisor scheduling effort s will increase under subjective schemes when supervisor alignment improves.
- **Hypothesis V:** Outreach e and quality q will both increase under subjective schemes when supervisor alignment improves.

- **Hypothesis VI:** Bonuses will less biased under subjective schemes when supervisor alignment improves.

We also investigate: (i) whether e and q are complements or substitutes; (ii) whether subjective and objective metrics act as complements or substitutes (Gibbs et al. 2004); and (iii) whether incentives operate on the extensive vs. intensive margins of outreach.

4 Design and Randomization

Building on this framework, we designed an experimental intervention to compare objective and subjective incentive schemes and to examine how the alignment of supervisor incentives shapes their effectiveness. In collaboration with the Punjab Agriculture Extension Department, we implemented three distinct incentive schemes targeting frontline extension agents (AOs/FAs) and their immediate supervisors (ADs). The program combined a performance-based bonus for staff with varying levels of supervisory discretion (see Figure 3 below).

Figure 3 about here

In the incentive arms, extension staff could quality for a monthly bonus that had four levels: a high bonus, a low bonus, a base payment or no bonus. The department predetermined the bonus amount at each level, with the base payment set at 7% of the monthly salary. This amount was bench-marked against historical average monthly TA claims made by field staff.[5] The low and high bonuses were set at an additional 7% and 15% of salary, respectively, and were to be paid for good performance (see Table C.1 in Appendix C for incentive amounts for each bonus category). A unique feature of the program was that it was designed to be budget neutral, with all bonus payments made through the preexisting annual TADA budget allocation. This was also intended to ensure that the existing recurrent budget allocation to the department was used more effectively to improve service delivery.

[5]It was initially bench-marked against FY15-16 but in FY18-19 the amounts were revised upwards.

4.1 Treatment and Control Arms

Control Arm: Control Arm: In tehsils assigned to the Control arm, no bonus payments were made. Staff continued to obtain travel and daily allowances as per the existing claim system. The ADs in this arm had access to the AgriSmart dashboard to upload farmer training programs and approve leaves, but could not access staff performance summaries, unlike the treatment arms.

Objective Arm: In tehsils assigned to the Objective arm, the bonus was based solely on performance information available in AgriSmart. To qualify for a base payment, staff had to meet monthly attendance thresholds for days and hours worked. To qualify for a low or high bonus, staff had to meet additional thresholds related to village and farmer visits, distance traveled, time spent on extension activities, and the proportion of completed farmer training programs. The thresholds for meeting base, low and high bonus for AOs and FAs were set by the Extension Department leadership (see details regarding thresholds in Tables C.2 and C.3 in Appendix C). The performance indicators for both AOs and FAs were similar but differed in scale due to their respective geographical scope. A third-party audit system, including callbacks to farmers, was used to verify the veracity of AgriSmart data and discourage misreporting.

ADs (supervisors) in this arm could view the performance metrics and bonus categories of each worker in their tehsil but had no discretion over bonus assignments.

Subjective Arm: In tehsils assigned to the Subjective arm, the AD had access to the AgriSmart portal, so could view AgriSmart based performance indicators of staff in his tehsil. He was also required to use the *staff payment assignment page* on the AgriSmart dashboard for the monthly bonus assignment (see Figure C.4 in Appendix C for a snapshot of the bonus assignment page). However, he had complete discretion in assigning bonus payments.

The potential advantage of the Subjective arm comes from the AD's ability to include dimen-

sions of staff performance not easily captured by AgriSmart. On the other hand, subjectivity could reduce transparency and leave room for bias.

Subjective Plus Arm: In tehsils assigned to the Subjective Plus arm, the bonus assignment process mirrored the Subjective arm. The main difference was that a monthly report on the performance of the ADs' tehsil was provided to the AD's immediate supervisor, the Divisional Director, as well as to provincial management, including in particular, the Director General of Extension and Secretary Agriculture (see Figure C.5 in Appendix C for a snapshot of the report).

This report included AgriSmart based summary statistics on staff performance and data quality indicators from the AgriSmart farmer audit. Importantly, the report also included indicators of AD effort, specifically the ADs' compliance with AgriSmart Standard Operating Procedures (SOPs) such as the scheduling of FTPs, a key complementary input for frontline service delivery, and the timely approval of staff leave applications. The report was intended to trigger the career incentives of ADs by making the performance of their tehsil, and their own effort, visible to the top management of the department.

4.2 Randomization and Balance Checks

The experiment was implemented across Punjab province, including 36 districts, 131 tehsils, and 2,595 frontline extension workers. Within each district, tehsils were randomly assigned to one of the treatment arms, or held as control. [6]

Table A.2 provides balance checks across a range of FA/AO and tehsil characteristics. Only three tests fail the equality of means test - a failure rate of less than 5%, which is well within the acceptable range. All the joint F tests, whether between treatment arms and the control group or between treatment arms have a p-value greater than 0.12. During the course of the intervention (a 30-month period), 396 workers (about 15%) left the program due to transfers

[6] A district is the third administrative level within Punjab, and tehsil is the fourth. A district typically has four tehsils.

or retirement. Due to this attrition our endline sample has 2199 frontline workers. In a few of these cases, staff were transferred across treatment tehsils, but this was a rare occurrence since frontline field staff are typically not moved across tehsils. We are able to track such shifts but we report ITT estimates throughout so this has no impact on the results. Table A.3 shows that the characteristics of retained staff are also balanced. Table A.4 shows that attrition was not related to treatment.

Also, during this period, four tehsils were merged with another pre-existing tehsil. [7] This reduced the total number of tehsils in the intervention to 127. However, in all four cases, the receiving tehsil had the same treatment as the tehsil being merged into it.

4.3 Timeline

The department officially launched the AgriSmart portal in March 2018. In April 2018, the incentive program was rolled out. During this month, extensive training of field staff as well as ADs was conducted. All training was conducted in tehsils and was treatment-wise.[8]

The AgriSmart dashboard, which enabled ADs to shift to online review and approval of leave requests, online scheduling of FTPs and online bonus assignments (in the Subjective and Subjective Plus arms), was fully launched in August 2018. Figure 4 presents a detailed program timeline, highlighting key activities.

[7] This was not a change in the administrative structure of the Province but a departmental decision related to span of control.

[8] Training sessions were conducted sequentially, by treatment arm, across all districts. They included: (1) an overview and technical training on using the AgriSmart application for daily activity logging; and (2) a review of the performance-based incentive program. ADs received additional training based on their assigned treatment, focusing on the use of the dashboard for managerial tasks such as leave approvals, scheduling FTPs, and bonus assignments in the Subjective and Subjective Plus arms.

5 Measurement and Data Sources

In this sub-section, we describe our data sources and main outcomes that we use for measurement.

5.1 Data Sources

Our analysis relies on administrative and survey data from the following sources:

AgriSmart Field Staff Data. Our primary data on AO/FA extension outreach effort comes from the daily activity records submitted by all field staff through the AgriSmart smartphone application. Using this data, we construct a staff-month panel covering a 30-month period from April 2018 to January 2021, based on intention-to-treat (ITT) assignment of field staff to tehsils including staff who submitted data to the Agrismart system in each month. We exclude the peak COVID-19 period (April–July 2020) when extension activities were suspended.

AgriSmart Dashboard Data. Data on supervisor (AD) effort is drawn from the record of AD activities on the AgriSmart dashboard. We construct an AD-month panel covering 26 months, from August 2018 (dashboard launch) to January 2021, excluding April–July 2020 due to COVID-19 disruptions

AgriSmart Farmer Audit Data.To deter misreporting in the AgriSmart application and verify data accuracy, we conducted a monthly phone audit with farmers. These were farmers whose numbers were reported in the AgriSmart application by extension workers over the past month, either because they participated in a group training activity or had a one-on-one meeting with the relevant extension worker. For each extension worker, one such farmer was randomly selected each month over an 18-month period from March 2019 to December 2020.[9]

[9]This excludes the four months from April–July 2020 when extension activities were suspended. The panel comprises 30,542 observations, fewer than the expected 39,582. This is due to: 1) some staff not reporting

Punjab-Wide Farmer Registration Database. Measures of farmer access to extension services and agricultural outcomes come from a farmer phone survey. Importantly, this survey targeted farmers who were not self-reported by extension staff on the AgriSmart application, ensuring unbiased farmer outcome data.

In 2019, the Punjab Agriculture Department assembled a database of farmers across Punjab for the rollout of several departmental initiatives. 1.8 million farmers had been registered in this database by 2020.[10] We use this database to sample one farmer from the catchment area of each extension worker for a phone survey. Nine rounds of phone surveys were conducted between January 2020 and January 2021.[11]

6 Main Results

6.1 Empirical Strategy

In this section, we present the empirical strategy used to for estimating treatment effects. We look at effects on three main categories of outcomes: extension outreach effort, farmer outcomes (including access to extension and agricultural outcomes), and supervisor effort. Sections 5 details the data sources used and the construction of the main outcome variables.

To estimate the impact of treatments on staff extension outreach, we use the following ANCOVA specification:

$$y_{itdm} = \alpha + \gamma \underline{y}_{itd} + \sum_j \beta_j T^j + \nu_d + \mu_{my} + \epsilon_{itdm} \tag{4}$$

where y_{itdm} is an AgriSmart-based extension outreach indicator (e.g., days worked, hours worked, village visits, farmer visits, FTPs conducted) for extension worker i in tehsil t,

on the Agrismart system in certain months due to leave or other work assignment by the department; or 2) staff not reporting farmer numbers against their activities in certain months.

[10] In Online Appendix A, we show that the characteristics of farmers in this database align closely with those of Punjab farmers in the 2010 Agriculture Census.

district d, and month m. $\underline{y}_{itd}$ is the corresponding baseline measure of y_{itdm}, included for ANCOVA adjustment. T^j is a treatment indicator equal to 1 if tehsil t is assigned to treatment j, and 0 otherwise. ν_d and μ_{my} denote district fixed effects and year-month dummies, respectively. Standard errors are clustered at the tehsil level, the unit of randomization (**?**, **?**). The coefficient of interest, β_j, captures the treatment effect for each treatment arm.

To estimate treatment effects on farmer outcomes, we use the following specification:

$$y_{itdm}^{farmer} = \alpha + \sum_j \beta_j T^j + \nu_d + \mu_{my} + \epsilon_{itdm} \tag{5}$$

where y_{itdm}^{farmer} is a farmer-reported outcome (e.g., farmer access to extension, knowledge score, wheat yield) for farmer i in tehsil t, district d, and month m. T^j, ν_d, and μ_{my} are defined as above, with standard errors clustered at the tehsil level.[12]The treatment effect is given by β_j for each arm.

Finally, to assess the impact on supervisor effort, we estimate:

$$y_{tdm}^{AD} = \alpha + \sum_j \beta_j T^j + \nu_d + \mu_{my} + \epsilon_{tdm} \tag{6}$$

where y_{tdm}^{AD} is an indicator of AD effort (e.g., proportion of staff with scheduled FTPs, proportion of leave requests approved) in tehsil t, district d, and month m. T^j, ν_d, and μ_{my} are defined as before, and standard errors are clustered at the tehsil level. The treatment effect is given by β_j for each arm.

We also compute FDR adjusted q-values following Anderson (2008) for each set of results.

[12]Although we lack baseline data for these outcomes, we test for balance on fixed farmer characteristics and find no significant differences across treatment arms. See Table A.10 in the supplementary appendix for details.

6.2 Treatment Effects on Extension Effort

Extension outreach. We begin by examining the impact of treatment on AO and FA extension outreach effort, as captured by AgriSmart indicators. These indicators reflect dimensions of effort directly incentivized under the Objective arm; thus, we expect to observe positive effects in this arm. As discussed in Section 3, supervisors in the Subjective and Subjective Plus arms were also free to observe staff performance through the lens of AgriSmart indicators if they chose to do so.

