NBER WORKING PAPER SERIES

HIERARCHY AND PERFORMANCE PAY: EXPERIMENTAL EVIDENCE FROM THE PUBLIC SECTOR

Garance Genicot
Zahra Mansoor
Ghazala Mansuri

Working Paper 34825
http://www.nber.org/papers/w34825

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2026

We are indebted to Captain Mehmood (former Secretary Agriculture, Punjab) for championing the project, Zubair Bhatti (World Bank Project Lead) for helping us initiate it and Dr. Anjum Ali (Director General Extension Services, Punjab) for his support during project implementation. We thank Ali Cheema, Imran Rasul, Clare Leaver, Christopher Woodruff, Simon Quinn, Julien Labonne, Stefan Dercon, Martin Williams, and participants at APSA (2020), CSAE (2019, 2020), SEEDEC 2025), BREAD South Asia (2025), KU Leuven and the University of Delaware for helpful comments. Ayesha Shahid and Usman Zahid provided invaluable implementation support. Research assistance by Saad Imtiaz, Maria Qazi and Bhavya Srivastava is gratefully acknowledged. Funding was provided by the World Bank i2i, World Bank Strategic Research Programme, Georgetown Center for Economic Research, and the International Growth Center. AEA registry: AEARCTR-0002430. Ethical approval: Georgetown University IRB. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Hierarchy and Performance Pay: Experimental Evidence from the Public Sector
Garance Genicot, Zahra Mansoor, and Ghazala Mansuri
NBER Working Paper No. 34825
February 2026
JEL No. D23, O12, Q16

ABSTRACT

Frontline workers in public bureaucracies perform multidimensional tasks and rely on supervisors who both monitor performance and supply essential managerial inputs. We study how incentive design should account for these interactions using a conceptual framework and a province-wide randomized experiment with the Punjab Agriculture Extension Department in Pakistan. Across 131 tehsils, we compare: (i) an Objective pay-for-performance scheme tied to digital activity metrics; (ii) a Subjective scheme in which supervisors allocate bonuses; and (iii) a Subjective Plus scheme that preserves discretion but introduces light-touch oversight of supervisors.

All three schemes increase outreach, but through distinct channels. Objective incentives primarily raise intensive effort—repeat visits to the same farmers—while Subjective Plus expands the extensive margin by inducing supervisors to schedule more farmer trainings and broaden geographic coverage. Contrary to standard multitasking concerns, meeting length—a proxy for engagement quality—increases across all arms.

Oversight of supervisors also reduces favoritism in bonus allocation and improves merit-based evaluation, leading to greater access and yield gains among marginalized farmers. The results show that light-touch monitoring of middle management can substantially amplify the effectiveness and equity of performance pay in multi-layered bureaucracies.

Garance Genicot
Georgetown University
Department of Economics
and NBER
gg58@georgetown.edu

Zahra Mansoor
Bristol University
zahra.mansoor@bristol.ac.uk

Ghazala Mansuri
The World Bank
gmansuri@gmail.com

A randomized controlled trials registry entry is available at AEA registry: AEARCTR-0002430

1 Introduction

In hierarchical public organizations, frontline workers perform multidimensional tasks and depend on supervisors for both monitoring and critical managerial inputs. Designing incentives in such settings is challenging. Service delivery typically requires complementary effort across multiple hierarchical layers (Wilson, 1989; Garicano, 2000), and frontline roles involve tasks that vary widely in observability and measurability (Holmstrom and Milgrom, 1991). Supervisors therefore play a pivotal role – they both evaluate frontline performance and supply their own managerial inputs such as planning, supervision, and coordination that shape production. By influencing how effort is allocated across measured and unmeasured dimensions, supervisors can amplify or attenuate the effects of incentives. This raises a central question for the economics of organizations and public administration: *how should incentives be structured in systems where supervisors play a pivotal role in shaping the effectiveness of frontline workers?*

Theory offers competing predictions. Linking rewards to objective, verifiable metrics can increase performance on targeted dimensions but may distort effort away from harder-to-measure tasks (Holmstrom and Milgrom, 1991). Allowing supervisors discretion in allocating rewards can incorporate richer information about quality, outreach, or contextual constraints, but also creates scope for favoritism, bias, or weak managerial engagement (Gibbs et al., 2004; Prendergast, 2007; Neal, 2011). A third approach combines discretion with oversight, aiming to discipline supervisors' use of private information while preserving incentives to provide complementary managerial inputs (Prendergast and Topel, 1996; Dal Bó et al., 2021). Empirically, performance pay tied to objective metrics has improved targeted outcomes in public-sector settings, including teacher attendance (Duflo et al., 2012), student learning (Glewwe et al., 2010; Muralidharan and Sundararaman, 2011; Leaver et al., 2021), health service provision (Basinga et al., 2011), and tax collection (Khan, Khwaja, and Olken, 2016). However, empirical evidence directly comparing objective and subjective incentive schemes is limited, largely confined to settings without hierarchical production roles, such as private firms or schools (Andrabi and Brown, 2022). Moreover, evidence on incentive systems that

operate across multiple hierarchical layers – where supervisors both evaluate and supply complementary inputs – remains scarce (Behrman et al., 2015; Deserranno et al., 2022).

This paper addresses this gap by experimentally evaluating how incentive schemes that differentially target frontline workers and supervisors affect service delivery in a large public bureaucracy. We study the Agriculture Extension Department in Punjab, Pakistan, where Agricultural Officers and Field Assistants deliver agronomic advice, conduct group trainings, and provide individualized support to millions of smallholder farmers. Their tasks span easily monitorable actions such as the number of farmer visits, and harder-to-measure dimensions such as engagement quality or the breadth of outreach. Supervisors (Assistant Directors, or ADs) play a central role in shaping performance: they schedule Farmer Training Programs (FTPs), allocate workers across villages, monitor staff activity, and approve leave. This structure makes the setting ideal for examining how incentive design interacts with supervisory behavior in a multi-layered hierarchy.

In partnership with the department, we implement a province-wide performance pay program built on *AgriSmart*, a digital monitoring platform that records daily activities through geotagged mobile reporting and provides supervisors with dashboards to plan and verify fieldwork. We randomly assign 131 tehsils and 2,595 workers to one of three pay-for-performance schemes or a control group. In the *Objective* arm, bonuses are tied to verifiable digital metrics. In the *Subjective* arm, supervisors allocate bonuses at their discretion. In the *Subjective Plus* arm, supervisors retain discretion but are subject to light-touch top-down oversight: monthly reports summarizing staff performance and the supervisor's own managerial inputs are shared with senior management.

We find that all three schemes increase frontline outreach, but along distinct margins. Objective incentives generate the largest gains in observable activity, particularly repeat visits to the same farmers (the intensive margin). Subjective Plus produces the largest gains on the extensive margin, expanding the number of villages and farmers reached, driven by substantial increases in supervisory scheduling of FTPs in comparison to the control group and other treatment arms. Crucially, meeting length—a proxy for engagement quality—increases

by 20–23% across all arms, suggesting that higher outreach does not come at the expense of quality. These patterns highlight the importance of complementary managerial inputs: the Subjective Plus arm induces supervisors to supply more planning effort and to spread outreach more equitably across space.

Farmer outcomes echo these distinctions. Access to extension services increases only in the subjective arms, while the Objective arm's gains in repeat visits do not translate into broader farmer reach. Benefits under Subjective Plus are concentrated among marginalized farmers: both access and yields rise for this group, and quantile treatment effects show stronger impacts in the lower tail of the yield distribution. Together, these results show that neither relying solely on verifiable metrics (Objective) nor solely on discretionary evaluation (Subjective) is sufficient; instead, aligning supervisors with the principal's objectives reshapes both the scale and distribution of service delivery.

The incentive schemes also produce distinct patterns in bonus allocation. Bonuses are larger on average in the subjective arms, consistent with leniency bias. Yet in comparison with the Subejctive arm, supervisors under Subjective Plus condition bonuses more strongly on observable performance and worker knowledge, and substantially less on social ties such as shared *zaat* or *biradari* affiliation. Even minimal but salient oversight thus reorients discretion toward merit-based evaluation.

This paper makes three contributions. First, we develop and test a conceptual framework linking multitasking concerns with supervisory alignment, showing how light-touch oversight can amplify the effectiveness of subjective performance evaluation by increasing managerial inputs and reducing bias. Second, using a randomized experiment in a multi-layered bureaucracy, we provide causal evidence on how alternative incentive regimes reallocate effort across intensive and extensive margins, affect engagement quality, and shape supervisory behavior. Third, we show that minimal but salient monitoring of supervisors improves equity and targeting by expanding outreach to marginalized farmers and reducing favoritism in bonus allocation. Together, our results demonstrate that supervising the supervisors is a scalable and empirically tractable lever for strengthening performance pay in public bureaucracies.

This paper contributes to several strands of literature. First, it advances research on performance pay in the public sector, where production is hierarchical and task performance spans multiple goals and observable and unobservable dimensions (see Manning et al. 2012; Giné, Mansuri, and Shrestha 2022; Burgess and Ratto 2003; Deserranno, Kastrau, and León-Ciliotta 2025a). Building on foundational theories of incentives and delegated authority, recent empirical work documents how monitoring and discretion interact within large bureaucracies (e.g., Best et al., 2022; Bertrand et al., 2019), and we extend this literature by showing that aligning supervisor incentives with the principal's objectives systematically reshapes both managerial inputs and how discretion is exercised. Second, we contribute to the agricultural extension literature by shifting the focus from how information is transmitted to farmers (Fernando, 2016; Beaman et al., 2021) to how incentives for the agents who deliver that information affect outreach, quality, and equity, complementing broader work on the organization of extension systems and the role of social and managerial frictions in technology adoption (Anderson and Feder, 2007; BenYishay and Mobarak, 2019). Finally, we contribute to research on management and organizational practices in the public sector (Rasul and Rogger, 2018; Rasul, Rogger, and Williams, 2021), providing causal evidence that improving supervisory inputs at the subnational level can raise frontline performance and reduce favoritism. By experimentally varying both supervisor discretion and the oversight applied to it, we show that "supervising the supervisors" is a powerful mechanism for strengthening performance pay in multi-layered bureaucracies.

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting. Section 3 presents the conceptual framework. Section 4 details the experimental design. Section 5 describes the data and main outcomes. Section 6 presents the results and Section 7 examines mechanisms. Finally, and Section 8 concludes.

2 Context

Agricultural Context. Punjab accounts for roughly 57% of Pakistan's cultivated land, making agriculture central to provincial policy.[1] Yet agricultural productivity has stagnated in recent years.[2] This slowdown is often linked to poor farming practices, low technology adoption, and limited crop diversification. Smallholder farmers — those cultivating less than five acres and comprising over 60% of all farm households in Punjab — are particularly affected. Improving their access to timely and relevant information could raise productivity, yet outreach to such farmers remains constrained.

Agriculture Extension in Punjab. In Punjab, the Provincial Agriculture Extension Directorate is the main institution responsible for delivering agricultural information to farmers. The department employs roughly 2,600 frontline staff operating down to the village level. Field Assistants (FAs), stationed at the union-council level are the lowest rung of the hierarchy, typically covering 7–8 villages. Agricultural Officers (AOs) work closely with FAs, each supervising 7–8 union councils. These frontline workers are directly overseen by Assistant Directors (ADs), at the tehsil level. Higher tiers include district- and divisional-level directors and provincial headquarters (see Figure 1).

AOs and FAs serve as core frontline staff, organizing Farmer Training Programs (FTPs), providing individual consultations, managing demonstration plots, and completing administrative tasks. FTPs are typically led by AOs, while FAs provide logistical support. Both report to ADs, who manage monthly schedules, assign trainings, approve leave, and conduct unannounced field visits to monitor quality.

Despite this structure, extension services were often perceived as captured by large landowners, with limited investment in staff monitoring or information quality. The department therefore faced a dual challenge: expanding outreach and improving content relevance.

The AgriSmart Program. To address these challenges, the Department of Agriculture

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

[2] https://www.finance.gov.pk/survey/chapter_25/1_Growth_and_Investment.pdf

collaborated with the World Bank to introduce a digital management system.[3] Under this project, our research team began a collaboration with the department and PITB to design *AgriSmart*. The system comprised a smartphone-based staff reporting application and a web-based management dashboard for supervisors.

