Working Paper UGA-22057

March 2024

Supervision at work: Evidence from a field experiment

This paper examines the effects of frontline supervision on the speed and quality of production in a Ugandan data-collection firm.

Ritwika Sen

IGC

DIRECTED BY

LSE

UNIVERSITY OF OXFORD

FUNDED BY

UKaid from the British people

Supervision at Work: Evidence from a Field Experiment

Ritwika Sen

This draft: March 11, 2024.
Click here for the most recent version

Abstract

How does supervision impact worker performance? This paper examines the effects of frontline supervision on the speed and quality of production in a Ugandan data-collection firm. We conducted a field experiment, varying the timing and intensity of supervision among workers, to study the effects of supervision on worker performance. We find that, following increased supervision, workers demonstrate improvements in the dimensions of performance where they were initially weakest. These effects persist, even on days, tasks, and performance metrics that are not directly supervised. Our findings suggest that supervision facilitates targeted on-the-job coaching of workers, which results in economically meaningful improvements in their performance. This underscores the significance of supervisors as conduits for the transmission of tacit production knowledge within organizations.

Keywords: Knowledge, On-the-job Coaching, Performance Monitoring, Multitasking
JEL Codes: D83, D86, J24, M53, M54

Northwestern University; Email: ritwika.sen@kellogg.northwestern.edu.
I thank Chris Udry, Mike Powell, Lori Beaman, Nicola Bianchi and Ameet Morjaria for their invaluable guidance and encouragement throughout this project. This paper would not have been possible without backing from Eliana La Ferrara, Anders Jacobsen, Andrew Dillon, and Gloria Ayesiga. I am particularly grateful to the supervisors and staff at the partner research organization for their support. I also thank Luis Rayo, Meghan Busse, Silvia Vannutelli, Vittorio Bassi, Tommaso Porzio, Gharad Bryan, Jason Kerwin, Joshua Dean, Daniel Martin, Federico Bugni, Joel Mokyr and Marie-Louise Decamps for helpful suggestions that contributed to substantive changes, along with participants at the NU Development Seminar, Kellogg Strategy Brown Bag & Development Day Conference 2023. I gratefully acknowledge financial support from the International Growth Centre Grant: UGA-22057. All errors are my own.

1. Introduction

Supervisors are very common. In the United States, frontline supervisors constitute just under 10% of the labor force across a range of vital production sectors.[1] While the precise nature of their roles vary, an important function of supervisors is often to enhance worker performance by investing in their development. Supervisors impart production tips or “know how", enabling workers to solve problems they encounter at the workplace more effectively. Thereby they serve as conduits for the transmission of *tacit knowledge* in organizations (Polanyi 1962). To fulfill this role, supervisors gather information on worker performance, diagnose areas for improvement, and provide feedback. In short, supervisors coach workers.

This view of supervisors is not intended to be controversial and hopefully resonates with anyone who has encountered supervisors at work. Yet it is not what we tend to focus on while discussing the *economic role of supervision* in organizations. The predominant discourse highlights performance measurement and incentive provision (Alchian and Demsetz 1972; Mookherjee 2013). While supervisors evaluate worker performance and administer rewards and penalties, this is not all they do. This matters, because if coaching is an important function, then the way in which supervisors add value will be different than if they were primarily monitors. For example, supervisors (as coaches) can ramp up the pace at which workers acquire production skills on-the-job. This holds different implications for how organizations should go about selecting, deploying, and planning career paths for supervisors.

We design and run a field experiment to study the causal effects of supervision on the performance of workers who conduct household interviews for a research organization in Uganda. We show that the effects of supervision are consistent with the notion that supervisors provide *targeted coaching*. Through the experiment, we uncover the unobserved actions of supervisors by testing if the effects of supervision are *persistent* and systematically *targeted* to characteristics that predict workers' initial weaknesses. We provide evidence that workers who receive supervision show persistent improvements in performance, even on days, production tasks, and performance metrics where they are not directly supervised. Moreover, these effects exhibit a multidimensional and predictably heterogeneous pattern. Workers who were initially predicted to be slow-paced, increase the speed at which they conduct interviews. Conversely, workers who

[1] Data from the US Bureau of Labor Statistics indicates that first-line supervisors account for 8% of employment in production, 10% in sales, and 9% in food preparation and service occupations (BLS 2022).

were initially predicted to be less attentive consistently record higher quality data.