The impact of treatment on outreach measures is presented in Table 1. In all three treatment arms, there is a positive and significant effect on attendance (number of days and hours worked) and outreach (village and farmer visits).

Attendance measures (columns 1 and 2) show that, relative to the control group, AOs/FAs in the Objective arm, on average, work nearly one additional day and approximately 54 minutes longer per day each month. In the Subjective arm, staff work slightly more than half a day more and about 30 minutes longer per day, while in the Subjective Plus arm, they work about three-quarters of a day more and roughly one hour longer per day.

For other outreach indicators, such as village visits and individual farmer visits (columns 3 and 4), the Objective arm significantly outperforms both the Subjective and Subjective Plus arms (P-value difference$<$0.01). Compared to the control group, AOs/FAs in the Objective arm make 10 additional village visits and 19 more farmer visits per month, 13 of which are to small farmers. Meanwhile, AOs/FAs in the Subjective arm make 3 more village visits, and 6 additional farmer visits, while while those in the Subjective Plus arm make 4 more village visits and 8 additional farmer visits per month.

Column 6 indicates that AOs/FAs in the Subjective Plus arm complete nearly 2 additional Farmer Training Programs (FTPs) per month relative to the control group. The effects on FTPs are insignificant in both the Objective and Subjective arms.

We report FDR adjusted q-values following Anderson (2008) in parenthesis; these are closely

aligned with the unadjusted p-values for all the results presented above.

Table 1 about here

The stronger effects observed in the Objective arm are expected, given that incentives were explicitly tied to AgriSmart outreach measures. At the same time, the positive effects in the Subjective and Subjective Plus arms suggest that, even though bonuses in these arms were not directly linked to AgriSmart data, the visibility of these measures to ADs may have motivated staff to improve their performance.

Finally, note that scheduling FTPs is an AD-level responsibility, whereas conducting scheduled FTPs is the responsibility of field staff. We further explore the role of supervisory inputs in frontline extension service delivery, particularly regarding FTPs, in Section 6.4.

Quality of extension outreach and effort margin. As outlined in our conceptual framework (Section 3), higher outreach effort in the Objective arm—which directly incentivizes the number of farmer and village visits—may come at the cost of lower visit quality, as quality itself is not incentivized. Furthermore, since bonuses in the Objective arm are based on achieving a minimum number of visits rather than expanding outreach to new farmers, we might expect greater effort on the intensive margin rather than the extensive margin. In contrast, incentives in the Subjective and Subjective Plus arms rely on supervisor evaluation, potentially reducing these trade-offs.

Columns (1)–(2) in Table 2 present treatment effects on extension quality, proxied by the length of meetings. While the number of meetings was incentivized in the Objective arm, meeting length was not, making it a suitable measure of interaction quality. We find that higher outreach effort does not come at the expense of lower quality. Column (1) shows positive coefficients for meeting length across all treatment arms. Staff in the Subjective Plus arm spend approximately 8 more minutes in meetings compared to the control group, although this effect is not statistically significant after adjusting for multiple comparisons (see adjusted q-value in parenthesis). Column (2) shows that, relative to the control group (where staff spend about 62 minutes on group activities), staff in the Objective, Subjective,

and Subjective Plus arms spend 13, 12, and 14 additional minutes, respectively—representing a 20–23% increase in meeting length.

Columns (3)–(6) report treatment effects on effort along the extensive margin. Column (3) shows that staff in the Objective and Subjective Plus arms visit more than 2 additional unique villages per month compared to the control group, while the effect in the Subjective arm is not significant. Column (4) indicates that in the Subjective Plus arm, one of these additional village visits involves conducting group activities such as FTPs. In contrast, additional village visits in the Objective arm are primarily for individual farmer visits, rather than for group activities.

In terms of farmers reached, Column (5) shows that all treatment arms engage with more farmers than the control group. The significant effect in the Subjective arm suggests that staff in this group may have increased outreach to new farmers within the same villages (intensive margin), rather than expanding to new villages (extensive margin). Column (6) shows that only the Subjective and Subjective Plus arms significantly increase the number of farmers reached through group activities, suggesting that staff in the Objective arm rely mainly on individual visits to expand their outreach.

Finally, Columns (7)–(8) show that the Objective arm induces more effort on the intensive margin relative to both the control group and other treatment arms. Over the course of the program, AOs/FAs in the Objective arm visited the same farmer at least 3 more times than those in the control group, while staff in the Subjective and Subjective Plus arms conducted only one additional repeat visit (Column 7). Moreover, these repeat visits in the Objective arm are not driven by group activities (Column 8), consistent with the earlier findings in Column (6).

Overall, these results highlight two key points. First, increased outreach across all treatment arms does not reduce quality, as measured by meeting length; in fact, group activity durations improve. Second, while both the Objective and Subjective Plus arms increase effort on the extensive margin (i.e., unique villages visited and farmers reached), the Objective arm achieves this primarily through individual visits, whereas the Subjective Plus arm relies

on both individual and group visits. Additionally, staff in the Objective arm demonstrate greater effort on the intensive margin compared to other treatment and control groups.

Table 2 about here

Staff reporting behavior. An important question is whether the outreach effects observed in Table 1 are driven by staff reporting behavior. Reporting could affect results in three ways. First, treatment effects could be biased if the share of non-reporting staff is imbalanced across treatment arms. Second, staff in treatment arms may report more diligently due to incentives being either directly tied to AgriSmart data or subject to supervisor review for bonus decisions. Third, staff in the Objective arm may over-report activities to meet incentive thresholds.

We address each concern in turn.

Table A.7 shows that activity reporting on AgriSmart is not related to treatment assignment. The monthly reporting rate is also quite high across all arms. On average, only 11% of staff in the control group did not report their activities on AgriSmart over the 30-month intervention period. This is a high compliance rate for a cadre newly introduced to smartphone-based reporting.

Second, we assess whether the observed treatment effects in Table 1 reflect genuine effort or simply more frequent reporting by treatment staff. Using endline survey data, where FAs and AOs reported how often they entered data into AgriSmart, we find that while treatment staff report slightly higher frequencies of reporting (Table A.8), most coefficients are statistically insignificant. Moreover, even where significant, the effects are small and insufficient to account for the treatment effects observed.[13]

Third, we investigate whether staff in the Objective arm attempted to game the system by inflating reported activities near incentive thresholds. We examine the distribution of

[13]For instance, Column 5 in Table A.8 shows that Subjective Plus staff report 4.3% more FTPs each month. Even if we assume this fully reflects reporting bias, it would account for only 0.28 FTPs of the total 6.7 FTPs reported in Table 1 in the Subjective Plus arm, leaving nearly 1.53 additional FTPs in the Subjective Plus arm.

each incentivized indicator to detect any bunching around the low and high bonus cutoffs. Figure 7 shows no evidence of bunching for most indicators, including days worked, hours worked, village visits, farmer visits, distance traveled, and FTPs conducted. However, we do observe bunching in the 'number of farmers per FTP'—the only fully self-reported indicator in the Objective arm bonus criteria. Other indicators rely on automated time tracking or geo-tagging, which are more difficult to falsify. As this pattern appears in only one measure, we conclude there is no substantial evidence of successful systematic gaming in the Objective arm.

6.3 Treatment Effects on Farmer Outcomes

In this section, we examine whether the improved extension outreach observed in Section 6.2 translates into greater farmer access to extension services, as well as downstream outcomes such as farmer knowledge and agricultural yields.

Table 3 presents these results. Column 1 shows that, compared to the control group, the Subjective and Subjective Plus arms increase farmer access to extension services by approximately 0.06 standard deviations. The coefficient for the Objective arm is positive but statistically insignificant.

To understand which components of the access index drive these effects, Table A.9 disaggregates treatment impacts across individual indicators. Relative to the control group, farmers in the Subjective and Subjective Plus arms are 3.5 and 3.7 percentage points more likely, respectively, to be informed about extension activities, and 2.3 and 1.5 percentage points more likely to receive one-on-one visits from their AOs/FAs (Columns 2 and 5 in Panel A). In the Subjective Plus arm, there is also indicative evidence that farmers are 2.2 percentage points more likely to report group meetings being held in their villages. The coefficient for farmer attendance at these group meetings is also positive in the Subjective Plus arm (while negative in the other arms), though it remains statistically insignificant.

Ultimately, the goal of extension services is to improve long-term agricultural outcomes such as farm practices and yields. While the treatment arms in this program incentivized staff effort rather than direct agricultural outcomes, theoretically, higher staff effort should translate into improvements in these downstream outcomes. However, several practical constraints limit this relationship. First, improvements in farmer outcomes requires access to relevant and up to date information in addition to field staff outreach. The department had initiated a program to improve the quality of advisory materials, but this did not take off in the end, at least within the timeline of this intervention. Second, changes in agricultural practices and yields often takes several seasons to materialize. Despite these challenges, Columns (2) and (3) of Table 3 explore treatment effects on farmer knowledge and wheat yields. Column 2 shows that treatment effects on farmer knowledge are statistically insignificant and close to zero. While the coefficients for wheat yield are positive in the Subjective and Subjective Plus arms, they are also insignificant. These findings are unsurprising, given that the intervention primarily incentivized outreach effort. The expected updating of staff skills and advisory content did not materialize.

Interestingly, while Table 1 showed that the strongest treatment effects on extension outreach effort were observed in the Objective arm, improvements in farmer-reported access to extension are only evident in the Subjective arms. One possible explanation, supported by our findings on extensive and intensive margins of effort in Table 2, is that the Subjective arms engaged more farmers through group activities such as FTPs. In the Subjective Plus arm, this effect was particularly pronounced, with more FTPs conducted in unique villages compared to both the control group and other treatment arms. By contrast, the Objective arm relied more heavily on individual farmer visits and showed evidence of repeated visits to the same farmers more frequently than other arms.

In the next section, we investigate why staff in the Subjective Plus arm may have been able to conduct more FTPs in a greater number of unique villages.

Table 3 about here

6.4 Treatment Effects on Supervisor Effort

Our conceptual framework in Section 3 highlights the important role that supervisors can play through the provision of complementary inputs. In this context, ADs' complementary inputs to frontline service delivery include key managerial tasks such as planning and scheduling FTPs, assigning FTPs to frontline staff, and a slew of administrative and supervisory tasks, including the timely approval of staff leave requests.

Supervisor incentives are activated differently across treatment arms. Although supervisors are not directly incentivized in the Objective and Subjective arms, they may still respond to increased staff effort or face staff pressure to complete some managerial duties, as in the objective arm on the issue of staff leave request approvals, which must be entered in a timely way for the FA/AO to qualify for the monthly bonus. In contrast, in the Subjective Plus arm, the monthly reports on tehsil performance, which include metrics of the ADs own effort, could have triggered greater effort from ADs in a bid to better align their actions with the objectives of the principal.

In this section, we examine the impact of the three treatment arms on supervisor effort. We look specifically at two aspects of supervisor effort: (1) the proportion of AOs/FAs assigned an FTP each month; and (2) the proportion of leave requests submitted by AOs/FAs that are addressed in a timely way (i.e., approved or rejected).

Table 4 presents these results. Column (3) shows that ADs in the Subjective Plus arm assign 6.7 percentage points more FTPs to staff (AO and FAs combined) relative to the control group (about a 16% increase over the control group), and these effects are significantly higher than those in the Objective and Subjective arms at the 10% level. However, most of this increase is concentrated in a higher assignment of FTPs to FAs. Column (1) indicates that 7.4 percentage points more FTPs were assigned to FAs in the Subjective Plus arm compared to the control group (about an 18% increase over the control group where about 40% of FAs are assigned FTPs) while column (2) shows a positive but insignificant effect for FTPs assigned to AOs in the Subjective Plus arm. In both cases, the effects for the Objective and

Subjective arms are insignificant.