AgriSmart was deployed province-wide in April 2018.[4] It required staff to submit geo-tagged, time-stamped reports of daily activities, enabling automated tracking of attendance, village and farmer visits, and FTP completions. Supervisors were required to use the dashboard to assign FTPs, approve leave, and manage staff records. To strengthen data integrity, the system incorporated multiple safeguards against misreporting, including geo-verification features and an independent farmer callback audit. Figure 2 displays the *AgriSmart* application interface.

As part of the rollout, the department also introduced Standard Operating Procedures (SOPs) for Assistant Directors (ADs), covering regular FTP scheduling, timely leave approvals, and systematic staff record management at the tehsil level.

Figure 2 about here

Travel and Daily Allowance (TADA). The department receives an annual allocation of a non-salary 'TADA' budget used to reimburse staff for work-related travel expenses. *De jure*, reimbursements are permitted only for out-of-station travel exceeding a 16 km radius from an officer's assigned jurisdiction. *De facto*, however, TADA disbursements have been ad hoc and inconsistently administered. With the improved monitoring capabilities introduced under *AgriSmart* and flexibility within the existing TADA budget, the department repurposed this budget, for the duration of the project, to finance staff incentives. This allowed the department to test incentives without requiring additional funds. It also provided a proof

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

[4]In 2016, extension services were offered in all 36 districts in Punjab. For extension purposes, a district is the third administrative level (following division and sub-division), while a tehsil is the fourth. A district typically has four tehsils. Extension services are not offered in urban tehsils and neighboring tehsils are sometimes merged for administrative purposes. As a result, total extension administrative tehsils in 2016 were 131 while the total number of tehsils in Punjab was 145.

of concept to support a shift toward more accountable, performance-based use of the TADA budget

3 Conceptual Framework

The design of incentive schemes in bureaucracies is inherently complex. The ultimate objective—social welfare—is elusive and difficult to measure. Even if regular measurement were feasible, linking incentives to beneficiary outcomes such as farmers' welfare or agricultural output would introduce considerable noise, resulting in a low signal-to-noise ratio (Baker, 1992; Gibbons, 1998). Consequently, most public-sector incentive schemes reward employee performance rather than final outcomes.[5]

While *AgriSmart* enables precise monitoring of attendance, activities, and locations, other aspects of performance—such as engagement with farmers or the quality of advice—remain difficult to assess. Supervisors, who occupy middle-management positions, typically possess richer information about these dimensions of frontline performance, yet their objectives may diverge from those of the principal. Moreover, supervisors contribute essential complementary inputs such as scheduling, planning, and oversight that are central to worker effectiveness but are not costless. Hence, their own incentives constitute a key component of the system's overall performance.

A principal seeking to improve service delivery must therefore choose between an *objective* scheme, which rewards measurable effort, and a *subjective* scheme, which relies on supervisor discretion. The framework below illustrates the trade-offs between these approaches and the conditions under which aligning middle-management (in our case, the AD) incentives with those of the principal enhances performance under subjective schemes. This framework motivates the design of our field experiment.

[5] An important exception is the education sector, where teachers' incentives are often based on students' test scores (see, among many others, Andrabi and Brown, 2022).

A Model

Assume that worker effort has two dimensions: an observable component e and a harder-to-measure component q. For simplicity, we refer to these as *outreach* (e) and *quality* (q), but the key distinction lies in their observability. The first dimension captures activities that are easily monitored—such as attendance or the number of farmer visits—while q represents aspects of performance that are less observable, such as the quality of advice or the effort required to reach remote or marginalized farmers.

Together with supervisory effort s, which represents managerial inputs such as scheduling, planning, and oversight, these dimensions generate a social benefit $f(e, q, s)$ that is not directly observable. The marginal products of all three inputs are positive and decreasing.

The principal seeks to maximize welfare: the social benefit $f(e, q, s)$ net of workers' cost of effort $c(e, q)$, supervisors' cost $k(s)$, and the cost $\kappa(B)$ of total bonus payments. The principal chooses an incentive scheme anticipating workers' and supervisors' responses. Specifically, an incentive scheme defines the probability $p_i(e_i, q_i, s_{j(i)})$ that worker i receives a bonus b when supervisor $j(i)$ exerts scheduling effort $s_{j(i)}$.

The principal's objective can therefore be written as:

$$\max_{p_i} \sum_i \Big[f(e_i, q_i, s_{j(i)}) - c(e_i, q_i) - k(s_{j(i)}) \Big] - \kappa(B), \tag{1}$$

where $B = \sum_i p_i(e_i, q_i, s_{j(i)})\, b$ is total bonus expenditure.

Workers derive utility from both monetary rewards (*extrinsic motivation*) and the social value of their activities (*intrinsic motivation*). Let $\alpha_i \in [0, 1]$ capture the relative weight on extrinsic motivation. Given a bonus plan $p_i(e, q, s)$ and supervisor effort s, worker i solves:

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

Under standard assumptions of differentiability and concavity, the worker's first-order con-

ditions are:

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

These conditions highlight that incentives operate through two distinct channels: by increasing the marginal return to observable effort e, and by influencing how workers substitute between e and q depending on their technological relationship. If e and q are complements, incentives raising e will also increase q; if they are substitutes, stronger incentives on e may crowd out q.

Denote the resulting optimal effort choices $(\widehat{e}, \widehat{q})$. When $\alpha_i = 0$, the worker is fully intrinsically motivated and no bonus is required. For $\alpha_i > 0$, the worker exerts too little effort without incentives, and performance pay can increase both margins depending on how $p_i(e, q, s)$ is structured.

Objective Incentives. Under an objective scheme, $p_i = p(e)$ depends only on observable effort e. When $\alpha_i > 0$, a higher bonus b increases $\widehat{e}_i$. The effect on $\widehat{q}$ depends on whether e and q are complements or substitutes. If $c_{eq} < 0$, raising b increases both e and q; if $c_{eq} > 0$, raising b increases e but lowers q.

Although supervisors are not directly incentivized in this scheme, their scheduling effort s could nonetheless increase when e and s are complementary in production ($f_{es} > 0$), indirectly strengthening performance.

Subjective Incentives. Under a subjective scheme, supervisors allocate bonuses among workers. They observe e as the principal does but also possess private information about q.[6] Subject to feasibility constraints, supervisor j assigns bonuses and chooses a mapping $p_i(e, q, s)$, while also selecting scheduling effort s, anticipating workers' optimal responses.

Supervisor j derives utility from both private preferences and alignment with the principal's

[6] Alternatively, information on q may be available to supervisors only at a cost.

objective. Let $\beta_j \in (0,1)$ denote the weight on private preferences, with $1-\beta_j$ capturing alignment with the principal. Lower β_j may reflect stronger mission motivation or tighter monitoring. The supervisor's private preferences are $V_j(\mathbf{p}) = \sum_i \omega_{ji}\, u(p_i(e_i, q_i)b)$, allowing for favoritism or centrality bias through the weights ω_{ji} (Gibbs et al., 2004; Prendergast, 2007; Neal, 2011). Supervisor j solves:

$$\max_{s_j, \mathbf{p}}; \beta_j V_j(\mathbf{p}) + (1-\beta_j) \sum_i \big[f(e_i, q_i, s_j) - c(e_i, q_i) - \kappa\, p_i(e_i, q_i) b \big] - k(s_j). \tag{3}$$

When β_j is high, private preferences dominate and bonus allocations may reflect bias or leniency. When β_j is low, the supervisor is more aligned with the principal and allocates bonuses to maximize welfare, raising both e and q and exerting greater scheduling effort s.

Thus, subjective schemes create both opportunities (better use of information on q) and risks (bias, weak managerial effort), depending on the extent of supervisory alignment.

Testable Implications

The framework clarifies the central trade-off: subjective schemes allow conditioning on hard-to-observe performance (q) but create scope for bias, whereas objective schemes ensure comparability but may overlook important qualitative dimensions of performance. The model yields the following predictions:

- **H1:** All incentive schemes increase observable effort e.
- **H2:** Observable effort e increases more under objective than under subjective schemes.
- **H3:** Hard-to-measure performance q improves under subjective schemes relative to no incentives.
- **H4:** Supervisory scheduling effort s increases under subjective schemes when supervisor alignment is high (i.e., low β).
- **H5:** Both e and q are higher under subjective schemes with stronger supervisor alignment.
- **H6:** Bonus allocations are less biased under subjective schemes when supervisory alignment is stronger.

The framework also motivates three additional questions. First, whether e (outreach) and

q (quality) function as technological complements or substitutes—that is, whether expanding outreach reinforces or crowds out engagement quality. Second, whether subjective and objective performance measures interact as complements or substitutes (Gibbs et al., 2004), determining whether combining them strengthens or weakens incentives. Third, whether incentives operate primarily on the extensive margin (raising the likelihood that workers undertake outreach) or the intensive margin (increasing effort among those already active).

4 Design and Randomization

The field experiment builds on the framework in Section 3. Within each district, tehsils are randomly assigned to treatment arms or control. In the three incentive arms, extension staff can qualify for a monthly bonus at four levels: high, low, base, or none. The department predetermines bonus amounts, with the base payment set at 7% of monthly salary. This amount is benchmarked against historical average monthly TA claims by field staff.[7] Low and high bonuses are set at an additional 7% and 15% of salary (see Appendix B).

4.1 Treatment and Control Arms

Control Arm: In tehsils assigned to the Control arm, no bonus payments are made. Staff continue to receive TADA allowances under the existing claim system. ADs in this arm have access to the AgriSmart dashboard to upload Farmer Training Programs (FTPs) and approve leave, but cannot view staff performance summaries, unlike their counterparts in the treatment arms.

Objective Arm: In tehsils assigned to the Objective arm, bonuses are based solely on verifiable performance information recorded in *AgriSmart*. To qualify for a base payment, staff must meet monthly attendance thresholds for days and hours worked. To qualify for a low or high bonus, staff must also meet additional thresholds related to village and farmer visits, distance traveled, time spent on extension activities, and the proportion of completed FTPs. These thresholds are set by the department's leadership (see Appendix Tables C.2

[7]The benchmark is based on FY15–16 levels initially but revised upwards in FY18–19.

and C.3). The indicators are similar for AOs and FAs but differ in scale due to their distinct geographic coverage. A third-party audit system, including callbacks to farmers, is used to verify the accuracy of *AgriSmart* data and discourage misreporting.

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

Subjective Arm: In tehsils assigned to the Subjective arm, ADs have access to the *AgriSmart* portal and can view staff performance indicators for their tehsil. They are required to use the staff *payment assignment page* on the *AgriSmart* dashboard for monthly bonus assignment, but they retain full discretion over bonus allocation.

The potential advantage of the Subjective arm lies in the AD's ability to consider performance dimensions not captured by *AgriSmart*. However, subjectivity may also reduce transparency and introduce bias.

Subjective Plus Arm: In tehsils assigned to the Subjective Plus arm, the bonus assignment process mirrors that of the Subjective arm. The key difference is that a monthly report on the performance of each AD's tehsil is provided to the AD's immediate supervisors at the district and divisional levels as well as to top provincial management, specifically the Director General of Extension and Secretary Agriculture.

The report includes *AgriSmart*-based summary statistics on staff performance and data-quality indicators from the farmer audit. Importantly, it also includes indicators of AD effort, specifically the AD's compliance with Standard Operating Procedures (SOPs) such as FTP scheduling (a key complementary input in frontline service delivery) and timely approval of staff leave applications. Figure 3 provides a snapshot of the report used in the Subjective Plus arm.

Under Subjective Plus, supervisors retain discretion, but their bonus allocations and managerial activities become observable to senior officials, increasing the expected cost of favoritism and low engagement.

Figure 3 about here

Figure 4 provides a summary of the treatment arms, including the supervisor's role in each arm. Figure 5 illustrates reported extension activity points over a three-month period in 2019 across all treatment arms and the control group.

Figure 4 and 5 about here

4.2 Randomization and Balance Checks

Table A.2 reports balance checks across FA/AO and tehsil characteristics. Five tests fail the equality-of-means test at the 5% level—consistent with chance. All joint F-tests have p-values above 0.110, except for the difference between the Objective and Subjective Plus arms ($p = 0.083$). For robustness, we report estimates with and without controls.