Identifying the effects of worker supervision in empirical settings is challenging for several reasons. First, estimating persistent effects requires a dataset that tracks workers' performance over time. Second, studying multidimensional and heterogeneous treatment effects requires consistent and comparable measures of worker performance along multiple dimensions. Third, we need precise records of which workers receive supervision and when. Simply documenting general supervision practices in a firm is insufficient; the granularity of performance data is crucial to discern the sources of performance improvements. Lastly, estimating the causal effects of worker supervision requires exogenous variation in the timing and intensity of supervision across different workers.

The above requirements are a tall order for observational data alone. To overcome these challenges, we conducted a field experiment *within* a leading non-profit research organization. We collaborated with our partner organization to embed an experiment within a 30-day household agricultural survey project set in the Central Region of Uganda. The household survey was conducted for purposes of a distinct research study on the agricultural knowledge, practices, and output of farming households. The questionnaire comprised of detailed questions designed primarily to estimate household crop production from the preceding agricultural season. The goal of our partner organization was to collect high-quality data from over 3,330 farming households quickly and cost-effectively. We embedded the experiment within the survey to study the effects of supervision on the performance of the 68 workers employed to conduct these household interviews. The setting permits us to study the *speed* and *quality* of the interviews that they each administer over the course of the project.

An important feature of this setting is that the quality of data collected by these workers is not observable to the firm, whereas quantity (and speed) is easier to observe. Specifically, the firm does not directly observe whether workers cut corners to expedite the completion of lengthy household interviews or attend to details and pose follow-up queries to ensure the accuracy of survey data they collect. The key trade-off faced by these workers is between interview speed and the depth of probing questions they pose to interview respondents. As an example of the time-consuming probing questions we have in mind, an interviewer might ask farmers to confirm their estimates of crop volumes harvested in a previous agricultural season.[2] Worker performance is motivated by the threat of dismissals, combined with prospects for employment and promotions

[2] It is often challenging for farmers to recall how much they harvested, especially in a context where home consumption of farm produce is commonplace.

in future projects. They are paid a fixed wage each day with no incentive component as the firm is unable to systematically observe and contract on data quality. Supervisors fill this important informational gap by overseeing worker performance in the field.

While it is typically challenging to measure supervisors' inputs into the performance of any *one* worker, the predominant field supervision practices in this production context enables workers to receive individualized attention.[3] A key responsibility of frontline supervisors is to monitor worker performance by conducting unannounced visits ('spot checks') during household interviews. During spot checks, supervisors silently monitor a worker conducting a household interview and often advise them on areas for improvement afterwards. To investigate the effects of supervision, our experiment consists of three components that build on this prevailing practice.

The first component experimentally varies the assignment of workers to spot checks throughout the 30-day survey project. This daily assignment procedure, which involves sampling with replacement, introduces random variation in worker supervision on any given day, as well as in the history of their *prior* exposure to supervision throughout the project cycle. Our identification strategy leverages the independence in assignments over time to estimate the persistent effects of prior supervision exposure on subsequent performance, even on days when workers are not supervised. We estimate the effects of cumulative supervision exposure during the first week of the project on worker performance in weeks two to five.[4]

The second component of our experiment leverages the multiple tasks involved in conducting a lengthy (approximately 2-hour long) household interview. The interview questionnaire comprises of twenty modules, or groups of related questions, that are administered for purposes of the underlying study. In practice, this implies that workers (and supervisors) typically allocate their attention and effort across survey modules. We equipped supervisors with *checklists* for eight out of the twenty survey modules during their spot check visits, which prompted them to monitor worker performance more carefully on this *subset* of survey modules.

By assessing module-specific performance measures, we evaluate how workers adapt to the firm's supervision policy. In particular, we investigate whether workers lower their performance on survey modules subject to less intensive supervision, testing

[3] Production activities are spatially dispersed, with an average distance of 111 meters across households. This ensures that interviews and supervisor checks are carried out independently of any other workers.