Finally, Column (4) shows a significant increase in leave approvals in the Objective arm, which supports the interpretation that staff may have exerted pressure on ADs to approve leaves, which impacted their qualification for a bonus, while the scheduling of more FTPs did not.

Table 4 about here

These results help explain the higher FTP completion rate by staff in the Subjective Plus arm observed in Table 1. Moreover, they indicate that because ADs were monitored through tehsil reports in the Subjective Plus arm, they not only assigned more FTPs to staff but also did so across more unique villages which helps explain our results on the extensive margin in Table 2. Overall, this highlights the importance of complementary inputs that supervisors provide to frontline service delivery and suggests the potential benefits of directly incentivizing supervisors.

7 Mechanisms

Our conceptual framework in Section 3 argued that when supervisors' incentives are better aligned with the objectives of the principal (for example, through top-down monitoring), they will exert more effort on complementary inputs and make more efficient bonus assignments. In the previous section, we showed that top-down monitoring under the Subjective Plus arm led supervisors to exert greater effort on their complementary inputs by assigning more FTPs to staff. However, several key mechanisms remain to be explored. First, did the tehsil reports succeed in aligning AD incentives with the principal's objectives as intended? Second, did supervisors acquire additional information to make more informed bonus assignments? Finally, did they use this information when making assignment decisions? In this section, we present evidence on each of these questions.

Legitimacy of the tehsil reports. The design of the Subjective Plus arm aimed to introduce top-down monitoring to align supervisors' incentives. Here, we assess whether ADs perceived the tehsil reports as legitimate and motivating.

Descriptive evidence from the endline AD survey offers support. Approximately 89% of ADs in the Subjective Plus arm reported that the tehsil reports were either very important (77%) or somewhat important (11%) to their senior leadership. Moreover, about 50% of ADs indicated that their seniors had taken action based on the reports—either by contacting them in person, over the phone, or via written communication regarding specific performance indicators. When asked "how do you think good performance against the indicators in the tehsil report can benefit you?", ADs most frequently cited either opportunities for promotion (21%) or personal motivation and satisfaction (79%). Taken together, this evidence suggests that the tehsil reports were perceived as legitimate within the department and were successful, at least partially, in aligning AD incentives with the principal's objectives.

Acquisition of information by supervisors. If supervisor incentives were better aligned with the principal's objectives, we might expect supervisors in the Subjective arms to acquire additional information about staff performance through more intensive efforts—for instance, by conducting more field monitoring visits or using the AgriSmart dashboard more regularly.

Table 5 presents the impact of our treatments on ADs' field visits (Column 1) and dashboard usage (Columns 2–3). We find no evidence that ADs in the Subjective Plus arm conducted more field visits relative to other arms (Column 1). However, as Column (3) shows, ADs in the Subjective Plus arm were more likely than ADs in the Subjective arm, to view staff performance summaries on the dashboard.

Thus, while the introduction of tehsil reports did not lead to more field monitoring visits, it appears to have encouraged greater use of AgriSmart dashboard data for monitoring staff performance.

Table 5 about here

Supervisor bonus assignment. In our final analysis, we examine how bonuses were assigned across treatment arms, with a specific focus on the staff characteristics that predicted bonus assignments in the Subjective arms. Supervisors in the Subjective and Subjective Plus arms could base their decisions on observable outreach metrics available through AgriSmart, or on less easily measurable factors such as staff ability and service quality, which may only be privately observed. However, while the Subjective arm opens the possibility of supervisory bias in bonus assignment, the top-down monitoring in the Subjective Plus arm could mitigate such biases. We test these theoretical predictions below.

We begin by examining overall bonus assignment patterns across arms. Figure 8 shows that average monthly bonus amounts were higher in the Subjective and Subjective Plus arms (PKR 4,845 and 4,994, respectively) compared to the Objective arm (PKR 2,642). This is consistent with the predictions from our conceptual framework in Section 3, which suggests that subjective schemes may induce *centrality bias* (similar bonuses for all) or *leniency bias* (generally higher bonuses) (Gibbs et al., 2004).[14]

Figure 8 about here

We then investigate which staff characteristics predict bonus allocation. We group these characteristics into four categories: (1) AgriSmart performance metrics (e.g., work attendance, outreach visits, FTPs completed), (2) worker ability (e.g., education, experience, knowledge scores), (3) worker connections with the AD (e.g., shared zaat/caste), and (4) worker personality (measured by BFI and PSM indices). The first group is observable through AgriSmart, while the latter three are likely known only to supervisors. For comparability, all predictors are standardized as z-scores. We estimate the following specification:

$$y_{itdm} = \alpha + \sum_j \beta_j C^j + \nu_t + \mu_{my} + \epsilon_{itdm},$$

[14]We rule out the possibility that these differences reflect differential payment rates across arms. According to our endline survey of field staff, Figure B.9a shows that workers in all arms reported receiving bonuses in 3–4 out of 6 months from January to June 2020. Table A.11 confirms that the number of bonuses paid was similar across arms, with only a marginally significant difference (p = 0.08) between the Objective and Subjective arms. Beliefs about whether unpaid bonuses would eventually be paid also did not differ significantly across treatments.

where y_{itdm} is the bonus amount assigned to worker i in tehsil t, district d, and month m; C^j represents staff characteristics; and ν_t and μ_{my} are tehsil fixed effects and month-year dummies. Standard errors are clustered at the tehsil level.[15]

Table 6 presents the results. Panel A reports estimates for the full sample, and Panel B focuses on staff assigned FTPs.

Table 6 about here

In line with expectations, bonus assignment in the Objective arm (Column 1, Panel A) is primarily driven by AgriSmart performance metrics, especially days worked, hours worked, and farmer visits. Columns (4) and (5) confirm that these metrics are significantly more predictive in the Objective arm than in the Subjective or Subjective Plus arms.

In contrast, Columns (2) and (3) show that while AgriSmart metrics like attendance and outreach remain relevant in the Subjective and Subjective Plus arms, other factors also influence bonus decisions. In the Subjective arm (Column 2), bonuses are more strongly linked to years of schooling, experience, and—importantly—whether the worker shares the same zaat as the AD, suggesting favoritism or social bias. In the Subjective Plus arm (Column 3), supervisors place greater weight on AgriSmart performance, staff knowledge, and personality traits (Columns 3 & 6), while relying significantly less on shared zaat. This suggests that top-down monitoring helps reduce favoritism and refocuses bonus allocation on performance-related criteria.

Panel B shows similar patterns among staff assigned FTPs, with an additional finding that Subjective Plus supervisors place significantly more weight on FTP completion compared to the other arms.

Taken together, these findings suggest two actionable lessons for incentive design. First, granting supervisors unmonitored discretion risks rewarding social ties over effort or per-

[15] We focus on FAs, who make up about 90% of the workforce and for whom we have complete covariate data. While AgriSmart indicators vary monthly, ability and personality measures are time-invariant. Including both types of measures improves precision and allows us to control for AD fixed effects.

formance. Second, even light-touch monitoring shifts bonus allocation toward objectively desirable traits—such as outreach, knowledge, and actual fieldwork—while reducing the influence of social bias.

8 Conclusion

This paper evaluates the impact of a large-scale randomized controlled trial in the Agricultural Extension Department in Punjab, Pakistan that aims to measure the impact of three different pay-for-performance schemes for frontline extension agents on extension service delivery. These include an Objective scheme in which incentives are based on performance metrics via a digital platform, a Subjective scheme in which incentives are based on supervisors' subjective evaluation, and a Subjective Plus scheme which is a hybrid approach combining a supervisor's subjective evaluation with top-down monitoring of the supervisor's inputs into frontline extension.

Our results highlight that monitoring supervisors can enhance the impact of performance incentives for frontline staff if supervisors make important complementary inputs into frontline service delivery, and hold private information about staff quality. Our results show positive treatment effects in the Subjective Plus arm on extension outreach with evidence of outreach to more unique farmers and villages, extension quality, and farmer access to extension.

Our analysis shows that these effects are primarily driven by supervisors provision of their complementary inputs into frontline service delivery. In particular, ADs in Subjective Plus schedule more Farmer Training Programmes (FTPs) which translates into staff visiting more unique villages and expanding outreach to more unique farmers for conducting these trainings. We also show that ADs in Subjective Plus are less likely to show favoritism and more likely to use their private information on staff quality as well as their performance on outreach measures (such as FTPs completed) compared to the Subjective arm ADs.

Note that the monitoring of ADs in Subjective Plus was fairly light-touch with no monthly or

quarterly high-stakes meetings based on the tehsil reports. This implies that our estimates of such an intervention are, at best, conservative and could be strengthened through a monitoring system where stronger feedback mechanisms are established with the supervisor's supervisors.

A final consideration is the cost and trade-offs involved in the various incentive schemes. The department's main objective in introducing the incentive schemes was to make the use of its large TADA budget more efficient and performance-driven. A related objective was to ensure that at least 70% of the total budget was disbursed to frontline extension staff. Prior to the incentive programme in FY 17-18, only 47% of the budget was distributed among frontline staff. Our hypothetical calculations of total expected budget requirement based on actual bonus assignments in each incentive scheme are presented in Table 8. While the Subjective and Subjective Plus arms cost almost twice as much as the Objective arm, the total expected requirement under these schemes remains within the total TADA budget ceiling of PKR 350 million. Moreover, the Subjective Plus scheme which shows the most promising results is closer to the departmental objective of disbursing at least 70% of budget to frontline staff in comparison to the Objective and Subjective scheme.

We conclude by noting considerations for policy. Our results have direct policy implications for the Extension department. They indicate important trade-offs to consider in the design of their incentive schemes – the Objective arm improves extension outreach with no unintended consequences for extension quality or staff motivation. It has the smallest budget requirement in comparison to the other schemes. On the other hand, the Subjective Plus arm improves both extension outreach and farmer access to extension, but has almost twice the budget requirement. The department's choice of a potential scheme will depend on its valuation of these costs and benefits. More broadly, our results are relevant for policymakers across other frontline service delivery departments in Pakistan, and other low-middle-income country contexts where services are delivered through multi-layered organisations.

References

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: a Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484):1481–1495.

Baker, George. 1992. "Incentive Contracts and Performance Measurement." *Journal of Political Economy* 100:598–614.

Baker, George et al. 1994. "Subjective Performance Measures in Optimal Incentive Contrac." *The Quarterly Journal of Economics* 109 (4):1125.

Bandiera, Oriana et al. 2007. "Incentives for Managers and Inequality Among Workers: Evidence from a Firm-level Experiment*." *The Quarterly Journal of Economics* 122 (2):729.

Beaman, Lori, Ariel Benyishay, Jeremy Magruder, and Ahmed Mobarak. 2018. "Can Network Theory-based Targeting Increase Technology Adoption?" *NBER Working Paper Series* :24912.

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.

Burgess, Simon and Marisa Ratto. 2003. "The Role of Incentives in the Public Sector Issues and Evidence." *Oxford Review of Economic Policy* 19 (2):285–300.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *The Journal of Economic Perspectives* 20 (1):91–116.

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.

Deserranno, Erika, Stefano Caria, Philipp Kastrau, and Gianmarco León-Ciliotta. 2022. "The allocation of incentives in multi-layered organizations." Economics Working Papers 1838, Department of Economics and Business, Universitat Pompeu Fabra.

Deserranno, Erika et al. ???? "Promotions and productivity: The role of meritocracy and pay progression in the public sector." *American Economic Review: Insights* .

Duflo, Esther et al. 2012. "Incentives Work: Getting Teachers to Come to School." *The American Economic Review* 102 (4):1241–1278.