Over the 30-month intervention period, 396 workers (about 15%) exit due to transfers or retirement, reducing the endline sample to 2,199 frontline workers. Cross-tehsil transfers are rare; we track these cases and report intention-to-treat estimates throughout. Table A.3 shows that retained staff are balanced (all joint F-tests have $p > 0.115$), and Table A.4 confirms that attrition is not correlated with treatment status.

During the study period, four tehsils are merged into pre-existing tehsils due to departmental span-of-control decisions.[8] This reduces the sample to 127 tehsils; in all cases, merged tehsils share the same treatment status, limiting spillovers.

Linking Experimental Design to Conceptual Framework

The three incentive arms map directly to the model's theoretical dimensions. The Objective arm isolates incentives based on observable effort e (H1–H2), allowing us to assess whether rewards tied to measurable outputs raise outreach and potentially crowd out unobservable effort q. The Subjective arm introduces supervisor discretion, testing H3–H5 by allowing bonuses to reflect qualitative performance not captured by *AgriSmart*. The Subjective Plus

[8]This does not reflect a change in the province's administrative structure.

arm increases transparency and oversight, aligning supervisors' incentives with those of the principal and testing H4–H6 on how supervisory alignment (β_j) shapes staff effort and bias in bonus allocation. Together, these arms identify trade-offs between objective and subjective incentives and assess whether stronger supervisor alignment improves overall performance.

4.3 Timeline

The department launched the *AgriSmart* portal in March 2018, followed by the incentive program in April 2018. Extensive training sessions for field staff and Assistant Directors (ADs) were conducted at the tehsil level and organized by treatment arm.[9] The *AgriSmart* dashboard enabling online leave approvals, FTP scheduling, and bonus assignment was fully launched in August 2018.

5 Measurement and Data Sources

5.1 Data Sources

Our analysis combines administrative and survey data from five main 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. We exclude the peak COVID-19 period (April–July 2020) when extension activities were suspended.

***AgriSmart* Dashboard Data**. Data on supervisor (AD) effort are 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

[9] Training sessions were conducted sequentially by treatment arm across districts and covered use of the *AgriSmart* application and the incentive program. ADs received additional training specific to their treatment, including use of the management dashboard for leave approvals, FTP scheduling, and bonus assignments in the Subjective and Subjective Plus arms.

due to COVID-19 disruptions.

***AgriSmart* Farmer Audit Data**. To deter misreporting and verify the accuracy of *AgriSmart* entries, we conducted monthly phone audits with farmers identified as having participated in a group training or one-on-one meeting in the previous month. For each extension worker, one farmer was randomly selected per month over an 18-month period from March 2019 to December 2020, excluding April–July 2020 when extension activities were suspended due to COVID-19.[10]

Punjab-Wide Farmer Registration Database and Farmer Phone Survey. Measures of farmer access to extension services and agricultural outcomes come from a farmer phone survey. Importantly, the survey targeted farmers *not* self-reported by extension staff in the *AgriSmart* application, ensuring that outcomes are not biased by staff reporting behavior.

In 2019, the Punjab Agriculture Department created a province-wide farmer registration database to support departmental initiatives; by 2020, it included 1.8 million farmers.[11] Using this database, we sampled one farmer from the catchment area of each extension worker for a phone survey. Nine survey rounds were conducted between January 2020 and January 2021.[12]

Endline Survey of Field Staff and ADs. An endline phone survey was conducted in January 2021 and completed by 1,884 frontline staff and 116 ADs, corresponding to response rates of 86 and 91 percent, respectively.[13] We use these data to assess AO and FA satisfaction with the performance pay schemes and treatment effects on intrinsic motivation. For ADs, the survey provides measures of engagement with the *AgriSmart* dashboard. In addition, we collect information on caste (*zaat*) and lineage (*biradri*) of field staff and ADs, which we

[10]The resulting panel includes 30,542 observations, fewer than the expected 39,582 (2,199 staff $\times$ 18 months), due to staff leave or reassignment and to cases where activities were reported without farmer contact information.

[11]Online Appendix A, Table C shows that farmers in this database closely match Punjab farmers in the 2010 Agriculture Census.

[12]Data collection was paused from March to May 2020 due to COVID-19 and again in August 2020 for logistical reasons. The final data form a staff-month panel over nine rounds, with a completion rate of approximately 67% and responses from 13,312 farmers linked to 2,199 AO/FAs. Table A.14 shows that farmer characteristics and response rates are balanced.

[13]Non-response was primarily due to unanswered calls or staff being on leave.

use to construct measures of social connectedness between ADs and frontline staff.

5.2 Main Outcomes

Our primary outcomes correspond directly to the key dimensions of the conceptual framework: observable outreach effort (e), non-observable outreach effort (q), farmer outcomes (including access to extension and agricultural performance), and supervisor effort (s).

Observable effort. Observable outreach effort corresponds to e in our conceptual framework and is measured using administrative data from *AgriSmart*. Key indicators include the number of days worked, hours worked, villages visited, farmers visited, and small farmers visited by each extension worker. We also record the number of Farmer Training Programs (FTPs) conducted by staff for whom an FTP was scheduled by the AD. To capture overall outreach intensity, we construct an *outreach index* by normalizing each component and averaging them with equal weights.

Non-observable effort. Non-observable outreach effort corresponds to q in the framework and captures qualitative dimensions of performance that are less easily monitored. We focus on two sets of measures: (i) *quality of engagement*, proxied by meeting length; and (ii) *spread of individual extension activities*, which reflects effort on the extensive margin directly under the worker's control.

High-quality extension typically requires deeper engagement with farmers—understanding their specific problems and tailoring advice accordingly. While ideal measures would include aspects such as time allocation, communication quality, and relevance of advice, we use meeting length as a proxy for engagement quality. Although recorded in *AgriSmart*, meeting length was neither incentivized nor visible to supervisors on the dashboard.

The spread of extension activities refers to the geographic and client reach of an AO/FA's *individual* meetings with farmers. We measure it as the number of unique villages and farmers reached per month through one-on-one visits. Because group trainings (FTPs) are scheduled by ADs, variation in their geographic spread is more reflective of supervisory effort

and is therefore analyzed separately. To capture the potential trade-off between breadth and depth of engagement, we also measure *intensive-margin effort* as the frequency of repeat visits to the same farmers over the course of the program.

Farmer Outcomes. Farmer access to extension services and agricultural outcomes are measured via the farmer phone survey.

Our primary indicator of extension access is an *access index*, constructed as the equally weighted average of four standardized (z-score) binary measures: whether the farmer (i) knows their AO/FA, (ii) was informed about extension activities, (iii) attended a one-on-one meeting, and (iv) participated in a group training (FTP). These indicators are collected monthly and combined into a composite measure of access.

The main downstream agricultural outcome is wheat yields, one of Punjab's principal crops. Yield is based on farmer-reported production during the 2019–2020 *rabi* season (October 2019–March 2020), and is measured in kilograms per acre. This period provides farmers with at least one year of exposure to the treatment - enough to capture short-run effects , though not long enough for longer-run impacts to fully materialize.

In addition to average treatment effects, we examine whether incentive schemes affect the *targeting* of marginalized farmers. We classify farmers by their predicted likelihood of extension access using control-group data, where the strongest predictors are landholdings above 25 acres and at least grade-10 education. Using a logit model, we generate predicted access probabilities for all surveyed farmers and define those below the median as "marginalized." We then estimate treatment effects for this group to assess whether incentives expand outreach to farmers least likely to be served.

Supervisor effort. Supervisor effort (s) is measured using monthly indicators derived from the *AgriSmart* dashboard that capture both planning and administrative engagement. Specifically, we track: (i) the share of frontline staff assigned to conduct Farmer Training Programs (FTPs)—reflecting the supervisor's role in scheduling and ensuring the geographic coverage of group extension activities; and (ii) the share of staff leave requests acted upon

(approved or rejected). Both measures are constructed at the tehsil (supervisor) level and capture managerial effort in planning, oversight, and responsiveness.

6 Main Results

6.1 Empirical Strategy

This section describes the empirical strategy used to estimate treatment effects. We study impacts on three outcome categories: extension outreach effort (observable and non-observable), farmer outcomes (including access to extension and agricultural outcomes), and supervisor effort. Section 5 describes the data sources and 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 outreach indicator (e.g., days worked, hours worked, village visits, farmer visits, FTPs conducted) for extension worker i in tehsil t, district d, and month m. $\underline{y}_{itd}$ is the corresponding baseline measure included for ANCOVA adjustment. T^j is an indicator for treatment assignment at the tehsil level. ν_d and μ_{my} denote district fixed effects and year–month dummies. Standard errors are clustered at the tehsil level, the unit of randomization (Abadie et al., 2017).

To estimate treatment effects on farmer outcomes, we estimate:

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

where y^{farmer}_{itdm} is a farmer-reported outcome (e.g., access to extension, knowledge scores, wheat yields) for farmer i in tehsil t, district d, and month m. Standard errors are clustered

at the tehsil level.[14]

Finally, to assess 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} measures AD effort (e.g., the share of staff with scheduled FTPs or approved leave requests) in tehsil t, district d, and month m. Standard errors are clustered at the tehsil level.

In all specifications, β_j captures the treatment effect for arm j. For each outcome family, we report q-values—p-values adjusted for multiple hypothesis testing using the False Discovery Rate procedure of Anderson (2008).

6.2 Impact on Extension Effort

Observable Extension Outreach. We begin by examining the impact of treatment on AO and FA extension outreach effort, as captured by observable *AgriSmart* indicators. These measures correspond to the 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 are also free to observe staff performance through the lens of *AgriSmart* indicators if they choose to do so.

Table 1 presents the impact of treatment on outreach measures. Across all three treatment arms, we find positive and significant effects on attendance (number of days and hours worked) and outreach (village and farmer visits).

Attendance measures (Columns 1–2) show that, relative to the control group, AOs/FAs in the Objective arm 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 roughly 30 minutes longer per day, while in the Subjective Plus arm they work about three-

[14]Baseline data are unavailable for these outcomes; however, we test balance on fixed farmer characteristics and find no significant differences across treatment arms (Table A.10).

quarters of a day more and approximately one hour longer per day. Staff in the Objective and Subjective Plus arm work significantly more hours than the Subjective arm (p-value difference < 0.05 in both cases).

For other outreach indicators, such as village visits and individual farmer visits (Columns 3–4), 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 about 10 additional village visits and 19 more farmer visits per month, 13 of which are to small farmers. In contrast, AOs/FAs in the Subjective arm make roughly 3 more village visits and 6 additional farmer visits, while those in the Subjective Plus arm make about 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 values are closely aligned with the unadjusted p-values for all reported results.

Column 7 shows treatment effects on the overall *AgriSmart* outreach index. Relative to the control group, the Objective arm improves outreach by 0.49σ, the Subjective arm by 0.18σ, and the Subjective Plus arm by 0.25σ, showing a consistent pattern across individual indicators. Figure A.1 in Appendix A further demonstrates that treatment effects on the outreach index remain stable across the ten quarters of the intervention cycle.

Table 1 about here

The stronger effects observed in the Objective arm are consistent with the design of the incentive scheme, which explicitly rewards *AgriSmart*-based measures. At the same time, the positive effects in the subjective arms, especially the Subjective Plus arm, suggest that the visibility of *AgriSmart* indicators to ADs may motivate staff even in the absence of direct linkage to bonuses. It is particularly noteworthy that hours worked and FTPs completed are significantly higher in the Subjective Plus arm relative to the Subjective arm (p-value difference < 0.05 in both cases).

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

Non-observable Extension Outreach. As outlined in our conceptual framework, higher outreach in the Objective arm may come at the cost of non-observable dimensions of effort, such as lower visit quality or the geographic spread of outreach, as these measures are not directly incentivized. Furthermore, because 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 of Table 2 present treatment effects on extension quality, proxied by the length of meetings. While the number of meetings is incentivized in the Objective arm, meeting length is not. We find no evidence that higher outreach effort comes 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 eight additional minutes per meeting relative 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, a 20–23% increase in meeting duration.