[4] By design, whether a worker is supervised on a particular day in weeks two to five is independent of their supervision exposure in the first week, which also enables us to distinguish the persistent and contemporaneous effects of supervision.

if they learned to game the firm's supervisory practices. Put differently, we determine if apparent enhancements in worker performance due to supervision also extend to tasks that supervisors did not monitor closely. This distinctive aspect of our experimental design allows us to explore whether workers genuinely improve their production skills or simply respond to incentives.

Lastly, for four out of the eight checklist (or high-intensity supervision) modules, we supplied supervisors with daily reports regarding one specific aspect of worker performance, namely, their speed. This intervention allows us to gauge if supervisors respond to information by providing feedback that is tailored to worker speed. Collectively, these features of the experiment shed light on the mechanisms through which supervision induces persistent changes in worker performance.

We develop empirical predictions about how supervisors impact worker performance by augmenting a standard multi-tasking model (Holmström and Milgrom 1991) with a supervisor who can make targeted investments in worker skills. Specifically, a worker chooses how much effort to put into production speed and output quality. The marginal costs of these efforts depend on their skills – how fast they are and how attentive they are. The two efforts are complements (i.e., the marginal cost of production speed is higher, the higher is output quality). The supervisor observes the worker's skills and "coaches" them, investing a fixed amount of time into improving one or both skills. The impacts of these investments are modeled as a shift in the worker's marginal costs, which leads to persistent changes in worker behavior. This model clarifies how a supervisor's optimal strategy tailors coaching to workers' performance weaknesses. Fast workers, who prioritize speed over quality, tend to slow down and improve on quality; slow workers (who prioritize quality over speed), in contrast, tend to speed up.

We test four empirical predictions from the model using two principal measures of worker performance: interview completion times (speed) and the number of plot-and-crop observations recorded by each worker (quality). The latter measure captures the incidence of shirking in this context due to the use of survey rosters (i.e., groups of questions that are repeated) in the interview questionnaire. Each additional plot-and-crop reported by a worker triggers four minutes' worth of interview questions, for example, regarding harvests. Hence, under-reporting the number of plot-crops is a well-understood technique for workers to cut corners in this context.[5]

If supervisors coach workers on-the-job then we expect to see persistent improvements

[5] We use baseline measures of plot-crops cultivated by these household in a previous season as well as re-interviews for a random sample to validate this quality measure. See Section 3 for further details.

in worker performance even on days when they are not supervised. We find that past supervision results in *persistent* improvements in the quality of data collected by workers, whereas there are no statistically significant effects on their production speed. Following a one-standard-deviation (1 SD) increase in week 1 supervision, workers record 2.9% (0.20 observations; p-value: 0.082) additional plot-crops during subsequent interviews through weeks 2 to 5. However, these average effects mask considerable heterogeneity - we find that the persistent effects of supervision are *targeted* to workers' initial performance weaknesses.

Given our focus on tacit knowledge, we consider heterogeneity by workers' prior experience of conducting a similar survey. Those without prior endowments of "know how" were predictably slower in administering interviews during the first week, even after controlling for differences in output quality (with a 7-minute difference in time per plot-crop reported). We refer to workers without task experience as "low-speed workers". Conversely, we refer to experienced workers as "high-speed" workers to signify their relative emphasis on speed at the outset of the survey.

Following a 1 SD increase in week 1 supervision exposure (or approximately one additional spot check), we find that low-speed workers tend to speed up without a corresponding deterioration in their output quality. These workers decrease their interview completion times by 6.8% (7 minutes; p-value: 0.000) through weeks 2 to 5 until the end of the survey project. Moreover, prior supervision results in a marginal, and not statistically significant, increase in the plots-and-crops recorded by these workers. Conversely, high-speed workers tend to improve on output quality while slowing down their production speeds. Specifically, a 1 SD increase in week 1 supervision increases the number of plots-and-crops they record in weeks 2-5 by 5.3% (0.36 observations; p-value: 0.012). However, this increase in output quality is accompanied by a statistically significant crowd-out of their interview completion times which increase by 3.9% (3.6 minutes; p-value: 0.03) in weeks 2-5. Importantly, these effects persist even on days when workers are not supervised, suggesting that workers internalize prior feedback.