Fernando, A. Shawn, A. Nilesh; Cole. 2016. "*M*obileizing Agricultural Advice: Technology Adoption, Diffusion and Sustainability."

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108 (5):874–904.

Gibbons, Robert. 1998. "Incentives in Organizations." *Journal of Economic Perspectives* 12 (4):115–132.

Gibbs, Michael et al. 2004. "Performance Measure Properties and Incentives." *Ideas Working Paper Series from Repec* .

Holmstrom, Bengt and Paul Milgrom. 1991. "Multitask Principal-agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, Organization* 7:24–52.

Khan, Adnan Q. et al. 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors.(report)." 131 (1):219.

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." *The American Economic Review* 111 (7):2213–2246.

Manning, Nick et al. 2012. *Public Sector Human Resource Practices to Drive Performance*. GET Note. World Bank, Washington, DC.

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.

Neal, Derek. 2011. "The Design of Performance Pay in Education." *NBER Working Paper Series* :16710.

Policy. 2018. "Punjab Agriculture Policy."

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1):7–63.

———. 2007. "The Motivation and Bias of Bureaucrats." *The American Economic Review* 97 (1):180–196.

Prendergast, Canice and Robert H. Topel. 1996. "Favoritism in Organizations." *Journal of Political Economy* 104 (5):958–978.

Rasul, Imran and Daniel Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *The Economic Journal* 128 (608):413–446.

Survey, Pakistan Economic. 2019. "Pakistan Economic Survey."

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It.* New York: Basic Books.

==========

Figures and Tables

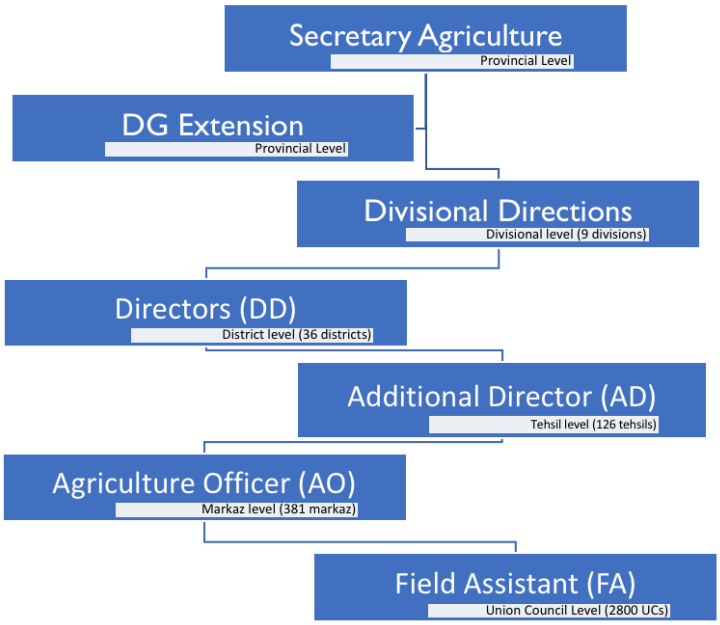
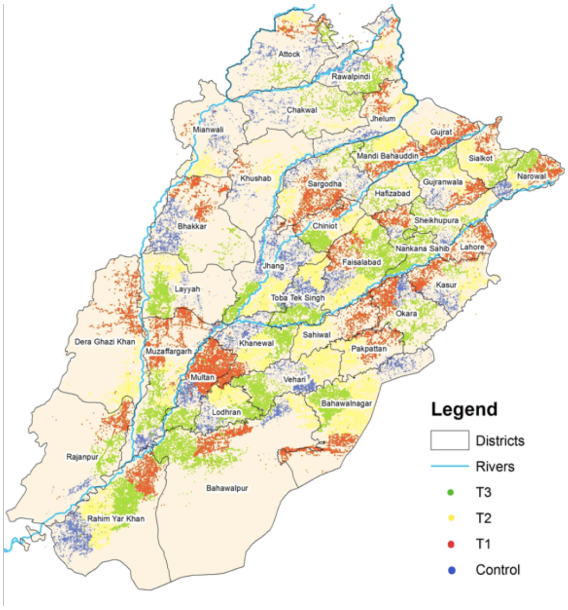


Figure 1: Hierarchical Structure



(a) AGRISMART DATA POINTS

(b) HOMEPAGE

Figure 2: AGRISMART SYSTEM

Supervisors (ADs) [Tehsil Level]		**Frontline Extension Staff (AOs and FAs)**			
		AgriSmart Based Data Entry			
		No Bonus *(TA payment possible)*	***Monthly Bonus Payment***		
No Access to AgriSmart	Can assign TA Funds to Staff	Control Tehsils (C)			
Access to AgriSmart	No Role in Staff Bonus Assignment		Objective Tehsils (T1)		
	Discretion in Staff Bonus Assignment			Subjective Tehsils (T2)	
	Discretion in Staff Bonus Assignment with Top-down Monitoring				Subjective Plus Tehsils (T3)

Figure 3: Agrismart Incentive Programme

Figure 4: Programme timeline

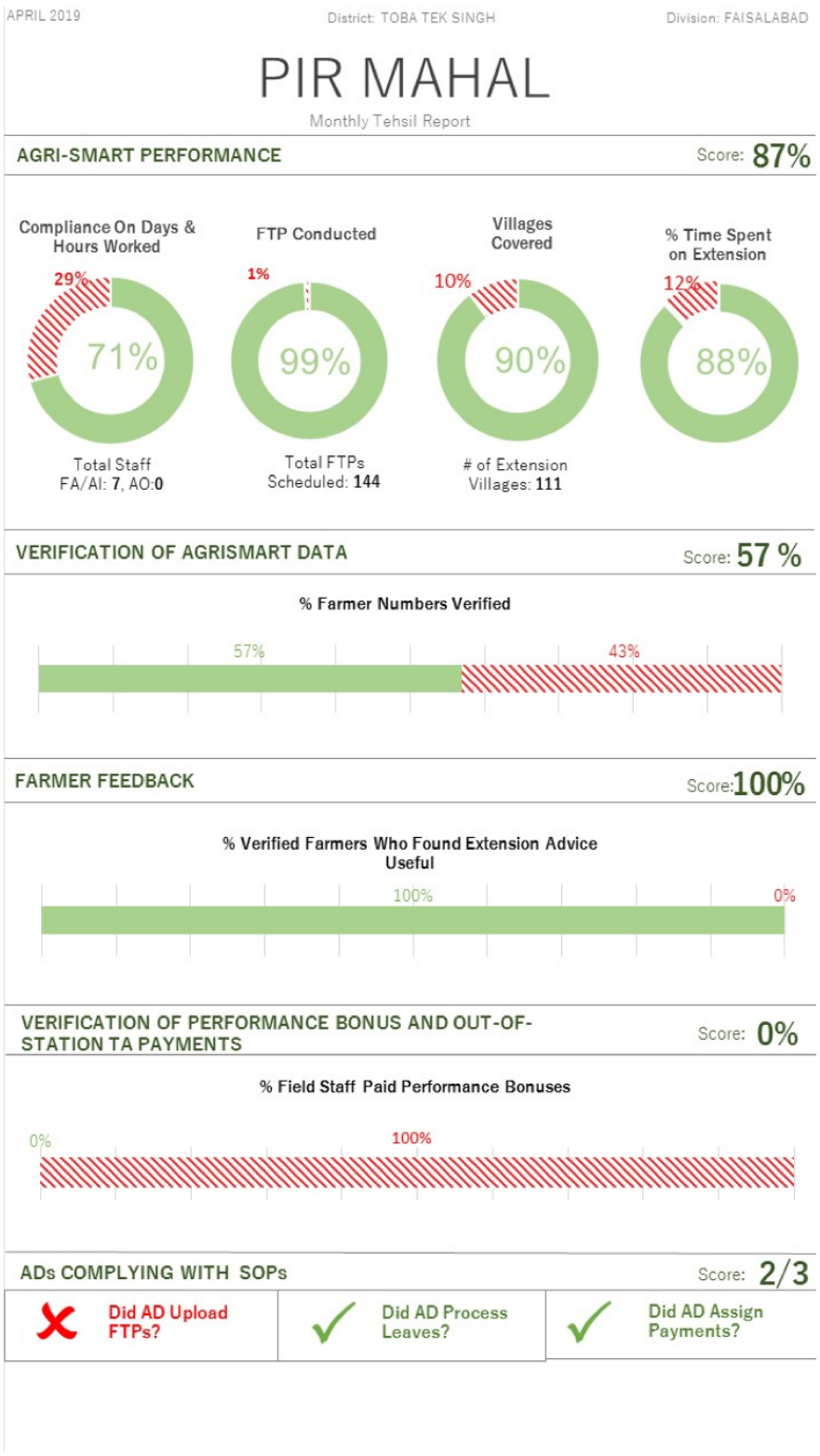


Figure 5: Tehsil report

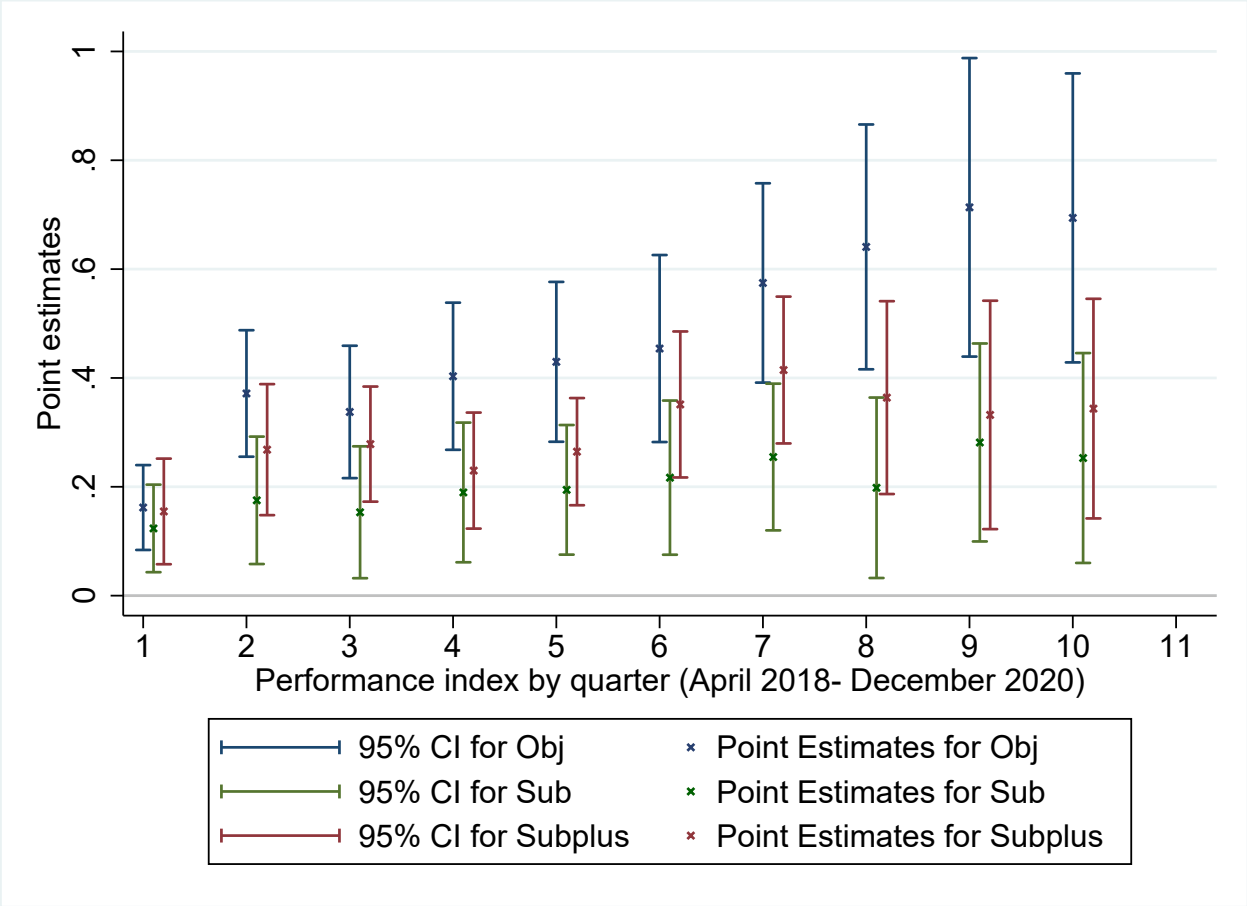


Figure 6: Treatment effects on performance index (by quarter)