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 focuses on villages visited specifically for conducting FTPs (which likely reflects supervisory engagement in FTP scheduling). Results highlight that field staff in the Subjective Plus arm visit about 1 additional village relative to the control group. The effect is also significantly different from the Objective and Subjective arm. This indicates

that additional village visits in the Objective arm are primarily for individual farmer visits, whereas in the Subjective Plus arm they involve both individual and 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 increase outreach to new farmers within the same villages (the intensive margin), rather than expanding to new villages (the extensive margin). Column 6 shows that only the Subjective and Subjective Plus arms significantly increase the number of farmers reached through group activities, indicating that staff in the Objective arm rely primarily on individual visits to expand their outreach.

Finally, Columns 7–8 report results for effort along the intensive margin. We find that the Objective arm induces greater 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 visit the same farmer about three more times than those in the control group, while staff in the Subjective and Subjective Plus arms conduct 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 three key insights. 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), this is mainly achieved through individual farmer visits in the Objective arm, whereas both individual and group visits increase in the Subjective Plus arm. Third, staff in the Objective arm demonstrate greater effort on the intensive margin.

Table 2 about here

Staff reporting behavior. An important concern is whether the outreach effects in Table 1 reflect reporting behavior. Reporting could affect the estimates through three channels: differential non-reporting across arms; greater reporting diligence in treatment arms due to

incentive links or supervisory review of *AgriSmart* data; and potential over-reporting near incentive thresholds in the Objective arm.

In practice, however, the scope for misreporting is limited. Most *AgriSmart* indicators rely on automated time-stamping and geo-tagging features that are difficult to manipulate. Moreover, the department implemented a third-party audit system (including farmer callbacks) to verify the accuracy of submitted reports and discourage misreporting. Nevertheless, we examine each of these concerns in turn.

Table A.7 shows that *AgriSmart* activity reporting is not correlated with treatment assignment. Monthly reporting rates are high across all arms: on average, only 11% of staff in the control group fail to report any activity during the 30-month intervention period—a high compliance rate for a cadre newly introduced to smartphone-based reporting.

Second, we assess whether the treatment effects in Table 1 reflect genuine effort or increased 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 (Table A.8), most coefficients are statistically insignificant. Even where significant, the differences are small and cannot account for the treatment effects observed.[15]

Third, we test whether staff in the Objective arm game the system by inflating reported activity near incentive thresholds. Examining the distribution of incentivized indicators, we find no evidence of bunching for most measures—including days and hours worked, village and farmer visits, distance traveled, and FTPs conducted (Figure A.2). The sole exception is modest bunching at the threshold for farmers per FTP, the only fully self-reported indicator; all others rely on automated or geo-tagged data.

Taken together, this evidence suggests no systematic indication of differential or inflated reporting across treatment arms. The observed treatment effects therefore likely reflect genuine differences in staff effort rather than reporting artifacts.

[15]For instance, Column 5 in Table A.8 shows that Subjective Plus staff report 4.3% more FTPs per month. Even if this fully reflects reporting bias, it would account for only 0.28 FTPs of the 6.7 FTPs reported in Table 1, leaving roughly 1.5 additional FTPs unexplained.

Staff Satisfaction. Finally, we examine how the incentive schemes affects staff motivation and satisfaction. While incentives can enhance performance by strengthening extrinsic motivation (Khan et al., 2016; Duflo et al., 2011), they may also have distortionary effects if they crowd out intrinsic or prosocial motivation (Benabou and Tirole, 2006; Giné, Mansuri, and Shrestha, 2022), or if they undermine self-efficacy and job satisfaction by being perceived as overly controlling. This discouragement could arise either from the rigidity of the Objective arm or from the uncertainty associated with supervisor discretion in the Subjective and Subjective Plus arms.

Table 3 presents treatment effects on three endline measures: (i) intention to leave the department (a proxy for job satisfaction), (ii) donations to an NGO in a dictator game (prosocial motivation), and (iii) self-efficacy.

Results show no significant treatment effects on intention to leave in the Objective and Subjective Plus arm. In contrast, we find that staff in the Subjective arm are 5 percentage points more likely to leave the service. We find no evidence that incentives crowd out prosocial motivation (Column 2) or reduce self-efficacy (Column 3).

Overall, these findings suggest that although the Objective arm is the most rigid incentive scheme, this rigidity does not reduce job satisfaction. This is consistent with evidence that inequality is better tolerated when perceived as merit-based (Rustichini and Vostroknutov, 2014; Deserranno, Kastrau, and León-Ciliotta, 2025b). Clear and predictable objective measures may enhance goal clarity and transparency, strengthening motivation rather than dampening it. By contrast, the higher intention to leave in the Subjective arm likely reflects frustration with opaque or unpredictable evaluation.

Endline survey responses reinforce this interpretation. Nearly 80% of staff in the *Objective* arm agree that the criteria are fair and realistic. When asked to choose a preferred bonus system, 55% favor an objective scheme with pre-defined targets and 21% prefer a hybrid scheme, while only 22% favor a fully subjective system. The preference (76%) for incentives with an objective component suggests that in large bureaucracies—where individual performance is difficult to observe—goal clarity and transparency are valued more than discretion.

6.3 Farmer Outcomes

In this section, we examine whether the improved extension outreach translates into better farmer outcomes such as farmer access to extension services and agricultural yields. We estimate average treatment effects for all farmers, as well as for marginalized farmers, to assess whether different incentive schemes affect the targeting of extension services.

All farmers. Table 4 presents estimated treatment effects on farmer outcomes. Column 1 shows that, relative to the control group, both 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.

Effective extension services are expected to improve farmer outreach as well as farm practices and productivity. However, an improvement in farm outcomes requires not just outreach but also high quality advisory content. This remained a critical gap in extension services in Punjab during the study period, despite the department's recognition of the need to update its advisory materials and discursive approach. Further, even under ideal conditions, changes in farm practices and productivity are not instantaneous. Farmers adapt over time with significant changes in farm outcomes evidenced only over multiple cropping seasons. Consistent with these constraints, Column 2 of Table4 shows positive but statistically insignificant effects on wheat yields in the Subjective and Subjective Plus arms.

Table 1 shows that while the Objective arm generates the largest gains in extension effort, improvements in farmer-reported access arise only under the Subjective arms. Consistent with Table 2, this reflects differences in effort allocation: Subjective arms reach more farmers through group activities such as FTPs, whereas the Objective arm relies more on individual and repeat visits. This pattern implies higher engagement intensity for a subset of farmers but broader coverage under Subjective incentives. Although yield effects are not statistically significant, their positive direction suggests that wider outreach does not come at the expense of productivity.

We next examine which farmers benefit from this broader outreach.

Table 4 about here

Marginalized Farmers. The analysis above shows that the Subjective Plus arm encourages greater effort along the extensive margin—expanding the number of unique villages and farmers visited (both overall and for FTPs), and farmer-reported access to extension. Whether such broader outreach is desirable remains an open question: it may extend services to typically excluded farmers but could reduce repeated interactions that support sustained adoption and yield gains. Given historically weak incentives to serve smaller, less-educated, and marginalized farmers, despite this being a core mandate, we next assess whether the treatment arms shift outreach toward these groups.

Table 5 reports results across three panels. Panel A presents heterogeneous effects by farmer type (marginalized vs. non-marginalized). Panel B restricts the sample to marginalized farmers, while Panel C focuses on non-marginalized farmers. We report treatment effects on the access index and yield outcomes in Columns 5–6, and the underlying access components in Columns 1–4 for completeness.

In Panel A, heterogeneous effects on the access index (Column 5) are insignificant across arms. However, Column 6 shows suggestive evidence that wheat yields of marginalized farmers in the Subjective Plus arm are higher than those of non-marginalized farmers (significant at the 10% level). We find no such heterogeneous yield effects in the other arms.

Panel B sharpens this pattern. Focusing on marginalized farmers, we find strong and consistent effects in the Subjective Plus arm: relative to the control group, marginalized farmers are 4.3 percentage points more likely to know their AO/FA, 3.9 percentage points more likely to be informed about government extension, 1.6 percentage points more likely to attend individual meetings, and report wheat yields that are 69 kg/acre higher (a 5.4% increase over the control mean).[16] Effects in the Subjective arm are generally smaller and statistically insignificant. Differences across treatment arms are not statistically significant, likely reflecting limited sample size.

By contrast, Panel C (non-marginalized subsample) reveals a different pattern. In the Subjective Plus arm, coefficients are smaller and insignificant, and the yield estimate is negative. In the Subjective arm, however, we see positive effects on several outcomes.

Table 5 about here

As a final check, we estimate quantile treatment effects (QTEs) of yields at $\tau = 0.20, 0.40, 0.60,$ and 0.80. QTEs allow us to assess whether treatment effects vary across the yield distribution, complementing heterogeneity analyses by farmer type. In particular, they indicate whether interventions disproportionately benefit lower-yield farmers, many of whom are marginalized. Figure A.3 shows that coefficients for the Objective and Subjective arms are negative across quantiles, suggesting limited effectiveness across the distribution. By contrast, the Subjective Plus arm exhibits consistently positive coefficients, with weakly significant effects at the lower end ($\tau = 0.20$).

Overall, these results suggest that the broader spread of effort in the Subjective Plus arm does not come at the expense of lower average yields (Table 4). Instead, there is indicative evidence that this arm improves both access to extension services and wheat yields specifically among marginalized farmers (Table 5). This pattern implies that expanding the reach of extension agents—particularly when supervisory incentives are well aligned—can raise aggregate productivity by enhancing outcomes for the most underserved farmers. Since

[16] Other information interventions report yield effects of 12–16% (e.g., Glennerster and Suri, 2015; Casaburi et al., 2014), but these provided specific, actionable advice on crop varieties. In our setting, extension agents lacked updated technical materials, making the observed yield gains encouraging.

these farmers have historically been excluded from public extension systems, they represent a significant yet underutilised source of potential productivity gains.

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 that enhance frontline performance. In this context, ADs' contributions include key managerial tasks such as planning and scheduling Farmer Training Programs (FTPs), assigning FTPs to frontline staff, and performing administrative duties such as timely leave approvals.

Supervisor incentives differ across treatment arms. Although ADs are not directly incentivized in the Objective and Subjective arms, they may still respond to increased staff effort or to pressure from their subordinates to complete managerial tasks. For instance, in the Objective arm, ADs' timely approval of staff leave requests is necessary for FAs/AOs to qualify for bonuses. By contrast, in the Subjective Plus arm, ADs receive monthly reports on their tehsil's performance, including indicators of their own effort. These reports may strengthen their alignment with the principal's objectives by increasing oversight and accountability.

We assess the impact of the three treatment arms on two aspects of supervisory effort: (1) the proportion of AOs/FAs assigned an FTP each month, and (2) the proportion of staff leave requests acted upon (approved or rejected) in a timely manner.

Table 6 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), and these effects are significantly higher than those in the Objective and Subjective arms at the 10% level. This increase is driven primarily by a higher assignment of FTPs to FAs. Column 1 indicates that 7.4 percentage points more FTPs are 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. Effects in the Objective and Subjective arms are insignificant throughout.

Column 4 shows a significant increase in leave approvals in the Objective arm, which supports the interpretation that staff may exert pressure on ADs to approve leaves, which impacts their qualification for a bonus, while the scheduling of more FTPs does not.

Table 6 about here

These findings help explain the higher FTP completion rates in the Subjective Plus arm (Table 1). Monitoring through tehsil-level performance reports leads ADs not only to schedule more FTPs but also to spread them across more villages, consistent with the expansion along the extensive margin in Table 2. Overall, the results underscore the role of complementary managerial inputs in enhancing frontline performance and the potential gains from incentivizing supervisors.

7 Mechanisms

Our conceptual framework in Section 3 predicts that better alignment of supervisors' incentives with the principals' objectives — such as through top-down monitoring — induces greater effort on complementary inputs and more merit-based and equitable bonus allocation. The previous section provides supporting evidence: under the Subjective Plus arm, tehsil reports increase supervisors' managerial effort, notably through greater FTP assignment. Here, we examine the mechanisms underlying this result by addressing three questions: (i) do tehsil reports align AD incentives with the principal's objectives; (ii) do supervisors acquire additional information for bonus decisions; and (iii) do they use this information when allocating bonuses?

Legitimacy of the tehsil reports. The Subjective Plus arm introduces top-down monitoring via tehsil-level performance reports. We first assess whether ADs perceive these reports as legitimate and motivating.