Our final set of results considers additional evidence to unpack the effects of past supervision on worker performance. We examine the possibility that workers internalize their supervisor's feedback solely because they fear sanctions for poor performance. This would imply: (a) complementarities between past and contemporaneous supervision effects; and (b) gaming on any survey modules that are monitored less intensely. We do not find any evidence in favor of this alternate mechanism. On the contrary, our evidence suggests that improvements in worker performance spillover to survey modules that are monitored less intensely. Moreover, workers also tend to improve their performance on measures that are not monitored by the firm.

Next, we consider whether prior supervision induces changes in behavior solely due to enhanced clarity on the firm's objectives rather than knowledge flows. Our findings show that coaching induces a reallocation of production time toward interview modules prioritized by the firm. However, these effects are driven by inexperienced workers to whom the prioritization of tasks is comparatively "new knowledge". We also find that the persistent effects of supervision are more closely tailored to a workers' speed performance when supervisors are given additional data on worker performance related to speed. This finding highlights the complementarity between information gathering through performance monitoring and knowledge transmission through coaching.

Lastly, we conduct a back-of-the-envelope calculation to show that the magnitude of supervision effects is economically meaningful. We estimate that for every supervisor-hour invested in coaching all workers during the first week, the firm obtains an increase in value worth 2.3 worker-hours.[6] This estimate includes data quality gains from recording 879 additional plots-and-crops (worth 216 worker-hours) and net interview time-savings of 264 worker-hours following an investment of 204 supervisor-hours in coaching workers during the first week. These results suggest that supervision is cost-effective as supervisors add value by transmitting tacit organizational knowledge to workers which contributes to persistent improvements in their performance.

Related literature and contribution. Our research is related to a core question in organizational economics, namely, "what do supervisors (or middle managers) do and how do they affect performance?" An active empirical literature studies this question using personnel data on worker performance and hierarchies within organizations. High-performing supervisors tend to improve how their subordinates perform in a variety of production settings (Bandiera et al. 2007; Lazear et al. 2015; Benson et al. 2019; Adhvaryu et al. 2023; Czura et al. 2023; Metcalfe et al. 2023). They do so by evaluating worker performance (Frederiksen et al. 2020), identifying talent (Minni 2023; Haegele 2024), and reducing turnover (Hoffman and Tadelis 2021; Friebel et al. 2022). In this paper we propose that supervisors are essential conduits to transmit tacit production knowledge within organizations through on-the-job coaching. Moreover, we design and run a field experiment to detect if targeted coaching is an important and effective role

[6] In our context, every supervisor-hour costs the firm 15% more than a worker-hour.

that supervisors perform as part of their daily activities.[7]

We also contribute to a literature on the effects of management practices on firm performance (Bloom and Van Reenen 2007; Bloom et al. 2013; Bruhn et al. 2018). Our study focuses on the subset of management practices associated with performance monitoring (Nagin et al. 2002; Gosnell et al. 2020; Jensen et al. 2020). We show that the contemporaneous effects of worker supervision vary with the visibility of distinct performance margins, particularly those (e.g. output quality) that become more apparent in the presence of supervisors. In this respect, we contribute to studies that empirically test for the presence of multitasking concerns (Fryer and Holden 2013; Hong et al. 2018; De Philippis 2021). A few papers also highlight the role of performance monitoring technologies as a potential input into managerial decision-making (Hubbard 2000, 2003; Adhvaryu et al. 2022; Kelley et al. 2023). We provide evidence of a new margin of managerial discretion that is informed by monitoring how workers perform — the choice of *how to coach* them. One implication of this decision margin is that supervisors not only influence how much effort workers exert, but also how they allocate effort across distinct performance margins.