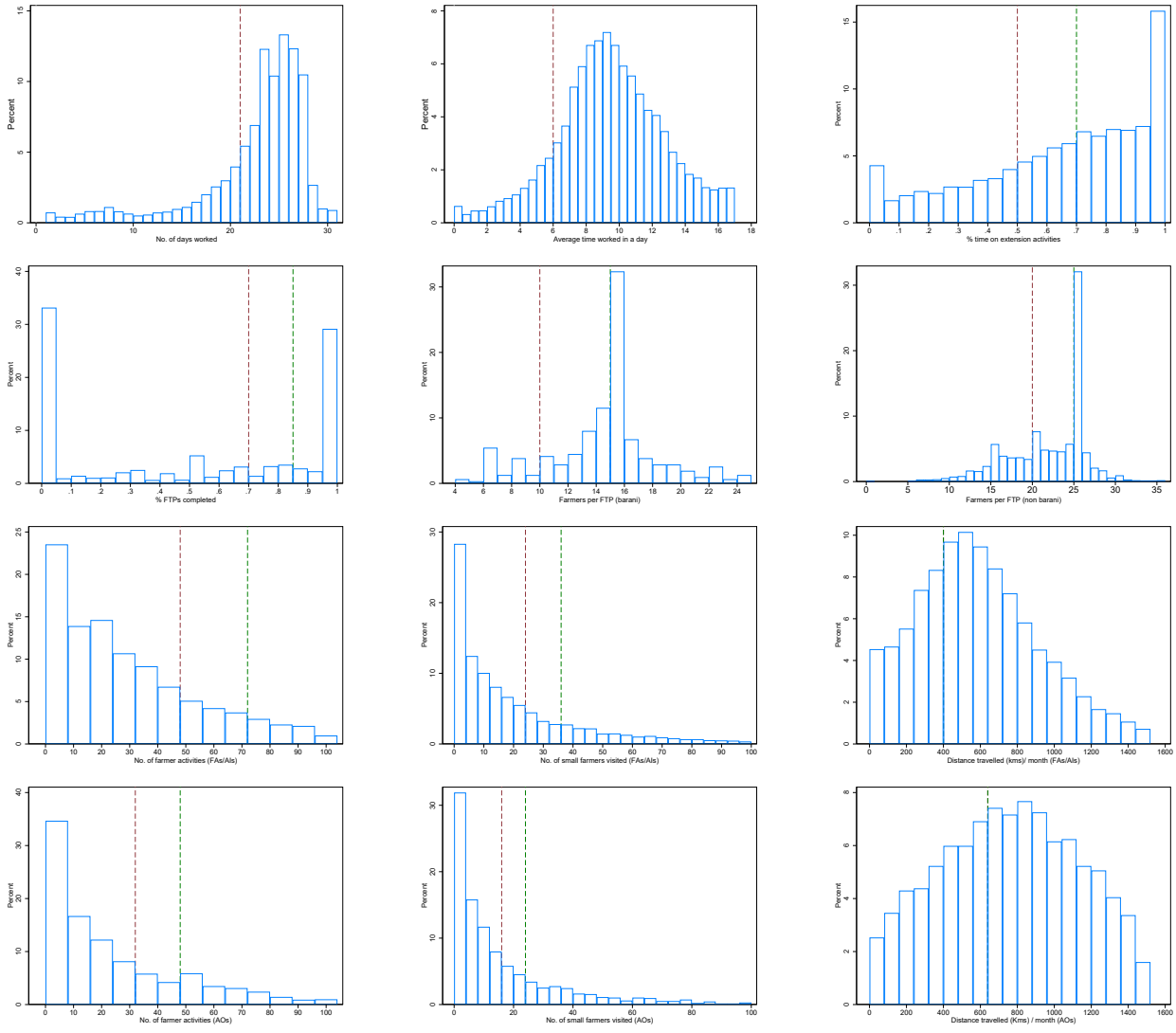
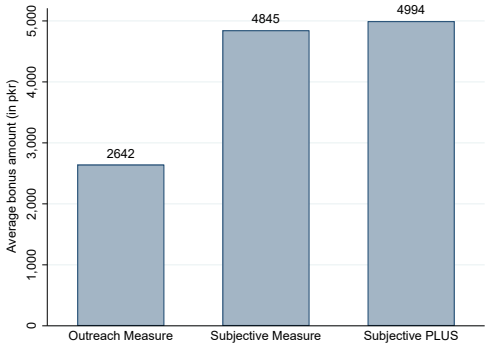
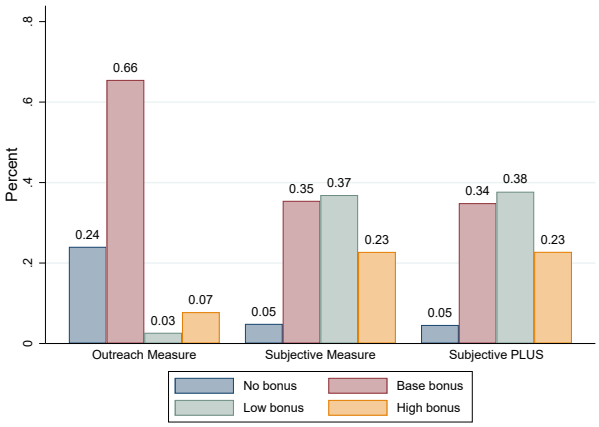


Figure 7: Activities against thresholds in Objective Arm

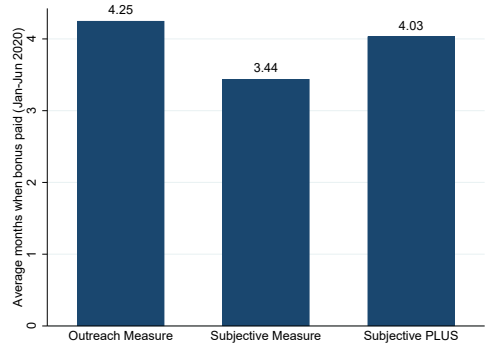


(a) Average bonus amounts (in PKR) by treatment arm.

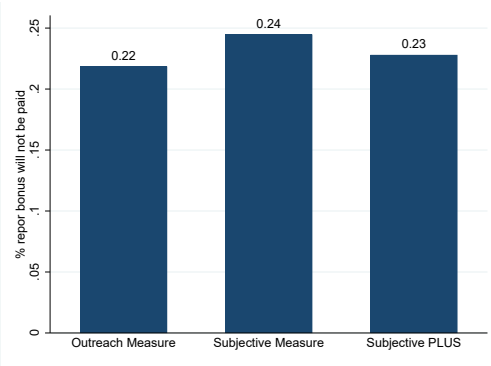


(b) Distribution of bonus levels across treatment.

Figure 8: Bonus assignment outcomes across treatment arms.



(a) Number of months with paid bonus (January–June 2020).



(b) Percentage of staff who believed their bonus would not be paid.

Figure 9: Treatment effects on bonus payments.

Table 1: Impact on Extension Outreach

	(1) Days Worked	(2) Hours Worked	(3) Village Visits	(4) Farmer Visits	(5) Small Farmer Visits	(6) No. FTPs completed
Objective	1.070***	0.928***	10.071***	19.048***	13.337***	0.851
	(0.275)	(0.238)	(1.866)	(3.223)	(2.273)	(0.572)
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.148]
Subjective	0.639**	0.493**	3.398***	6.480**	3.359*	0.671
	(0.269)	(0.198)	(1.100)	(2.636)	(1.704)	(0.525)
	[0.025]	[0.021]	[0.004]	[0.021]	[0.058]	[0.204]
Subjective	0.741**	0.961***	4.067***	7.867***	3.258**	1.819***
Plus	(0.286)	(0.218)	(1.125)	(2.359)	(1.469)	(0.555)
	[0.018]	[0.001]	[0.001]	[0.002]	[0.034]	[0.002]
Control Mean	21.046	9.282	20.014	22.485	11.450	4.884
P-value difference:						
Obj = Sub	0.108	0.016**	0.001***	0.001***	0.000***	0.763
Obj = Subplus	0.261	0.874	0.004***	0.003***	0.000***	0.122
Sub = Subplus	0.730	0.010**	0.602	0.667	0.952	0.036**
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-month dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	64,096	64,096	64,096	64,096	64,096	24,807

Notes: Standard errors are clustered at the tehsil level (unit of randomization) and reported in parentheses. Columns 1–5 are estimated using ANCOVA with baseline controls; Column 6 is an average treatment effect due to lack of baseline for FTPs. Outcomes are monthly averages from April 2018–January 2021 (excluding April–July 2020). The sample includes ITT workers active in AgriSmart. Fewer observations in Column 6 result from FTP scheduling coverage. FDR-adjusted q-values are in square brackets. Significance levels: * 10%, ** 5%, *** 1%.

Table 2: Impact on Extension Quality and Effort Margins

	Extension Quality		Extensive Margin				Intensive Margin	
	(1) Activity length (secs) overall (monthly)	(2) Activity length (secs) group (monthly)	(3) Villages visited overall (monthly)	(4) Villages visited group (monthly)	(5) Farmers reached overall (yearly)	(6) Farmers reached group (yearly)	(7) Farmer repeated overall (yearly)	(8) Farmer repeated group (yearly)
Objective	272.129	806.451**	2.277***	-0.121	169.026***	8.379	3.095***	0.260
	(299.615)	(352.500)	(0.652)	(0.406)	(46.407)	(28.101)	(0.672)	(0.282)
	[0.438]	[0.048]	[0.004]	[0.800]	[0.001]	[0.800]	[0.001]	[0.438]
Subjective	336.652	690.546**	0.759	0.405	128.118***	61.595**	1.429**	-0.013
	(254.062)	(289.860)	(0.527)	(0.416)	(46.862)	(29.310)	(0.572)	(0.200)
	[0.266]	[0.042]	[0.228]	[0.438]	[0.024]	[0.071]	[0.036]	[0.948]
Subjective Plus	465.475*	837.967***	2.141***	1.028**	187.903***	104.217***	1.088*	0.097
	(270.990)	(311.806)	(0.551)	(0.416)	(44.150)	(29.583)	(0.601)	(0.201)
	[0.141]	[0.024]	[0.001]	[0.036]	[0.001]	[0.004]	[0.124]	[0.719]
Control Mean	3142.619	3715.560	6.739	5.088	479.844	274.515	5.141	1.573
P-value difference:								
Obj = Sub	0.806	0.710	0.040**	0.220	0.476	0.075*	0.015**	0.138
Obj = Subplus	0.491	0.928	0.847	0.003***	0.740	0.004***	0.007**	0.454
Sub = Subplus	0.591	0.624	0.016**	0.084*	0.267	0.158	0.622	0.433
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-month dummies	Yes	Yes	Yes	Yes	No	No	No	No
N	56,805	20,222	56,805	20,222	2,171	2,089	2,171	2,089

Notes: Columns 1–2 use a staff-month panel from April 2018 to January 2021 (excluding April–July 2020). This panel includes ITT workers, active AgriSmart users, and those reporting village names. Given FTPs were scheduled for only 40% of staff, Column 2 has a smaller sample. Columns 3–8 are based on ITT employees who reported farmer phone numbers. Standard errors clustered at the tehsil level, in parentheses. Significance: * 10%, ** 5%, *** 1%.