Descriptive evidence from the endline AD survey provides strong support. Approximately 89% of ADs in the Subjective Plus arm report that the tehsil reports are either *very important* (77%) or *somewhat important* (11%) to senior leadership. Moreover, about half of ADs

indicate that their superiors have taken follow-up 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 cite either opportunities for promotion (21%) or personal motivation and satisfaction (79%). Taken together, this evidence suggests that the tehsil reports are perceived as legitimate within the department and are successful, at least partially, in aligning AD incentives with the principal's objectives.

Acquisition of information by supervisors. If alignment improves as intended, we would expect supervisors in the Subjective arms to acquire additional information about staff performance—either through increased field monitoring or more frequent use of the *AgriSmart* dashboard.

Table 7 presents treatment effects on two such dimensions: ADs' field visits (Column 1) and dashboard usage (Columns 2–3). We find no evidence that ADs in the Subjective Plus arm conduct more field visits relative to other arms (Column 1). However, as Column 3 shows, ADs in the Subjective Plus arm are more likely than ADs in the Subjective arm to view staff performance summaries on the dashboard.

Thus, while tehsil reports do not increase field-level supervision, they appear to encourage greater reliance on digital monitoring data in staff assessments.

Table 7 about here

Supervisor bonus assignment. Finally, we examine bonus assignment across treatment arms, in particular in the subjective schemes. Supervisors may rely on *AgriSmart* metrics or on less observable attributes such as staff ability, service quality, or personal connections; top-down monitoring in the Subjective Plus arm may mitigate resulting bias.

Figure 6 shows that average monthly bonus payments are higher in the Subjective and Subjective Plus arms (PKR 4,845 and PKR 4,994, respectively) than in 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).[17]

Figure 6 about here

We then examine which staff characteristics predict bonus allocation. These characteristics fall into four categories: (1) *AgriSmart* performance metrics (e.g., work attendance, outreach, FTPs completion), (2) worker ability (education, experience, knowledge scores), (3) social connections with the AD (e.g., shared *zaat* (caste)/*biradari* (lineage) ties), and (4) personality traits (PSM and Big Five indices). The first category is observable through *AgriSmart*, while the latter three are likely known to supervisors. For comparability, all predictors are standardized as z-scores. We estimate the following specification:

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

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.[18]

Table 8 presents the results. Panel A reports estimates for the full sample, and Panel B focuses on staff assigned FTPs.

Table 8 about here

As expected, bonus allocation in the Objective arm (Column 1, Panel A) is driven primarily by *AgriSmart* metrics, especially days worked, hours worked, and farmer visits. Columns 4–5 confirm that these metrics are significantly more predictive in the Objective arm than in the Subjective or Subjective Plus arms.

[17] We rule out the possibility that these differences reflect differential payment rates across arms. According to our endline survey of field staff, workers in all arms report receiving bonuses in 3–4 out of 6 months from January to June 2020. Table A.11 confirms that the number of bonuses paid is 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 do not differ significantly across treatments.

[18] 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.

In the Subjective and Subjective Plus arms (Columns 2–3), *AgriSmart* metrics remain relevant, but additional factors influence bonus allocation. In the Subjective arm (Column 2), bonuses are more strongly associated with years of schooling, experience, and—importantly—shared *zaat* with the AD, suggesting favoritism or social bias. By contrast, in the Subjective Plus arm (Column 3), supervisors place greater weight on *AgriSmart* performance, staff knowledge, and personality traits, while relying significantly less on social ties. This pattern indicates that top-down monitoring reduces 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.

Taken together, these findings yield two lessons for incentive design. First, granting supervisors unmonitored discretion risks rewarding social ties over effort or performance. 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 examines how incentive design in hierarchical bureaucracies can be improved by aligning supervisors' behavior with principals' objectives. Using a randomized controlled trial in the Punjab Agriculture Extension Department, we compare three pay-for-performance schemes that vary in how frontline workers and supervisors are monitored and rewarded.

We show that light-touch oversight of supervisors substantially amplifies the effectiveness of frontline performance pay. The *Subjective Plus* scheme—introducing minimal top-down monitoring without changing supervisors' financial incentives—generates large gains in outreach, engagement, and farmer access, matching or exceeding the effects of the *Objective* scheme. These gains arise because aligned supervisors supply complementary managerial inputs, including greater FTP scheduling, broader geographic coverage, and more merit-based bonus allocation.

Oversight in *Subjective Plus* is intentionally modest: monthly performance summaries are shared with senior management but not used for formal appraisal. Even so, this intervention realigns middle-management incentives, reduces favoritism, and strengthens the link between effort and reward.

All schemes remain within the department's existing TADA budget. While *Subjective Plus* entails slightly higher average payments, these reallocate resources toward frontline workers while delivering improvements in service quality and equity, implying efficiency gains rather than cost escalation.

More broadly, the findings highlight a central insight for organizational economics: in multi-tiered bureaucracies, performance pay is most effective when paired with mechanisms that align managerial inputs higher in the hierarchy. Even light-touch monitoring of supervisors can discipline discretion and leverage local information without undermining autonomy, offering a scalable strategy for improving performance in low- and middle-income settings.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" NBER Working Papers 24003, National Bureau of Economic Research, Inc.

Anderson, Jock R. and Gershon Feder. 2007. "Agricultural Extension." *World Bank Research Observer* 22 (2):199–236.

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.

Andrabi, Tahir and Christina Brown. 2022. "Subjective versus Objective Incentives and Teacher Productivity." Tech. Rep. 22/092, RISE Working Paper Series.

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

Basinga, Paulin, Paul Gertler, Agnes Binagwaho, Agnes Soucat, Jennifer Sturdy, and Christel Vermeersch. 2011. "Effect on Maternal and Child Health Services in Rwanda of Performance-Based Financing: A Randomised Controlled Trial." *The Lancet* 377 (9775):1421–1428.

Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak. 2021. "Can Network Theory-Based Targeting Increase Technology Adoption?" *American Economic Review* 111 (6):1918–43.

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.

Benabou, Roland and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *The American Economic Review* 96 (5):1652–1678.

BenYishay, Ariel and A Mushfiq Mobarak. 2019. "Social Learning and Incentives for Experimentation and Communication." *The Review of Economic Studies* 86 (3):976–1009.

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu. 2019. "The Glittering Prizes: Career Incentives and Bureaucrat Performance." *The Review of Economic Studies* 87 (2).

Best, Michael, Jonas Hjort, Francesco Molinari, and Gautam Rao. 2022. "Information, Incentives, and Targets." *Quarterly Journal of Economics* 137 (1):1–64.

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.

Casaburi, Lorenzo, Michael Kremer, Sendhil Mullainathan, and Rachna Ramrattan. 2014. "Harnessing ICT to Improve Agricultural Extension: Evidence from Kenya." Working paper, Abdul Latif Jameel Poverty Action Lab (J-PAL). J-PAL Agricultural Technology Adoption Initiative (ATAI) report.

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, Philipp Kastrau, and Gianmarco León-Ciliotta. 2025a. "Promotions and Productivity: The Role of Meritocracy and Pay Progression in the Public Sector." *The American Economic Review. Insights* 7 (1):71–89.

———. 2025b. "Promotions and Productivity: The Role of Meritocracy and Pay Progression in the Public Sector." *The American Economic Review. Insights* 7 (1):71–89.

Duflo, Esther et al. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *The American Economic Review* 101 (6):2350–2390.

———. 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* .

Giné, Xavier, Ghazala Mansuri, and Slesh A. Shrestha. 2022. "Mission and the Bottom Line: Performance Incentives in a Multigoal Organization." *The review of economics and statistics* 104 (4):748–763.

Glennerster, Rachel and Tavneet Suri. 2015. "The Economics of Microfinance and Agricultural Extension: Evidence from Kenya." *American Economic Journal: Applied Economics* 7 (3):1–32. Field experiment on agricultural information dissemination and technology adoption in Kenya.

Glewwe, Paul et al. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (3):205–227.

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., Asim I. Khwaja, and Benjamin A. Olken. 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors." 131:219?271.

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.

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

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

Prendergast, Canice. 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.

Rasul, Imran, Daniel Rogger, and Martin J Williams. 2021. "Management, Organizational Performance, and Task Clarity: Evidence from Ghana's Civil Service." *Journal of Public Administration Research and Theory* 31 (2):259–277.

Rustichini, Aldo and Alexander Vostroknutov. 2014. "Merit and Justice: An Experimental Analysis of Attitude to Inequality." *Plos One* 9 (12):1–19.