Finally, our work is connected to studies on the significance of knowledge within organizations (Garicano 2000; Garicano and Rossi-Hansberg 2006; Gibbons and Prusak 2020). The empirical literature in this domain has largely focused on learning-by-doing (Benkard 2000; Levitt et al. 2013), and horizontal knowledge sharing among peers (Chan et al. 2014; Sandvik et al. 2020; Emanuel et al. 2023) or supervisors (Menzel 2021). Instead, we quantify the impact of top-down knowledge flows from supervisors to workers on performance.[8] In so doing, we shed light on the important role of supervisors in transmitting and deploying the use of (otherwise dispersed) knowledge in production.

1.1. Structure of the Paper

The rest of the paper is organized as follows: Section 2 describes the institutional setting. Section 3 introduces our data sources and key empirical patterns. Section 4 introduces a stylized theoretical framework and develops several predictions. Section 5 lays out the research design and empirical strategy. Section 6 presents the main results, a discussion of effect magnitudes, and robustness checks. Section 7 explores the causal mechanisms that contribute to the persistent effects of supervision. Section 8 concludes the paper.

[7] On the applied theory side, organizational economists have proposed and drawn out some of the implications of "middle managers as coaches" (Friebel and Raith 2022).

[8] Cullen and Perez-Truglia (2023) examine the impact of *social interactions* between supervisors and workers on the subsequent career trajectories of workers. They do not find any effect on performance.

2. Institutional Details

We partnered with a leading research non-profit organization in Uganda to run a field experiment embedded within a large agricultural household survey project. A firm was constituted *within* the parent organization to carry out the underlying survey.[9] Personnel were recruited through advertisements in national newspapers and popular job search platforms. All applicants were required to meet basic screening criteria (e.g., possession of a high school diploma) and successfully complete a week-long training period and qualifying exam in order to be recruited. In this section, we describe key features of the production environment in this data-collection firm, including the production process, organizational structure and performance incentives.

2.1. The Production Process

Data collection took place through private, in-person interviews conducted by individual workers (rather than teams) in the households of survey respondents. Each worker was equipped with a Personal Digital Assistant (PDA) device to administer the interview questionnaire and record responses. Production activities were also spatially dispersed, with an average distance of 111 meters to the nearest household in our study sample. These features of the environment attenuate concerns regarding complementarities in the production technology across workers or spillovers from the observation of peers' actions. Indeed, workers report that no other worker was present *at any time* during 97% of household interviews in our study.

The interview questionnaire was organized into 20 distinct modules on various topics of interest to the underlying research project, ranging from farmers' knowledge of agricultural practices to crop harvests in the preceding agricultural season. A notable feature of the questionnaire design was the use of *survey rosters* (or repeated groups of interview questions), a standard way to collect farmer-reported measures of agricultural output at a granular plot-and-crop level. Figure 1A shows a picture from a typical agricultural household plot in Uganda. It is a common practice for households to grow different types of crops on the same plot of land in our setting, such as maize, beans, and cassava as shown in the picture. To compute a reliable record of crop output, interviewers were required to ask farmers questions regarding their crop harvests (and other related information) separately for each plot and crop. Figure 1B depicts a set of sample questions that would need to be repeated three times (i.e., once for each crop)

[9] It is standard practice in this setting to employ personnel on a project-by-project basis.

to record the agricultural output from the plot shown above in Figure 1A. This design feature created a one-to-one link between the number of crops on each plot recorded by an interviewer and the time required to complete the household interview. We will focus on this in-built trade off between the number of plot and crop observations reported by a worker and their interview completion speeds to construct a measure of data quality for this study (see Section 3 for details).[10]

2.2. Organizational Structure

The data-collection firm was organized in a four-tier hierarchy with an aim to maximize production *speed* and *quality*. This structure is commonly applied across projects run by the parent organization.

The lowest tier (tier 1) comprised sixty-eight Field Officers who were responsible for conducting household interviews. Although workers were required to conduct interviews independently, they were organized into six teams to facilitate accountability, swift problem solving, and travel to field sites. These teams were managed by six Team Leaders (in tier 2) who, in turn, were responsible for production and quality control in the field. Team Leaders assigned workers production *quantity* targets of two to three household interviews each day. It was mandatory for workers to report on the status of every assigned interview before the close of business, making it easy to monitor whether quantity targets were met. However, it was harder for the Team Leaders to verify the *quality* of data collected. To assess the quality dimension of worker performance, Team Leaders conducted random "spot checks" where they sat in on household interviews as silent monitors. Spot checks were typically followed by one-on-one conversations, where Team Leaders provided feedback to workers on areas for improvement at the end of the household interview. This shop-floor practice for quality monitoring and feedback constitutes a key feature of our experiment design discussed in Section 5. We shall focus on the supervisory practices of Team Leaders and their impact on Field Officer performance in this study. Going forward, we shall refer to the Team Leaders as "supervisors" and Field Officers as "workers".