Table 3: Impact on Farmer Outcomes

	(1) Access Index	(2) Knowledge Score	(3) Wheat Yield
Objective	0.026	-0.007	-15.676
	(0.017)	(0.010)	(27.956)
	[0.291]	[0.572]	[0.576]
Subjective	0.066***	-0.015	20.312
	(0.023)	(0.010)	(28.974)
	[0.036]	[0.286]	[0.572]
Subjective Plus	0.062***	-0.007	29.208
	(0.023)	(0.010)	(29.719)
	[0.036]	[0.572]	[0.572]
Control Mean	-0.069	0.332	1263.343
P-value difference:			
Obj = Sub	0.065*	0.426	0.224
Obj = Subplus	0.160	0.986	0.134
Sub = Subplus	0.882	0.374	0.783
District FE	Yes	Yes	Yes
Year-month dummies	Yes	Yes	Yes
N	13,312	9,980	8,102

Notes: The sample is based on a farmer phone survey, conducted over 8 rounds from January 2020 to January 2021 (one farmer randomly sampled per ITT employee per round). Column (1): The access index is an equally weighted average of z-scores for extension access variables, including whether the farmer knows their AO/FA, is informed about extension activities, and had individual or group-level contact. Column (2): The knowledge score is based on responses to three questions of varying difficulty about a major crop, with scores averaged across the questions. This column has fewer observations as knowledge questions were administered only to FAs. Column (3): Wheat yield is measured in kgs/acre, based on the 2019–2020 rabi season; fewer observations reflect that only farmers cultivating wheat were included. Standard errors clustered at the tehsil level are reported in parentheses. Significance: * 10%, ** 5%, *** 1%.

Table 4: Impact on Supervisor Effort

	(1) % FA Assigned FTPs	(2) % AO Assigned FTPs	(3) % Staff Assigned FTPs	(4) % Leaves Approved
Objective	0.011	0.024	0.009	0.130***
	(0.032)	(0.034)	(0.030)	(0.035)
	[0.920]	[0.816]	[0.920]	[0.001]
Subjective	0.028	-0.005	0.021	0.017
	(0.028)	(0.037)	(0.026)	(0.039)
	[0.742]	[0.924]	[0.816]	[0.920]
Subjective Plus	0.074**	0.033	0.067**	-0.003
	(0.028)	(0.031)	(0.026)	(0.034)
	[0.044]	[0.742]	[0.044]	[0.924]
Control Mean	0.403	0.635	0.425	0.667
P-value difference:				
Obj = Sub	0.563	0.428	0.683	0.002***
Obj = Subplus	0.047**	0.788	0.054*	0.000***
Sub = Subplus	0.119	0.297	0.091*	0.592
District FE	Yes	Yes	Yes	Yes
Year-month dummies	Yes	Yes	Yes	Yes
N	3,296	2,406	3,296	2,993

Notes: The sample is based on an AD-month panel from August 2018 to January 2021, covering 26 months (excluding April–July 2020 due to COVID-19). In columns 2 and 4, 6 observations are missing due to lower compliance in the final months. Column 4's dependent variable (% Leaves Approved) only applies where staff submitted leave requests; in some tehsils, no requests were submitted, reducing the sample. Some tehsils had vacant AO positions in certain months, explaining the smaller sample in column 2. All tehsils had FAs throughout. Standard errors are clustered at the tehsil level and reported in parentheses. Significance levels: * 10%, ** 5%, *** 1%.

Table 5: Impact on Supervisor Field Visits and Use of Dashboard

	(1) No. of AD Field visits (monthly)	(2) Use of dashboard: general monitoring (> twice/week)	(3) Use of dashboard: view staff performance (> twice/week)
Objective	-2.237	0.079	0.097
	(1.355)	(0.107)	(0.110)
Subjective	-0.589	0.038	-0.035
	(1.432)	(0.098)	(0.099)
Subjective Plus	-0.568	0.152	0.212*
	(1.220)	(0.123)	(0.125)
Control Mean	17.767		
P-value difference:			
Obj = Sub	0.236	0.735	0.298
Obj = Subplus	0.179	0.517	0.345
Sub = Subplus	0.987	0.329	0.047**
District FE	No	No	No
Division FE	Yes	Yes	Yes
N	116	86	83

Notes: The sample is based on the AD endline survey. In column 1, we include all treatment groups. In column 2 and 3, we exclude the control group given ADs in the control group did not have access to staff-level performance summaries. The estimations in columns 2 and 3 do not include a constant. Errors are clustered at the tehsil level which is the unit of randomization. Estimates are significant at the *10%, **5%, and ***1% level.

Table 6: Correlates of Bonus Assignment

		(1)	(2)	(3)	(4)	(5)	(6)
		Coefficients			P-value Difference		
		Obj	Sub	Subplus	Obj-Sub	Obj-Subplus	Sub-Subplus
Panel A: Full Sample							
Agrismart Indicators	Agrismart compliance	652.642***	288.499***	366.555***	0.000***	0.000***	0.065*
	(attendance and hours)	(18.274)	(24.571)	(29.220)			
	Agrismart visits	19.225***	7.430***	12.531***	0.000***	0.000***	0.000***
	(village and farmer)	(0.391)	(0.817)	(0.887)			
Ability	Years of schooling	15.336	98.239***	17.945	0.017***	0.942	0.046**
		(19.799)	(25.610)	(27.010)			
	Years of experience	-4.400	141.979***	140.024***	0.000***	0.000***	0.944
		(13.465)	(18.024)	(20.452)			
	Worker knowledge	4.809	-67.457***	64.447***	0.004***	0.028***	0.000***
		(15.670)	(20.029)	(22.242)			
Connection with AD	Same zaat as AD	5.775	94.032***	-40.770**	0.008***	0.073*	0.000***
		(15.140)	(28.889)	(18.794)			
Personality	PSM Index	-40.141	35.876	174.777***	0.238	0.006***	0.098*
		(42.401)	(53.317)	(65.475)			
	BFI Index	18.172	-41.562	105.821***	0.139	0.028**	0.002***
		(21.774)	(32.298)	(34.456)			
	Observations	11,580	6,274	7,066			
Panel B: Subsample with FTP Assignments							
Agrismart Indicators	Agrismart compliance	732.860***	329.000***	378.386***	0.000***	0.000***	0.484
	(attendance and hours)	(33.422)	(44.190)	(47.710)			
	Agrismart visits	24.387***	8.192***	12.628***	0.000***	0.000***	0.021**
	(village and farmer)	(0.609)	(1.192)	(1.426)			
	Agrismart FTPs completed	121.817***	126.958***	247.434***	0.915	0.030**	0.039**
		(34.732)	(34.446)	(45.314)			
Ability	Years of schooling	94.236***	60.095	-5.878	0.540	0.070*	0.287
		(31.504)	(41.562)	(40.316)			
	Years of experience	-67.861***	75.228**	102.249***	0.000***	0.000***	0.545
		(22.239)	(30.439)	(31.653)			
	Worker knowledge	5.034	-61.709*	54.995*	0.111	0.205	0.009***
		(25.665)	(32.172)	(32.162)			
Connection with AD	Same zaat as AD	21.231	54.925	4.970	0.473	0.663	0.319
		(24.192)	(40.985)	(27.481)			
Personality	PSM Index	121.996*	55.913	180.427*	0.530	0.637	0.340
		(70.121)	(86.388)	(100.203)			
	BFI Index	23.209	-81.393*	113.371**	0.100*	0.146	0.006***
		(36.374)	(48.861)	(51.867)			
	Observations	4,448	2,631	2,899			
	Tehsil FE	Yes	Yes	Yes			
	Year-Month FE	Yes	Yes	Yes			

Notes: The dependent variable is the bonus amount assigned to staff in a given month. Independent variables are normalized z-scores. Standard errors are clustered at the tehsil level. The dataset is a staff-month panel from August 2018 to January 2021, excluding April–July 2020. It is a subsample including FAs in treatment groups for whom bonuses were assigned. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Correlates of Bonus Assignment (based on AD perceptions of staff quality)

	(1)	(2)	(3)	(4)	(5)	(6)
	Coefficients			P-value difference		
	Obj	Sub	Subplus	Obj-Sub	Obj-Subplus	Sub - Subplus
Good knowledge about Agri.	-279.706**	-19.877	-102.669	0.034**	0.312	0.600
	(111.564)	(80.099)	(170.087)			
Hardworking	61.561	-39.831	152.759	0.395	0.601	0.278
	(80.749)	(101.987)	(193.487)			
Good communicator	378.216***	118.033	159.271	0.078*	0.177	0.785
	(111.025)	(100.910)	(129.306)			
Good at conducting FTPs	106.753	377.973***	790.530***	0.047**	0.000***	0.033**
	(81.510)	(104.941)	(181.042)			
Good at giving farmers advice	-53.630	-269.323**	-206.020	0.218	0.435	0.752
	(117.402)	(123.141)	(176.164)			
Good at using Agrismart	207.563***	310.514***	-84.691	0.389	0.028**	0.002***
	(79.265)	(91.874)	(113.545)			
Observations	2,649	1,958	1,649			
Tehsil FE	Yes	Yes	Yes			
Year-Month FE	Yes	Yes	Yes			

Notes: The sample is constructed based on two datasets - the Agrismart database which includes the bonus amount assigned to staff in a specific month and AD ranking of a selected sub-sample of FAs in the AD endline survey. The final sample used in these specifications is the staff-month panel for FAs who were ranked by ADs. The independent variables are normalised z-scores so their unit of analysis is comparable. Regressions include tehsil fixed effects. Estimates are significant at the *10%, **5%, and ***1% level.

Table 8: Estimated Financial Burden of the Incentive Programme

Item	**Amount (PKR)**
Provincial TADA Allocation FY19–20	350 million
Burden of Payments FY19–20: Objective scheme	65.2 million
Burden of Payments FY19–20: Subjective scheme	118.6 million
Burden of Payments FY19–20: Subjective Plus scheme	145.9 million
Estimated Burden for Out-of-Station TADA	59.2 million
Estimated Liabilities (FY18–19)	50.5 million
Total Expected Requirement: Objective scheme	174.0 million
Total Expected Requirement: Subjective scheme	228.3 million
Total Expected Requirement: Subjective Plus scheme	255.7 million
Total Expected Requirement: Business-as-usual	106.0 million
Total Requirement as % of Allocation: Objective scheme	49.7%
Total Requirement as % of Allocation: Subjective scheme	65.2%
Total Requirement as % of Allocation: Subjective Plus scheme	73.1%
Total Requirement as % of Allocation: Business-as-usual	30.3%

Notes: The expected requirement is the sum of bonus payments, estimated liabilities, and the out-of-station budget.

Appendix

Appendix A: Figures and Tables

A. 1: Descriptive Statistics of Staff Performance Variables

	Mean	SD	P25	Median	P75
Basic FA/AO Characteristics					
Age	42.9	11.5	33	46	53
Salary (in PKR)	39,689	28,088	27,000	37,200	43,500
Years of Experience	18.7	12.3	8	23	30
FA/AO Performance					
Days Worked/month	13.3	6.9	9	14	19
Hours Worked/month	7.6	4.6	4	7	11
Village Visits/month	13.3	11.9	4	11	19
Farmer Visits/month	17.7	23.2	3	11	24
Small Farmer Visits/month	9.4	12.7	1	5	13
Prop. Time on Extension/month	0.55	0.30	0.34	0.59	0.79
Distance Travelled (Kms)/month	368.5	1,172.5	108	226	413
Other Variables					
Avg Claim (yearly)	26,575	40,205	0	10,500	31,485
Avg Payment (yearly)	18,290	31,693	0	6,300	18,000
Prop. of Villages with Metal Road	0.83	0.12	0.76	0.84	0.91
Observations	2,595				

Notes: All variables are based on monthly averages unless otherwise indicated. Percentiles refer to the 25th (P25), 50th (Median), and 75th (P75) percentiles.