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

==========

Figures and Tables

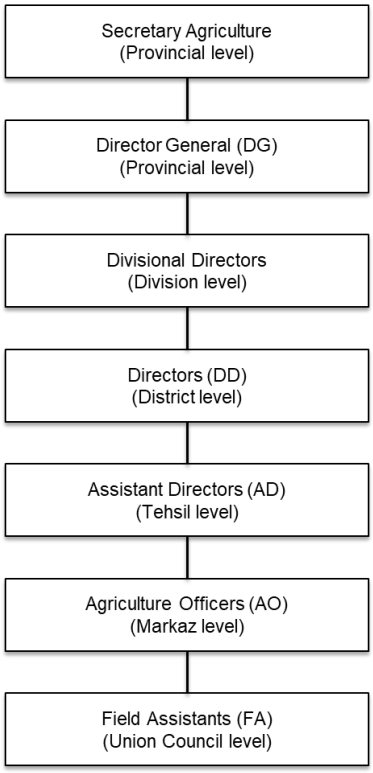


Figure 1: Agricultural Department Hierarchy from an Extension Perspective

Figure 2: Agrismart Application Interface

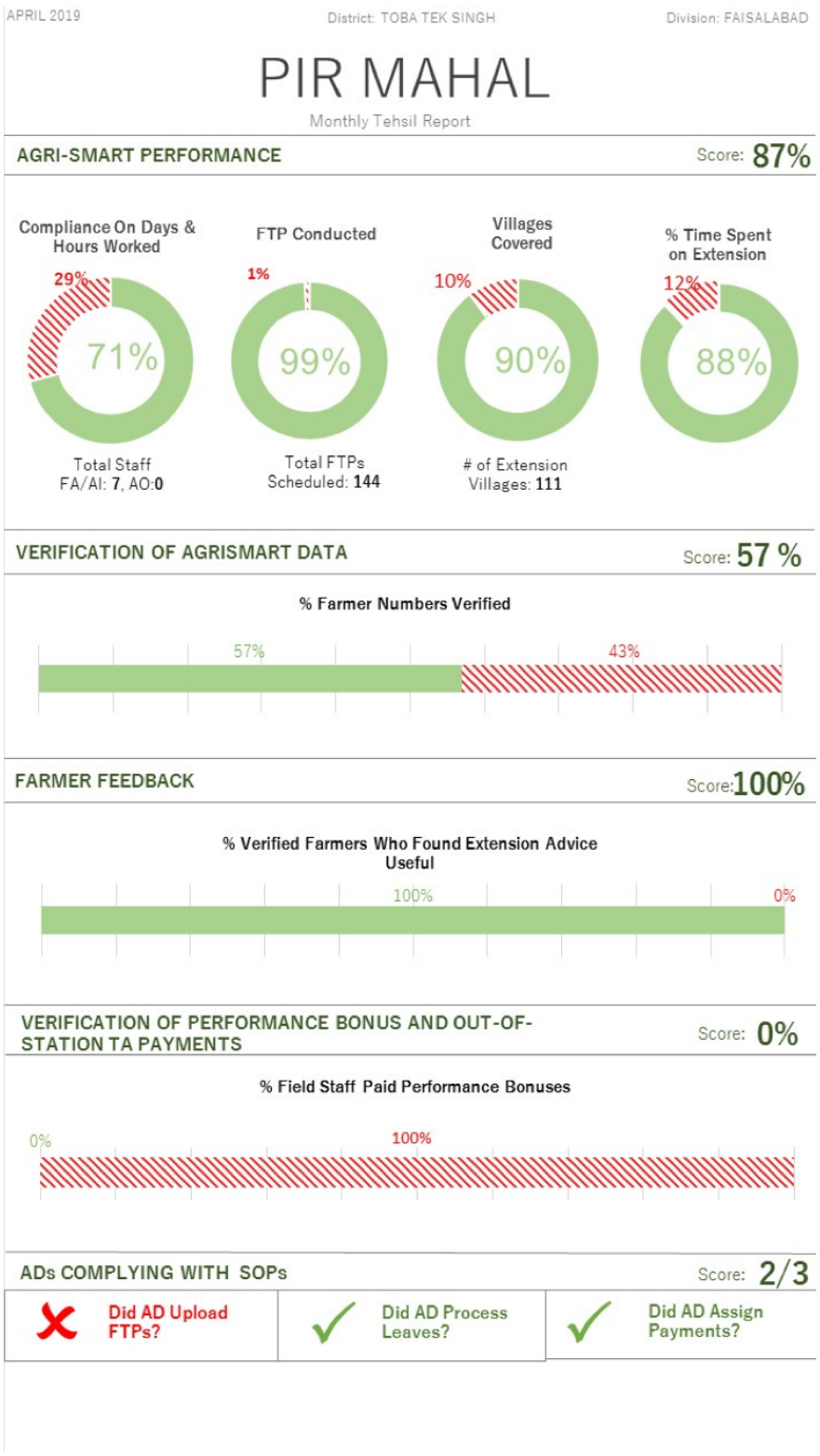


Figure 3: Tehsil report

<table>
<tr><td colspan="2" rowspan="3">Supervisors (ADs)
[Tehsil Level]</td><td colspan="3">Frontline Extension Staff (AOs and FAs)</td></tr>
<tr><td colspan="3">AgriSmart Based Data Entry</td></tr>
<tr><td>No Bonus
(TA payment possible)</td><td colspan="2">Monthly Bonus Payment</td></tr>
<tr><td>No Access to AgriSmart</td><td>Can assign TA Funds to Staff</td><td>Control Tehsils (C)</td><td colspan="2" rowspan="3"></td></tr>
<tr><td rowspan="3">Access to AgriSmart</td><td>No Role in Staff Bonus Assignment</td><td rowspan="3"></td><td>Objective Tehsils (T1)</td></tr>
<tr><td>Discretion in Staff Bonus Assignment</td><td>Subjective Tehsils (T2)</td></tr>
<tr><td>Discretion in Staff Bonus Assignment with Top-down Monitoring</td><td>Subjective Plus Tehsils (T3)</td></tr>
</table>

Figure 4: Agrismart Incentive Programme

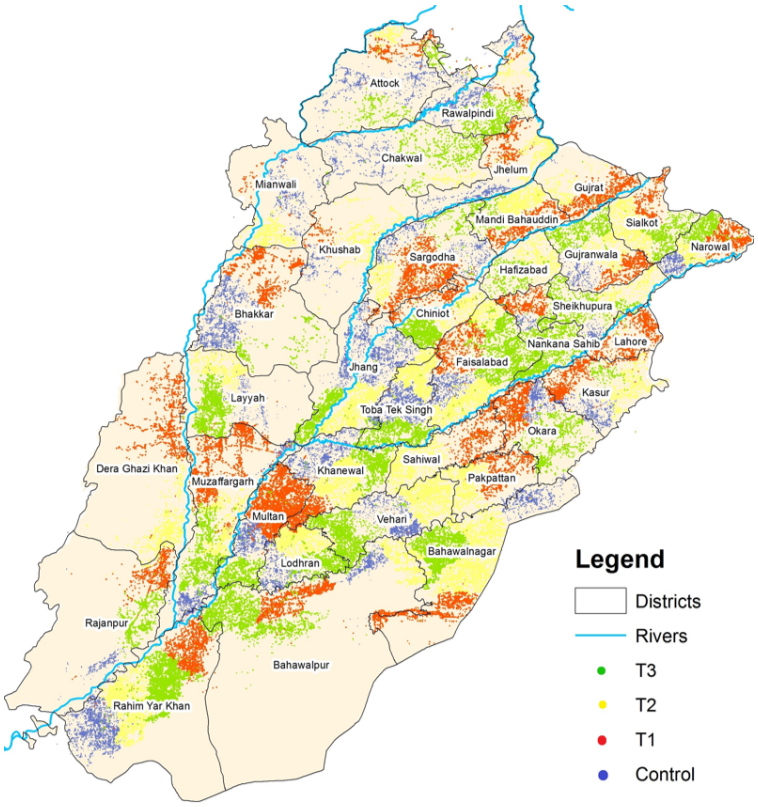
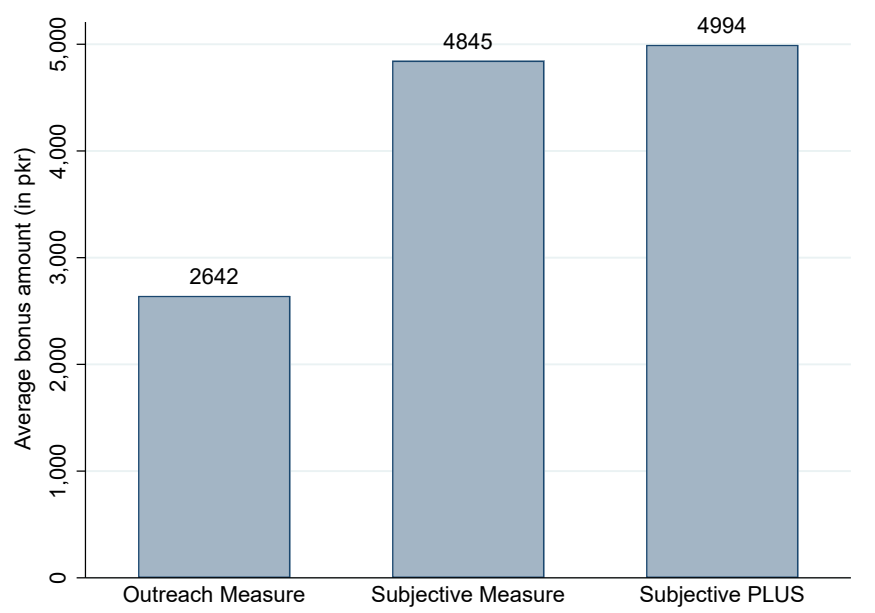
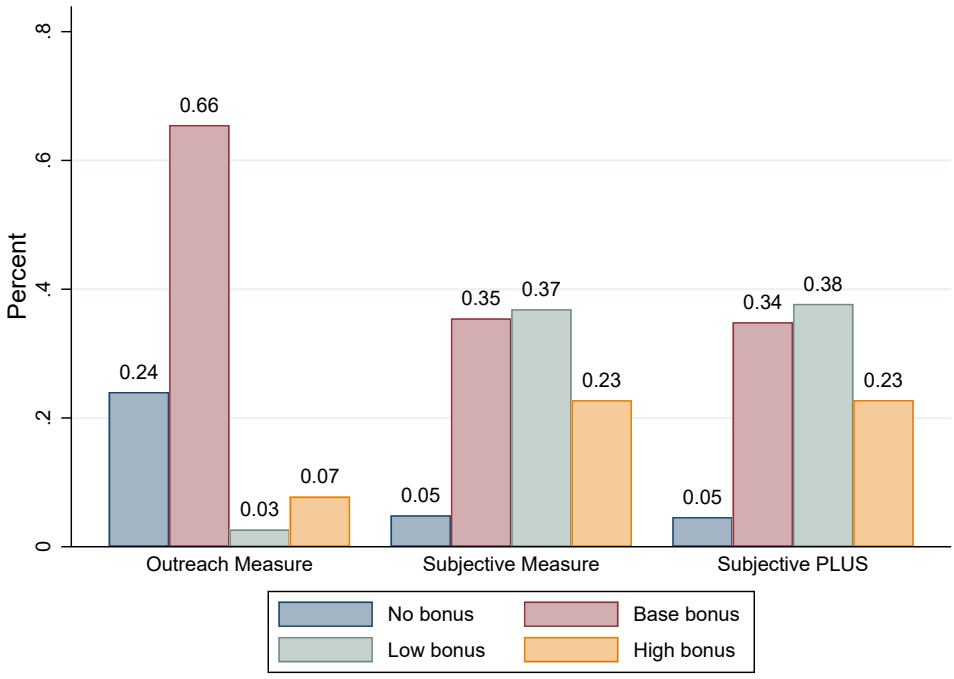


Figure 5: Agrismart Data Points



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



(b) Distribution of bonus levels across treatment.

Figure 6: Bonus assignment outcomes across treatment arms.

Table 1: Impact on Observable Extension Outreach

	(1) Days Worked	(2) Hours Worked	(3) Village Visits	(4) Farmer Visits	(5) Small Farmer Visits	(6) No. FTPs completed	(7) Outreach Index
Objective	1.070***	0.928***	10.071***	19.048***	13.337***	0.851	0.489***
	(0.275)	(0.238)	(1.866)	(3.223)	(2.273)	(0.572)	(0.072)
	[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	0.179***
	(0.269)	(0.198)	(1.100)	(2.636)	(1.704)	(0.525)	(0.057)
	[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***	0.249***
Plus	(0.286)	(0.218)	(1.125)	(2.359)	(1.469)	(0.555)	(0.052)
	[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	-0.003
P-value difference:							
Obj = Sub	0.108	0.016**	0.001***	0.001***	0.000***	0.763	0.000***
Obj = Subplus	0.261	0.874	0.004***	0.003***	0.000***	0.122	0.004***
Sub = Subplus	0.730	0.010**	0.602	0.667	0.952	0.036**	0.269
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	64,096	64,096	64,096	64,096	64,096	24,807	64,096

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 AOs/FAs 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 Non-observable Extension Effort

	Extension Quality		Extensive Margin (Spread)				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 AOs/FAs 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: Staff Satisfaction

	(1) Intention to Leave	(2) Dicatator Game Donation (out of PKR 100)	(3) Self Efficacy Index
Objective	0.017	4.097**	0.237***
	(0.017)	(1.837)	(0.063)
Subjective	0.051***	1.993	0.153**
	(0.018)	(1.857)	(0.064)
Subjective Plus	0.008	3.066	0.156**
	(0.015)	(2.113)	(0.067)
Control Mean	0.087	15.635	-0.114
P-value difference:			
Obj = Sub	0.100*	0.265	0.211
Obj = Subplus	0.617	0.620	0.252
Sub = Subplus	0.017**	0.597	0.966
District FE	Yes	Yes	Yes
N	2065	2065	2064

Notes: The sample is based on AOs and FAs at endline. Out of the total sample of 2199 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.

Table 4: Impact on Farmer Outcomes

	(1) Access Index	(2) Wheat Yield
Objective	0.026	-16.919
	(0.017)	(28.000)
Subjective	0.067***	18.922
	(0.023)	(29.081)
Subjective Plus	0.060***	27.417
	(0.023)	(29.642)
Control Mean	-0.069	1263.343
P-value difference:		
Obj = Sub	0.060*	0.228
Obj = Subplus	0.178	0.138
Sub = Subplus	0.817	0.793
District FE	Yes	Yes
Year-month dummies	Yes	Yes
N	13,228	8071

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 AO/FA 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): 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 5: Impact on Marginalized Farmers