Tiers 3 and 4 of the organization comprised senior management (Field Managers, a Research Associate) who operated from the field office and coordinated activities across production teams. They were also responsible for handling any production problems referred to them by the Team Leaders. Given the complexity of verifying

[10]As noted in Carletto et al. (2021), it is a common choice in agricultural surveys to measure crop output at the plot-level as production conditions, ownership, and management patterns often vary across plots.

data quality, the firm also recruited auditors to conduct back checks or re-interviews for a random subset of households. We use data generated from these back checks to validate our key performance outcomes in Section 3. Although back checks contributed to performance monitoring in this firm, we do not explicitly study their role in shaping worker performance and confine our focus to *frontline* supervision.

2.3. Performance Incentives

Incentives were closely linked to the parent organization's objectives of producing a high volume of good-quality data. Performance was primarily motivated by the threat of dismissals, combined with prospects for employment and promotions in future projects. Workers were paid a fixed wage each day with no incentive component as the firm was unable to systematically observe and contract on data quality. The firm did not pay workers piece-rates tied to production quantity, as that may have motivated workers to prioritize interview speed at the expense of data quality (Holmström and Milgrom 1991). Any available evidence of intentionally poor quality work, such as data falsification, was treated very seriously in this setting; workers could be dismissed from the ongoing project and *any* future projects run by the parent organization.

Promotions to supervisory or management roles in subsequent projects were based on performance reviews provided by field supervisors to the central Human Resources division.[11] This was an attractive prospect for workers, as promotions were accompanied by a sizeable pay raise. For example, the supervisors in tier 2 earned 15% more than production workers in tier 1. Supervisors faced similar incentives, as promotions to senior management roles in tier 3 (as well as demotions to tier 1) in subsequent projects are commonplace in this setting.

3. Data

We use three main sources of data for this study. The first source is survey data on the characteristics of workers we collected *prior* to the commencement of production activities. The second consists of daily reports from supervisors *and* workers on their communication in the workplace. The third source consists of production data that we compile (together with the data-collection firm) on the speed and quality of household

[11] To minimize concerns about favoritism and influence activities, the supervisors were: (i) assigned to production teams randomly; and, (ii) rotated across teams every 10 days. Moreover, workers were also asked to rate the performance of their supervisors at the end of each project.

interviews conducted by workers. In addition, we use rich information on the traits of households being interviewed from a prior ("baseline") survey wave conducted for purposes of the underlying study. In this section, we provide details on each of these data sources. This is followed by a description of three key empirical patterns observed in the data, setting the stage for subsequent analysis.

3.1. Worker Survey Data

The worker survey was administered in October 2022 during a week-long pre-training period for the data collection project.[12] The survey was designed to measure dimensions of worker quality that could predict their performance on-the-job such as work history, cognitive skills, communication skills and personality traits.

Table 1 presents descriptive statistics from our study sample of 68 workers recruited to conduct household surveys. Workers were predominantly female (65%) and 30 years old on average. Most workers had completed at least one post-secondary education degree at the time of the interview, with an average of 17 years of schooling. This is a highly educated workforce for the country setting - 30% of Uganda's population aged 18 or above had a post-secondary education degree as per the 2014 National Population and Housing Census (UBOS 2017, p. 46).

We collected measures of sector, firm, and task-specific experience to capture the potential impact of accrued human capital on production performance (Becker 1962; Gibbons and Waldman 2004). The average (median) worker had 39 (30) months of experience conducting surveys and 24 (9) months of experience working on projects run by the parent organization. This suggests that worker behavior observed in our relatively short-term (one month long) data collection project was representative of regular workplace conditions as these workers were accustomed to employment on a project-by-project basis, and several of them had shared work histories.