A. 2: Randomization Balance

	Mean				P-value Differences					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Control	Objective	Subjective	Subjective +	C - Obj	C - Sub	C - Sub+	Obj - Sub	Obj - Sub+	Sub - Sub+
FA/AO Characteristics										
Age	41.96	43.67	43.80	42.03	0.079	0.053	0.948	0.846	0.054	0.034**
	(0.82)	(0.49)	(0.48)	(0.66)						
Salary	38598	41073	39181	39792	0.215	0.754	0.580	0.199	0.479	0.711
	(1642)	(1112)	(852)	(1402)						
Years in Service	17.37	19.24	19.86	18.10	0.049**	0.010***	0.476	0.426	0.208	0.047**
	(0.83)	(0.56)	(0.55)	(0.68)						
Days Worked/month	14.17	12.82	12.91	13.54	0.222	0.252	0.599	0.941	0.581	0.629
	(0.77)	(0.54)	(0.55)	(0.67)						
Hours Worked/month	7.47	7.59	7.27	7.92	0.837	0.752	0.479	0.619	0.613	0.359
	(0.40)	(0.42)	(0.48)	(0.52)						
Village Visits/month	12.97	13.05	13.10	14.28	0.956	0.932	0.494	0.977	0.523	0.569
	(1.13)	(1.06)	(1.33)	(1.58)						
Farmer Visits/month	17.51	17.00	16.68	19.90	0.833	0.756	0.461	0.901	0.373	0.345
	(1.83)	(1.63)	(1.93)	(2.78)						
Small Farmer Visits/month	9.43	10.02	8.37	9.86	0.699	0.487	0.801	0.259	0.921	0.370
	(1.12)	(1.01)	(1.05)	(1.29)						
Prop. Time on Extension/month	0.56	0.57	0.53	0.54	0.784	0.436	0.403	0.310	0.212	0.841
	(0.02)	(0.02)	(0.04)	(0.02)						
Distance Travelled (Kms)/month	375.83	344.60	347.88	409.50	0.736	0.765	0.772	0.968	0.544	0.567
	(72.86)	(56.86)	(57.71)	(90.42)						
Avg Claim (yearly)	21859	30215	28178	25139	0.304	0.571	0.695	0.855	0.544	0.794
	(5693)	(5706)	(9820)	(6142)						
Avg Payment (yearly)	15824	21317	19401	16074	0.423	0.716	0.969	0.845	0.408	0.733
	(4845)	(4759)	(8772)	(4174)						
Joint F-Test					0.378	0.110	0.308	0.158	0.083*	0.524
Observations	610	682	675	628						
Tehsil Variables										
Prop. of cultivated area	0.765	0.772	0.795	0.808	0.830	0.341	0.143	0.457	0.206	0.601
	(0.02)	(0.02)	(0.02)	(0.02)						
Prop. of Villages with Metal Road	0.787	0.842	0.841	0.811	0.094*	0.089*	0.443	0.994	0.254	0.243
	(0.03)	(0.02)	(0.02)	(0.02)						
Prop. of Villages with Brick Street	0.329	0.499	0.516	0.432	0.030**	0.013**	0.160	0.828	0.384	0.254
	(0.05)	(0.06)	(0.05)	(0.05)						
Prop. of Villages with Housing	0.023	0.024	0.025	0.022	0.656	0.436	0.720	0.682	0.274	0.093*
	(0.002)	(0.001)	(0.001)	(0.05)						
Prop. of Villages with Canal	0.616	0.645	0.669	0.680	0.756	0.546	0.465	0.795	0.702	0.892
	(0.06)	(0.07)	(0.06)	(0.06)						
Joint F-Test					0.366	0.118	0.426	0.983	0.248	0.162
Observations	32	32	34	33						

Notes: The first four columns report means and standard deviations (in parentheses) of different variables across treatment groups. Columns (5)-(10) report p-values for equality of means between treatment arms. We also report the p-value of the joint F-test of individual characteristics between groups. Estimates are significant at the *10%, **5%, and ***1% levels.

A. 3: Treatment Balance (on Sample that Remained in the Study)

	Means				P-Values					
	Control	Objective	Subjective	Subjective+	C-Obj	C-Sub	C-Sub+	Obj-Sub	Obj-Sub+	Sub-Sub+
FA/AO Characteristics										
Age	40.689	42.136	42.314	40.921	0.148	0.091*	0.833	0.806	0.175	0.098*
	(0.818)	(0.542)	(0.479)	(0.685)						
Salary	36,640	38,676	36,925	38,041	0.293	0.875	0.516	0.753	0.550	
	(1478)	(1244)	(996)	(1572)						
Years of Experience	17.522	19.380	19.486	18.757	0.066*	0.046**	0.260	0.902	0.541	0.444
	(0.786)	(0.637)	(0.589)	(0.743)						
Days Worked/month	14.008	12.706	13.301	13.375	0.222	0.516	0.611	0.607	0.610	0.956
	(0.694)	(0.799)	(0.825)	(1.034)						
Hours Worked/month	7.532	7.636	7.385	7.895	0.858	0.814	0.598	0.690	0.711	0.493
	(0.407)	(0.410)	(0.476)	(0.568)						
Village Visits/month	12.527	12.939	13.558	14.369	0.773	0.523	0.331	0.709	0.469	0.701
	(0.958)	(1.045)	(1.300)	(1.661)						
Farmer Visits/month	17.246	16.555	17.614	20.397	0.770	0.891	0.358	0.669	0.263	0.444
	(1.829)	(1.500)	(1.967)	(3.033)						
Small Farmer Visits/month	8.902	9.763	8.912	10.222	0.550	0.995	0.449	0.559	0.792	0.463
	(1.049)	(0.970)	(1.076)	(1.423)						
Prop. Time on Extension/month	0.563	0.577	0.533	0.544	0.642	0.518	0.587	0.305	0.273	0.806
	(0.025)	(0.017)	(0.039)	(0.243)						
Distance Travelled (Kms)/month	379	357	369	401	0.840	0.927	0.854	0.896	0.686	0.764
	(85.289)	(66.620)	(63.193)	(86.805)						
Avg TA Payment (yearly)	16,260	20,866	18,424	15,168	0.537	0.830	0.872	0.805	0.375	0.735
	(5438)	(5011)	(8725)	(3991)						
Avg TA Claim (yearly)	22,128	29,267	27,172	23,778	0.414	0.653	0.847	0.850	0.514	0.764
	(6215)	(6011)	(9627)	(5887)						
Joint F-Test					0.155	0.169	0.115	0.265	0.260	0.648
Observations	520	575	568	536						

Notes: Columns 1-4 report means (standard deviations in parentheses) for each group. Columns 5-10 report p-values for differences between groups. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A. 4: Relationship between Attrition and Treatment groups

	(1)
	Attrition
Objective	0.009
	(0.027)
Subjective	0.011
	(0.026)
Subjective Plus	-0.001
	(0.027)
Observations	2,595

Notes: Attrition is defined as a dummy variable equal to 1 for staff (or tehsil) that left the programme. Estimates are significant at the *10%, **5%, and ***1% level.

A. 5: Difference between Exiting and Remaining Staff

	(1)	(2)	(3)
Variable	Exiting	Remaining	P-value difference
Age	50.455	41.544	0.000***
	(0.347)	(0.557)	
Salary (in PKR)	51,360	37,588	0.000***
	(659)	(1668)	
Experience (yrs in service)	27.382	18.801	0.000***
	(0.381)	(0.664)	
Days worked/month	13.354	13.331	0.960
	(0.447)	(0.598)	
Hours worked/day	7.266	7.610	0.195
	(0.247)	(0.319)	
Village visits/month	13.306	13.350	0.955
	(0.688)	(1.014)	
Farmer visits/month	16.692	17.929	0.405
	(1.134)	(1.692)	
Small farmer visits/month	9.207	9.452	0.798
	(0.597)	(1.038)	
Prop. time on extension/month	0.521	0.554	0.035**
	(0.014)	(0.020)	
Distance Traveled (in Kms)/month	326.607	376.044	0.144
	(37.780)	(31.492)	
Observations	396	2,199	2,595

Notes: Estimates are significant at the *10%, **5%, and ***1% level.

A. 6: Treatment Effects on Staff Outreach by Quarter

		Days worked	Hours worked	Village visits	Performance index
Quarter 1	Q1 Objective	0.100	1.064***	1.580	0.119*
		(0.351)	(0.406)	(1.913)	(0.069)
	Q1 Subjective	0.683**	0.282	0.335	0.026
		(0.298)	(0.352)	(1.298)	(0.054)
	Q1 Subjective Plus	0.184	0.638*	0.631	0.052
		(0.413)	(0.362)	(1.445)	(0.065)
Quarter 2	Q2 Objective	0.957**	1.321***	6.074***	0.329***
		(0.381)	(0.333)	(2.042)	(0.080)
	Q2 Subjective	0.982***	0.746***	3.658**	0.187***
		(0.361)	(0.285)	(1.496)	(0.065)
	Q2 Subjective Plus	0.554	0.989***	3.293*	0.203**
		(0.502)	(0.311)	(1.672)	(0.081)
Quarter 3	Q3 Objective	0.896**	0.584*	7.359***	0.303***
		(0.355)	(0.298)	(2.021)	(0.081)
	Q3 Subjective	0.902**	0.442*	4.260***	0.153**
		(0.364)	(0.263)	(1.571)	(0.064)
	Q3 Subjective Plus	1.112***	0.676**	4.905***	0.227***
		(0.407)	(0.315)	(1.395)	(0.064)
Quarter 4	Q4 Objective	0.715**	0.812***	9.141***	0.381***
		(0.353)	(0.302)	(2.003)	(0.075)
	Q4 Subjective	0.611	0.647***	4.279***	0.168**
		(0.370)	(0.230)	(1.490)	(0.064)
	Q4 Subjective Plus	0.417	1.164***	2.860*	0.164**
		(0.491)	(0.296)	(1.676)	(0.072)
Quarter 5	Q5 Objective	1.030***	1.155***	9.236***	0.390***
		(0.334)	(0.326)	(2.159)	(0.079)
	Q5 Subjective	0.634**	0.663***	3.268**	0.147**
		(0.314)	(0.239)	(1.349)	(0.061)
	Q5 Subjective Plus	0.658**	1.328***	3.505**	0.200***
		(0.295)	(0.268)	(1.746)	(0.065)
Quarter 6	Q6 Objective	0.974***	1.092***	8.753***	0.378***
		(0.313)	(0.334)	(2.262)	(0.092)
	Q6 Subjective	0.596*	0.593**	2.767**	0.157**
		(0.312)	(0.275)	(1.385)	(0.069)
	Q6 Subjective Plus	0.744*	1.192***	5.549***	0.270***
		(0.392)	(0.314)	(1.841)	(0.074)
Quarter 7	Q7 Objective	1.349***	0.855***	15.250***	0.545***
		(0.355)	(0.305)	(2.858)	(0.095)
	Q7 Subjective	0.769**	0.697**	3.935**	0.184***
		(0.355)	(0.269)	(1.572)	(0.068)
	Q7 Subjective Plus	1.225***	1.144***	8.632***	0.367***
		(0.326)	(0.292)	(1.950)	(0.070)
Quarter 8	Q8 Objective	1.543***	1.105***	17.640***	0.655***
		(0.455)	(0.334)	(3.742)	(0.125)
	Q8 Subjective	0.731	0.776**	3.645**	0.182**
		(0.492)	(0.306)	(1.779)	(0.081)
	Q8 Subjective Plus	1.282***	1.347***	8.411***	0.362***
		(0.429)	(0.315)	(3.049)	(0.106)
Quarter 9	Q9 Objective	1.771***	1.411***	18.377***	0.706***
		(0.604)	(0.393)	(4.136)	(0.157)
	Q9 Subjective	1.199*	0.809**	4.460***	0.253***
		(0.628)	(0.384)	(1.513)	(0.086)
	Q9 Subjective Plus	1.523**	1.436***	5.523***	0.317***
		(0.609)	(0.435)	(2.013)	(0.099)
Quarter 10	Q10 Objective	2.204***	1.176***	17.846***	0.670***
		(0.731)	(0.367)	(3.869)	(0.143)
	Q10 Subjective	0.941	0.674*	3.868**	0.195*
		(0.768)	(0.351)	(1.903)	(0.101)
	Q10 Subjective Plus	1.682**	1.412***	5.127**	0.301***
		(0.689)	(0.389)	(2.248)	(0.105)
Observations		58,768	58,768	58,768	58,768
District Fixed Effects		Yes	Yes	Yes	Yes