	(1) Know staff	(2) Informed abt. extension	(3) Had grp. meeting	(4) Had ind. meeting	(5) Access Index	(6) Wheat Yield (kgs/acre)
Panel A: Heterogeneous treatment effects by marginalized farmer						
Marginalized	-0.135***	-0.040**	-0.040***	-0.039***	-0.152***	-116.774***
	(0.021)	(0.019)	(0.014)	(0.009)	(0.026)	(26.708)
Objective	-0.006	0.001	0.022	0.007	0.018	-51.544
	(0.027)	(0.015)	(0.013)	(0.009)	(0.027)	(31.336)
Subjective	0.026	0.034*	0.032*	0.032**	0.082**	-8.751
	(0.030)	(0.018)	(0.017)	(0.013)	(0.037)	(30.500)
Subjective Plus	0.008	0.030	0.027	0.019	0.055	-12.534
	(0.033)	(0.021)	(0.020)	(0.012)	(0.041)	(35.790)
Objective x Marginalized	0.015	0.017	0.002	-0.000	0.017	62.587
	(0.028)	(0.023)	(0.020)	(0.012)	(0.036)	(39.813)
Subjective x Marginalized	0.001	-0.002	-0.031	-0.020	-0.040	45.308
	(0.031)	(0.029)	(0.021)	(0.016)	(0.044)	(36.377)
Subjective Plus x Marginalized	0.033	0.011	-0.007	-0.006	0.011	72.541*
	(0.030)	(0.026)	(0.022)	(0.013)	(0.042)	(37.513)
N	13,228	13,228	13,228	13,228	13,228	8,071
Panel B: Marginalized Farmer (subsample)						
Objective	0.009	0.020	0.022	0.007	0.037	21.769
	(0.018)	(0.017)	(0.014)	(0.008)	(0.022)	(38.084)
Subjective	0.031	0.034	0.003	0.015**	0.049*	36.826
	(0.022)	(0.022)	(0.011)	(0.007)	(0.027)	(35.631)
Subjective Plus	0.043**	0.039**	0.015	0.016**	0.066***	68.820**
	(0.020)	(0.017)	(0.012)	(0.006)	(0.023)	(33.554)
Control Mean	0.409	0.581	0.150	0.058	-0.069	1263.343
P-value difference:						
Obj = Sub	0.233	0.445	0.111	0.392	0.614	0.728
Obj = Subplus	0.057	0.237	0.614	0.235	0.188	0.224
Sub = Subplus	0.563	0.789	0.284	0.929	0.496	0.391
N	7,064	7,064	7,064	7,064	7,064	4,017
Panel C: Non-marginalised Farmer (subsample)						
Objective	-0.008	0.000	0.024*	0.006	0.017	-52.846*
	(0.027)	(0.017)	(0.013)	(0.009)	(0.029)	(30.385)
Subjective	0.021	0.034**	0.033*	0.029**	0.077**	-14.399
	(0.030)	(0.017)	(0.017)	(0.012)	(0.038)	(31.513)
Subjective Plus	0.006	0.037*	0.034	0.015	0.058	-17.426
	(0.033)	(0.020)	(0.021)	(0.012)	(0.041)	(37.006)
Control Mean	0.409	0.581	0.150	0.058	-0.069	1263.343
P-value difference:						
Obj = Sub	0.222	0.039	0.542	0.036	0.055	0.156
Obj = Subplus	0.629	0.083	0.649	0.443	0.311	0.276
Sub = Subplus	0.609	0.887	0.988	0.361	0.675	0.935
N	6,164	6,164	6,164	6,164	6,164	4,054
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-month dummies	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors clustered at the tehsil level (which is the unit of randomization) and presented in parenthesis. All regressions include district, year, and month FE. [a] The dependent variables required respondents to recall and report their activities in the past month. Estimates are significant at the *10%, **5%, and ***1% level

Table 6: 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). 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 7: 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.

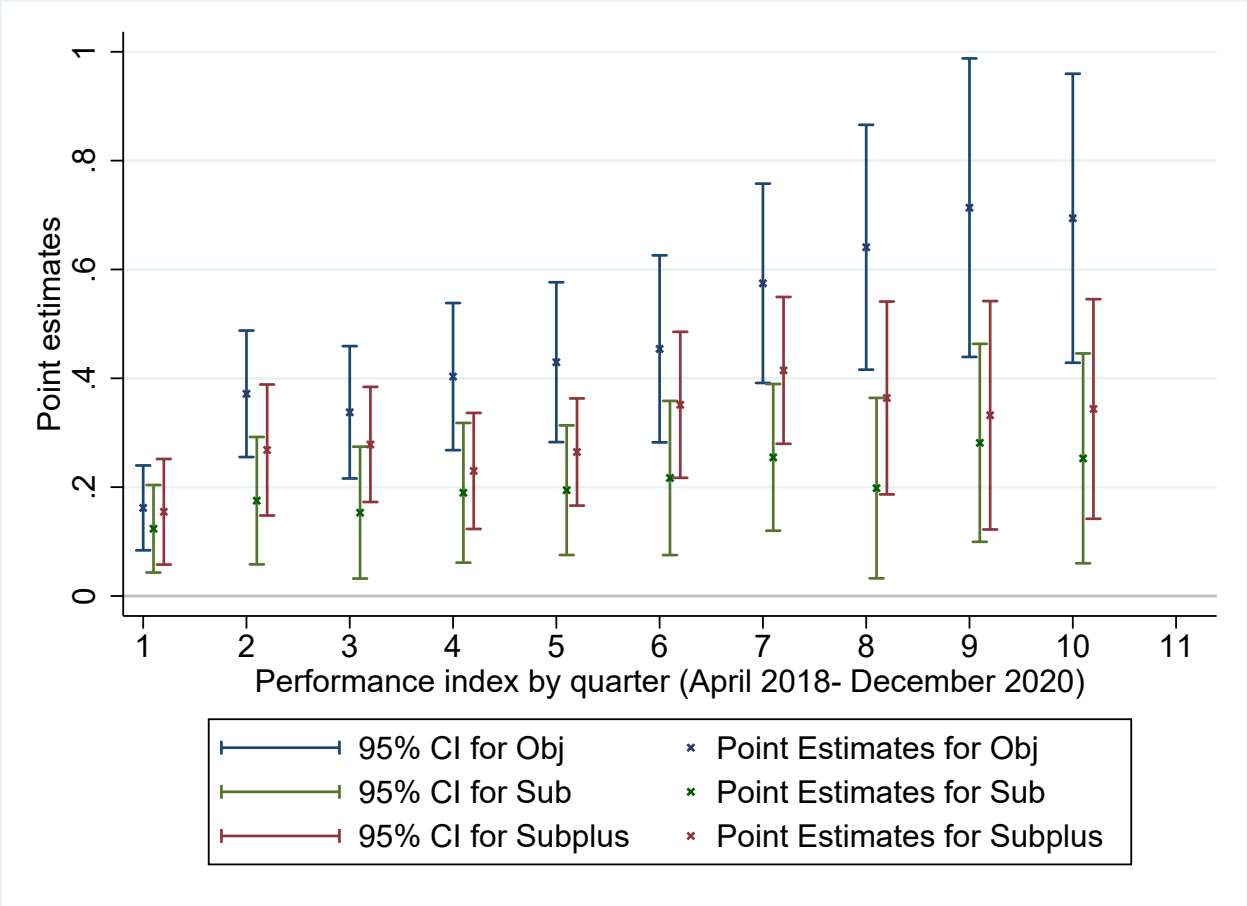
Table 8: Correlates of Bonus Assignment

		(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 FTPs completed	121.817***	126.958***	247.434***	0.915	0.030**	0.039**
		(34.732)	(34.446)	(45.314)			
	Observations	4,448	2,631	2,899			
	Tehsil FE	Yes	Yes	Yes			
	Year-Month FE	Yes	Yes	Yes			

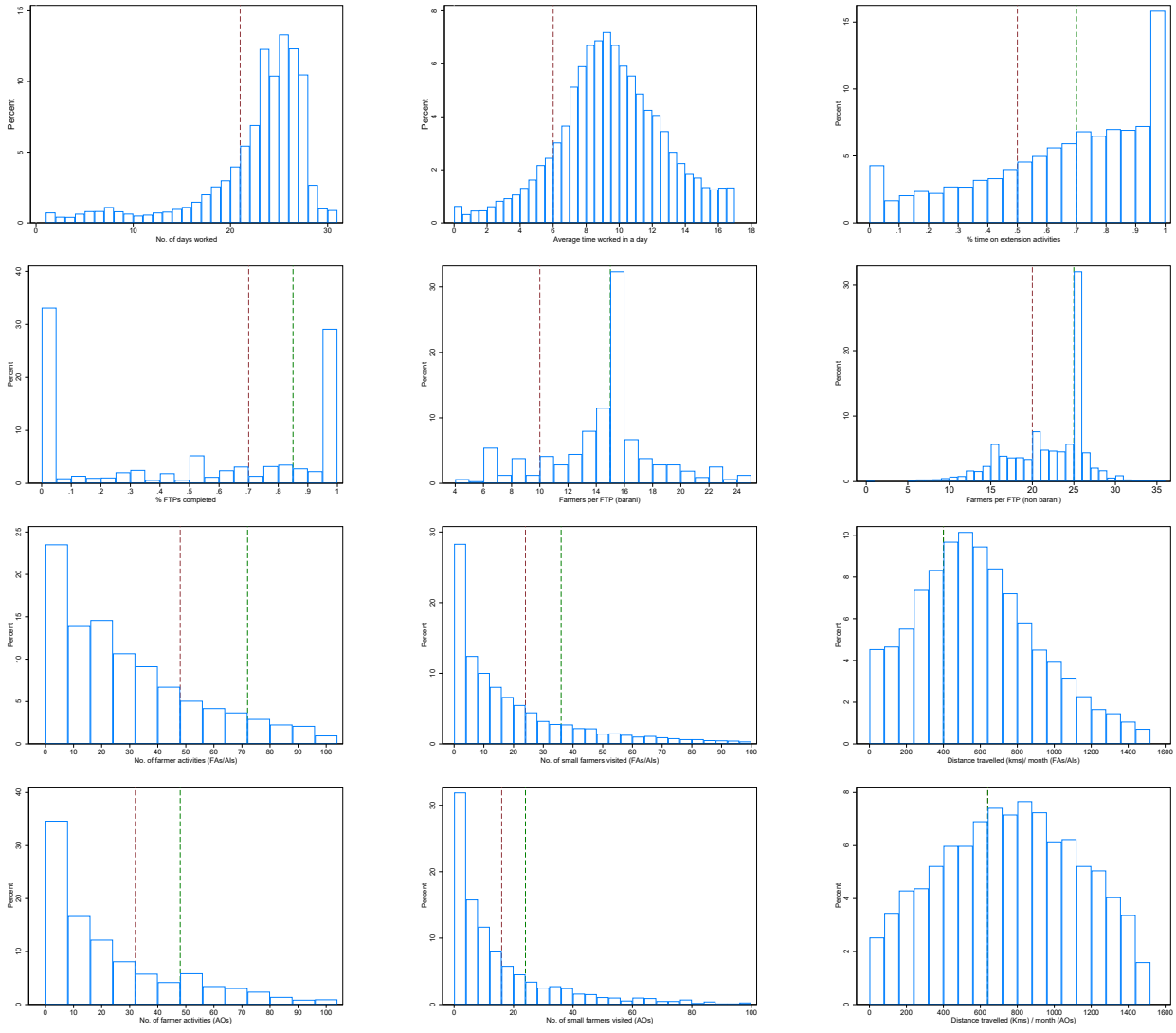
Notes: The dependent variable is the bonus amount assigned to staff in a given month. Independent variables are normalized z-scores. Standard errors are clustered at the tehsil level. The dataset is a staff-month panel from August 2018 to January 2021, excluding April–July 2020. It is a subsample including FAs in treatment groups for whom bonuses were assigned. Estimates are significant at the *10%, **5%, and ***1% level.