Notes: This table presents treatment effects on staff outreach by quarter, covering three quarters in 2018, four in 2019, and three in 2020, excluding the Covid-19 period (April–July 2020). Standard errors clustered at the tehsil level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A. 7: Inactive Users

	(1)
Objective	-0.012
	(0.035)
Subjective	-0.016
	(0.035)
Subjective Plus	-0.019
	(0.037)
Control Mean	0.114
P-value (difference tests):	
Obj = Sub	0.863
Obj = Subplus	0.740
Sub = Subplus	0.887
District FE	Yes
Month FE	Yes
Year FE	Yes
N	73,023

Notes: This sample is based on a staff-month panel covering a 30-month period from April 2018 to January 2021 (excluding the peak COVID-19 months from April–July 2020). The panel includes ITT workers who remained in the department, including inactive users. Approximately 11.5% of observations are missing, explaining the larger sample size compared to Table 1. Standard errors are clustered at the tehsil level (the unit of randomization) and shown in parentheses. All regressions include district, year, and month fixed effects. Estimates are significant at the *10%, **5%, and ***1% level.

A. 8: Reporting Behaviour of FAs and AOs

	(1) Days Not Logged In	(2) Monthly Under-reported Hours	(3) % Village Visits Reported	(4) % Farmer Visits Reported	(5) % FTPs Reported
Objective	-0.312	0.483	0.892	1.379	0.705
	(0.246)	(0.412)	(1.559)	(1.854)	(2.166)
Subjective	-0.297	-0.420	1.333	-0.258	1.323
	(0.245)	(0.420)	(1.316)	(1.700)	(2.197)
Subjective Plus	-0.581**	-0.499	1.869	0.620	4.319*
	(0.268)	(0.441)	(1.758)	(1.846)	(2.247)
Control Mean	0.792	5.176	88.890	88.035	84.826
Obj = Sub	0.937	0.008***	0.774	0.339	0.781
Obj = Subplus	0.205	0.006***	0.611	0.689	0.103
Sub = Subplus	0.183	0.828	0.742	0.595	0.168
District FE	Yes	Yes	Yes	Yes	Yes
N	1599	1598	1598	1598	1597

Notes: Errors are clustered at the tehsil level which is the unit of randomization. All estimations include district FE. Estimates are significant at the *10%, **5%, and ***1% level. The dependent variables are constructed from the staff endline survey where staff were asked about the number of hours or days that are under-reported in a month, as well as the proportion of village visits, farmer visits, and FTPs that are reported in a month.

A. 9: Impact on Farmer Access to Extension

	(1) Knows AO/FA	(2) Informed about activities	(3) Had Grp Meeting	(4) Attend Ind Meeting
Objective	0.002	0.012	0.020**	0.006
	(0.017)	(0.012)	(0.010)	(0.005)
Subjective	0.030	0.036**	0.016	0.023***
	(0.021)	(0.015)	(0.011)	(0.007)
Subjective Plus	0.026	0.039***	0.023*	0.015**
	(0.021)	(0.013)	(0.012)	(0.006)
Control Mean	0.409	0.581	0.150	0.058
P-value differences:				
Obj = Sub	0.125	0.065	0.744	0.017
Obj = Subplus	0.232	0.054	0.815	0.247
Sub = Subplus	0.874	0.876	0.644	0.362
District FE	Yes	Yes	Yes	Yes
Year-Month dummies	Yes	Yes	Yes	Yes
N	13,312	13,312	13,312	13,312

Notes: The sample is based on a farmer phone survey, including 8 different survey rounds between January 2020 - January 2021 (we randomly sampled one farmer against each ITT employee in each round). Estimates are significant at the *10%, **5%, and ***1% level

A. 10: Treatments Effects in Audit Data

	Valid Number	Knows Staff	Attend Grp Meeting	No in Grp Meeting	Attend Ind Meeting
Objective	-0.008	0.006	-0.008	-1.295	-0.012
	(0.012)	(0.011)	(0.024)	(5.324)	(0.012)
Subjective	-0.003	-0.002	0.001	4.967	0.010
	(0.011)	(0.010)	(0.022)	(6.342)	(0.012)
Subjective PLUS	0.026**	-0.020*	-0.029	-2.679	0.009
	(0.012)	(0.011)	(0.025)	(5.457)	(0.013)
Control Mean	0.692	0.816	0.630	41.306	0.734
P-value differences:					
Obj = Sub	0.643	0.367	0.697	0.203	0.070
Obj = Subplus	0.005	0.008	0.402	0.747	0.083
Sub = Subplus	0.005	0.051	0.160	0.167	0.884
District FE	Yes	Yes	Yes	Yes	Yes
N	30,543	20,034	4,903	3,078	15,131

Notes: Errors are clustered at the tehsil level which is the unit of randomization. Estimates are significant at the *10%, **5%, and ***1% level.

A. 11: Payments across treatment arms

	(1) Months with paid bonus (Jan-Jun 2020)	(2) % with belief bonus will not be paid
Objective	4.248***	0.218***
	(0.223)	(0.036)
Subjective	3.439***	0.245***
	(0.391)	(0.036)
Subjective Plus	4.032***	0.228***
	(0.361)	(0.037)
p-value difference		
Obj - Sub	0.08*	0.61
Obj - Subplus	0.61	0.85
Sub- Subplus	0.27	0.75
Observations	1418	1345

Notes: Each specification includes treatment arms only and is estimated without a constant. Since the specifications exclude the control staff, the observations are fewer than total ITT staff of 1,975. About 94 staff members did not respond to the question in column 1. In column 2, the observations reduce further since 73 employees report that their bonuses were fully paid hence the question did not apply to them. Errors are clustered at the tehsil level. Estimates are significant at the *10%, **5%, and ***1% level.

A. 12: Impact on AO and FA Motivation

	(1) Dictator game donation (out of PKR 100)	(2) Job satisfaction index	(3) Self efficacy index
Objective	4.096**	0.155*	0.248***
	(1.976)	(0.080)	(0.066)
Subjective	2.617	0.098	0.127*
	(2.034)	(0.064)	(0.070)
Subjective Plus	3.900*	0.042	0.138*
	(2.242)	(0.066)	(0.076)
Control Mean	15.635	0.009	-0.114
P-value difference:			
Obj = Sub	0.462	0.448	0.076*
Obj = Subplus	0.927	0.143	0.140
Sub = Subplus	0.546	0.381	0.890
District FE	Yes	Yes	Yes
N	1884	1884	1883

Notes: The sample is based on an ITT sample of AOs and FAs at endline. Out of the total sample of 2199 ITT employees, about 1884 responded to the phone survey (response rate of 86%). Standard errors are clustered at the tehsil level presented within parenthesis. Estimates are significant at the *10%, **5%, and ***1% level.

Appendix B: Design Details

Designation	Payment Categories			
	No Payment	Base Payment*	Low Bonus (Base + Bonus)	High Bonus (Base + Bonus)
FA	0	2500	4900 (2500 + **2400**)	7300 (2500 + **4800**)
AI	0	3000	6000 (3000 + **3000**)	9000 (3000 + **6000**)
AO	0	6000	10500 (6000 + **4500**)	15000 (6000 + **9000**)

Notes: All amounts are in PKR/month. The amounts mentioned here are the maximum possible amounts that can be received under Base Payment. The actual amount of Base Payment will be pro-rated based on the number of days worked in a month. The exact method of pro-rating is specified below.

B. 1: Bonus Payment Categories under the 3 Incentive Schemes

Indicator	No Payment	Base Payment*	Low Bonus	High Bonus
Number of days	<22	22	22	22
Hours per day	<6	6	6	6
Villages visited			20	20
% of time spent on Extension			50%	65%
% of scheduled FTP conducted			70%	85%
Farmers per FTP			Irrigated areas: 20 Barani: 10	Irrigated areas: 25 Barani: 15
Farmers visited			48	72
Small farmers visited			24	36
Distance travelled			400	400

Assumptions:
Farmers visited: This assumes that out of a week of 6 days, an FA spends 1 day on FTP and 1 day on office activities. Out of the remaining 4 days, if soil sampling is being conducted on one of the days, it is not possible to visit more than 3 farmers. For the other 2 days, a bare minimum of 3 farmers must be visited. Based on this, a minimum threshold of 12 farmers per week is suggested which leads to around 48 farmers a month. For high bonus, the threshold increases by a factor of 1.5, i.e. 18 farmers are visited per week leading to the threshold of 72 farmers. Note that an FTP often takes a full work day because an FA may have to bring farmers to the training and drop them back after the training is completed.

Village visited: Out of a total of 5 working days (with 1 day reserved for potential office tasks), an FA is expected to visit a different village every day. To do his job well, an FA must spend a full day in a village trying to reach out to as many farmers as possible. This gives the threshold 20 villages/month. The village visited threshold for high bonus remains the same since higher than 20 villages per month can induce the FA in not maximizing his outreach within each village.

Distance travelled: To do his/her basic tasks, an FA needs to travel a minimum of 20 kms daily on average, which gives a threshold of 400 kms per month (assuming FA is in the field 5 days in a week). The threshold remains the same for high bonus since more distance travelled will not necessarily imply higher performance.

B. 2: Payment category thresholds for FAs/AIs in Objective Measure Scheme

Indicator	No Payment	Base Payment*	Low Bonus	High Bonus
Number of days	<22	22	22	22
Hours per day	<6	6	6	6
Villages visited			20	20
% of time spent on Extension			50%	65%
% of scheduled FTP conducted			70%	85%
Farmers per FTP			Irrigated areas: 20 Barani: 10	Irrigated areas: 25 Barani: 15
Farmers visited			32	48
Small farmers visited			16	24
Distance travelled			640	640

Assumptions:
Farmers visited: This assumes that out of a week of 6 days, an AO spends at least 5 days doing FTPs, hence individual farmer outreach for AOs is lesser. It was proposed that a realistic low bonus threshold for AOs was around 8 farmers/week which leads to 32 farmers /month. The high bonus threshold is 1.5 times higher than this.

Village visited: Out of a total of 5 working days (with 1 day reserved for potential office tasks), an AO is expected to visit a different village every day. This gives the threshold 20 villages/month. The village visited threshold for high bonus remains the same since higher than 20 villages per month can induce the FA in not maximizing his outreach within each village.

Distance travelled: To do his/her basic tasks, an AO travels a minimum of 32 kms daily on average, which gives a threshold of 640 kms per month (assuming AOs are in the field 5 days a week). The threshold remains the same for the high bonus since more distance travelled will not necessarily imply higher performance.

B. 3: Payment category thresholds for AOs in Objective Measure Scheme