Appendix

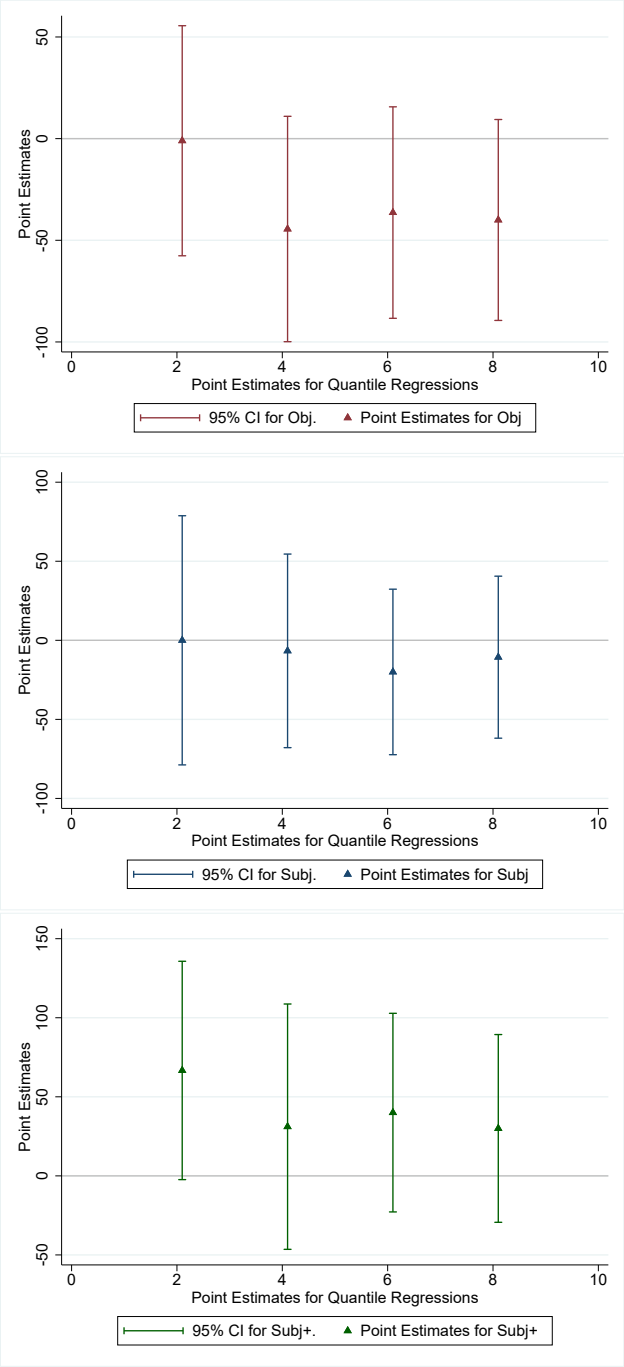
Appendix A: Figures and Tables



A. 1: Treatment effects on performance index (by quarter)



A. 2: Activities against thresholds in Objective Arm



A. 3: Quantile Treatment Effects

A. 1: Descriptive Statistics of Staff 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
TA Claim and Payment					
Avg Claim (yearly)	26,575	40,205	0	10,500	31,485
Avg Payment (yearly)	18,290	31,693	0	6,300	18,000
Observations	2,595				

Notes: All variables are based on monthly averages unless otherwise indicated. The performance indicators are based on *AgriSmart* data from March 2018 before the intervention roll-out in April 2018. 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	41.958	43.666	43.802	42.028	1.708*	1.844*	0.070	-0.136	1.639*	1.774**
	(11.809)	(11.162)	(11.226)	(11.548)	(0.081)	(0.055)	(0.948)	(0.846)	(0.056)	(0.034)
Salary	38,598	41,073	39,181	39,7912	2,475	583	1,194	1,892	1,281	-611
	(31,025)	(30,189)	(21,901)	(28,637)	(0.216)	(0.755)	(0.581)	(0.201)	(0.480)	(0.712)
Years of experience	17.366	19.239	19.864	18.100	1.873*	2.498***	0.735	-0.625	1.139	1.763**
	(12.677)	(11.991)	(12.208)	(12.233)	(0.050)	(0.009)	(0.477)	(0.427)	(0.209)	(0.049)
Days worked/month	14.170	12.821	12.908	13.537	-1.349	-1.262	-0.634	-0.087	-0.716	-0.628
	(6.737)	(6.753)	(6.968)	(6.931)	(0.224)	(0.254)	(0.600)	(0.941)	(0.582)	(0.629)
Hours worked/month	7.467	7.586	7.269	7.923	0.119	-0.198	0.455	0.317	-0.336	-0.653
	(4.672)	(4.607)	(4.690)	(4.596)	(0.838)	(0.753)	(0.481)	(0.620)	(0.613)	(0.360)
Village visits/month	12.969	13.054	13.104	14.279	0.085	0.135	1.310	-0.049	-1.224	-1.175
	(11.643)	(11.283)	(11.690)	(13.086)	(0.956)	(0.938)	(0.496)	(0.977)	(0.524)	(0.570)
Farmer visits/month	17.515	16.997	16.683	19.901	-0.518	-0.832	2.387	0.314	-2.904	-3.218
	(22.030)	(19.695)	(19.426)	(30.322)	(0.834)	(0.756)	(0.462)	(0.901)	(0.375)	(0.346)
Small farmer visits/month	9.434	10.023	8.367	9.858	0.589	-1.067	0.424	1.656	0.165	-1.491
	(12.981)	(12.333)	(11.294)	(14.168)	(0.700)	(0.489)	(0.802)	(0.261)	(0.921)	(0.371)
Prop. time on extension/month	0.563	0.571	0.527	0.536	0.008	-0.036	-0.027	0.044	0.035	-0.009
	(0.300)	(0.291)	(0.317)	(0.302)	(0.784)	(0.437)	(0.405)	(0.312)	(0.214)	(0.841)
Distance travelled (kms)/month	375.83	344.60	347.88	409.50	-31.230	-27.952	33.668	-3.278	-64.898	-61.620
	(1,312)	(1,298)	(880)	(1,161)	(0.736)	(0.765)	(0.773)	(0.968)	(0.545)	(0.567)
Avg. TA claim (yearly)	21,859	30,2158	28,178	25,134	8,357	6,320	3,280	2,037	5,076	3,039
	(41,172)	(40,071)	(40,129)	(39,271)	(0.307)	(0.572)	(0.696)	(0.855)	(0.545)	(0.794)
Avg. TA payment (yearly)	15,824	21,317	19,401	16,074	5,492	3,577	250.07	1,916	5,242	3,327
	(34,088)	(31,525)	(34,403)	(26,459)	(0.425)	(0.717)	(0.969)	(0.846)	(0.409)	(0.733)
Joint F-Test					0.377	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	0.007	0.030	0.043	-0.022	-0.036	-0.013
	(0.130)	(0.130)	(0.112)	(0.091)	(0.830)	(0.343)	(0.145)	(0.458)	(0.208)	(0.602)
Prop. of villages with metal road	0.787	0.842	0.841	0.811	0.055*	0.054*	0.024	0.000	0.031	0.031
	(0.137)	(0.114)	(0.109)	(0.101)	(0.097)	(0.091)	(0.445)	(0.994)	(0.256)	(0.245)
Prop. of villages with canal	0.616	0.645	0.669	0.680	0.029	0.053	0.064	-0.024	-0.035	-0.011
	(0.349)	(0.390)	(0.339)	(0.337)	(0.757)	(0.547)	(0.467)	(0.796)	(0.702)	(0.892)
Prop. of villages with housing	0.023	0.024	0.025	0.022	0.001	0.002	-0.001	-0.001	0.002	0.003*
	(0.012)	(0.008)	(0.007)	(0.007)	(0.657)	(0.437)	(0.721)	(0.682)	(0.276)	(0.095)
Prop. of villages with brick street	0.329	0.499	0.516	0.432	0.170**	0.187**	0.103	-0.017	0.067	0.084
	(0.282)	(0.323)	(0.300)	(0.289)	(0.032)	(0.014)	(0.163)	(0.828)	(0.386)	(0.255)
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 mean differences between groups with p-values in parentheses. 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	1.447	1.624*	0.231	-0.177	1.216	1.393*
	(11.454)	(10.817)	(10.941)	(11.259)	(0.150)	(0.092)	(0.833)	(0.806)	(0.176)	(0.100)
Salary	36,640	38,676	36,925	38,041	2,035	285	1,400	1,751	635	-1,116
	(31,015)	(29,826)	(18,341)	(28,511)	(0.295)	(0.876)	(0.519)	(0.295)	(0.754)	(0.550)
Years of Experience	16.117	17.877	18.486	17.116	1.759*	2.369**	0.998	-0.609	0.761	1.370
	(12.418)	(11.957)	(12.113)	(12.199)	(0.067)	(0.016)	(0.344)	(0.469)	(0.428)	(0.145)
Days worked/month	14.008	12.706	13.301	13.375	-1.302	-0.707	-0.633	-0.595	-0.669	-0.074
	(6.569)	(6.604)	(6.816)	(6.892)	(0.223)	(0.517)	(0.612)	(0.608)	(0.611)	(0.956)
Hours worked/month	7.532	7.636	7.385	7.895	0.104	-0.148	0.362	0.252	-0.259	-0.510
	(4.659)	(4.587)	(4.623)	(4.574)	(0.858)	(0.814)	(0.599)	(0.691)	(0.711)	(0.494)
Village visits/month	12.527	12.939	13.558	14.369	0.412	1.031	1.842	-0.619	-1.430	-0.811
	(10.715)	(11.011)	(11.538)	(13.041)	(0.773)	(0.524)	(0.333)	(0.710)	(0.470)	(0.701)
Farmer visits/month	17.246	16.555	17.614	20.397	-0.691	0.368	3.151	-1.060	-3.843	-2.783
	(21.986)	(16.643)	(19.938)	(31.253)	(0.771)	(0.892)	(0.360)	(0.670)	(0.264)	(0.445)
Small farmer visits/month	8.902	9.763	8.912	10.222	0.862	0.010	1.320	0.852	-0.459	-1.310
	(12.069)	(11.058)	(11.504)	(14.596)	(0.551)	(0.995)	(0.450)	(0.560)	(0.792)	(0.465)
Prop. time on extension/month	0.563	0.577	0.533	0.544	0.014	-0.030	-0.019	0.044	0.033	-0.011
	(0.300)	(0.287)	(0.311)	(0.300)	(0.642)	(0.519)	(0.588)	(0.307)	(0.274)	(0.807)
Distance travelled (kms)/month	378.785	356.969	369.016	401.297	-21.816	-9.769	22.513	-12.047	-44.328	-32.281
	(1,413)	(1,410)	(949)	(1,174)	(0.841)	(0.927)	(0.854)	(0.896)	(0.686)	(0.765)
Avg. TA claim (yearly)	22,128	29,267	27,172	23,778	7,139	5,044	1,650	2,095	5,489	3,394
	(43,216)	(40,618)	(40,008)	(38,803.742)	(0.415)	(0.654)	(0.848)	(0.850)	(0.515)	(0.765)
Avg. TA payment(yearly)	16,260	20,866	18,424	15,168	4,606	2,163	-1,092	2,442	5,698	3,256
	(36,079)	(32,208)	(34,584)	(26,015)	(0.538)	(0.831)	(0.872)	(0.805)	(0.376)	(0.736)
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 mean differences between groups with p-values in parentheses. 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% level.

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

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 AO/FA 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 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.

Appendix B: Design Details

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

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

Notes: All amounts are in PKR/month. The amounts mentioned under base payment are the maximum possible that can be received when staff attendance (in terms of both days and hours is complete). The actual amount of base payment will be pro-rated based on the number of days worked in a month. The amounts in bold under the low and high bonus columns indicate the additional bonus on top of base payment.

B. 2: Payment category thresholds for FAs in the Objective Arm

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 traveled			400	400

Assumptions:

Farmers visited: Out of a 6-day workweek, an FA typically spends 1 day conducting FTPs and 1 day on office-related activities. This leaves 4 days for direct outreach. If soil sampling occurs on one of these days, it limits outreach to 3 farmers. On the remaining days, an FA should visit at least 3 farmers per day. This yields a realistic minimum of 12 farmers per week, or 48 farmers per month. For high bonus, this threshold increases by a factor of 1.5, i.e., 18 farmers per week or 72 per month. Note that FTPs can require a full working day due to responsibilities like transporting farmers to and from training locations.

Villages visited: Of the 5 active fieldwork days (with 1 day reserved for administrative duties), an FA is expected to visit one village per day. The base expectation is 20 villages per month. This threshold remains constant for high bonus to avoid incentivizing overly shallow engagement across too many villages.

Distance traveled: An FA is expected to travel an average of 20 km daily over 5 field days per week, resulting in a monthly threshold of 400 km. This remains unchanged for high bonus, since additional travel does not necessarily equate to better field performance.

B. 3: Payment category thresholds for AOs in the Objective Arm

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: 20 Barani: 10	Irrigated: 25 Barani: 15
Farmers visited			32	48
Small farmers visited			16	24
Distance travelled			640	640

Assumptions:
Farmers visited: Out of a week of 6 days, an AO spends at least 5 days doing FTPs, hence individual farmer outreach is less. A realistic threshold for AOs is 8 farmers/week, or 32/month. High bonus threshold is 1.5x this.
Villages visited: Out of 5 working days (1 day reserved for office tasks), an AO visits a different village each day, setting a threshold of 20 villages/month. This stays the same for high bonus to avoid shallow outreach.
Distance travelled: AOs travel a minimum of 32 km daily (5 days/week), giving a monthly threshold of 640 km. This remains unchanged for high bonus.