Furthermore, 43% of the workers had relevant task-specific exposure from experience working on a previous survey wave for the underlying study. The baseline household survey was conducted approximately six months before the current (midline) survey, and several aspects of the household questionnaire (such as the methodology for recording crop harvests or specific survey questions) were repeated across the two survey waves. A binary variable for experience in the baseline survey wave will constitute

[12]The consent script for the survey clearly stated that (un)willingness to participate in the survey would have no bearing on employment outcomes for workers. Moreover, survey responses would be used solely for research purposes: no personally identifiable information would be made available to their employer.

a key measure of task-specific experience in this paper.

The worker survey consisted of modules to measure cognitive skills that could contribute to performance differences across workers. The nature of production activities (e.g., accurately recording crop harvests across agricultural plots) required a broad set of cognitive skills including memory, arithmetic, and analytical thinking. To measure these aspects of workers' cognitive skills, we selected three different cognitive tests commonly used in the literature (Laajaj and Macours 2021): (i) The digit span test (both forwards and backwards); (ii) a timed arithmetic test with basic addition, subtraction, multiplication and division questions; and, (iii) the 10-item Raven's matrices test. We combine these metrics to construct a standardized index of cognitive ability for purposes of the experiment design and analysis. We also collected measures of non-cognitive skills following the literature on the predictive power of personality traits for worker performance including: (i) locus of control and (ii) self esteem (Adhvaryu et al. 2023).[13] Summary statistics for these variables are presented in Table 1.

The worker questionnaire was supplemented by a communication skills assessment that was carried out during the pre-training week. Supervisors were provided a set of assessment criteria to rank the communication skills of workers while they performed mock interviews. The supervisors used a five-point Likert Scale to rate performance on uniformity (i.e., following the household survey script), intonation, presence, and effectiveness of communication. The criteria for good communication in this setting were identified through prior focus-group interviews with the supervisors and senior management team. We aggregate these categories to construct a standardized index of communication skills, with appropriate corrections for grader fixed effects.

3.2. Workplace Communication Data

We collect records of workplace communication through brief post-interview reports recorded by workers and spot-check reports filled by supervisors. Our data allows us to trace how often supervisors monitored interviews in the field, whether they provided any feedback to workers following a spot check, and what topics they discussed during these conversations. The data shows that 13% of interviews were spot checked, and 57% of checks were followed by production tips from supervisors.

Figure 2 highlights the prevalence of knowledge sharing in these exchanges. In

[13] The survey also comprised of questions to measure the Big Five Personality Traits. However, we do not utilize this data as we found negative correlations between measures designed to study the same latent trait.

Panel A we see that conversations between workers and their supervisors regarding an interview largely took place face-to-face. Panel B plots supervisors' assessments of whether they could share any tips (following a spot check) to help a worker improve their performance. We see that opportunities for knowledge sharing were front-loaded in the project cycle, with supervisors identifying potential production tips 81% of the time in the first week and only 38% of the time in the last week. Moreover, the advice that supervisors shared was often linked to the firm's objectives related to production speed and quality, as illustrated by comments in supervisors' spot check reports. On the one hand, the supervisors advised some workers to *work faster* noting that "Time management is not good yet" or that "The interviewer is good, but still slow in one way or the other". On the other hand, supervisors advised some workers to *slow down* and pay attention to details observing that "She needs enough time to go through all plots and the inter-crops on each of the plots ..." or "She reads faster, I advised her to slow down".

3.3. Production Data

We gathered production records from 3,330 household interviews to assess production speeds and quality. This dataset covers 1,777 worker-days, representing the level of experimental variation. It enables us to monitor the performance of 68 workers throughout a 30-day project cycle.[14] While our worker sample size is limited, a notable aspect of our data is its capacity to examine the workers' performance across a wide range of performance indicators, some of which were not consistently observed by their supervisors over time. Table 2 provides an overview of these outcome measures and the frequency at which they were measured. We discuss each group of outcomes in turn.

3.3.1. Speed: Outcomes

We compute the time spent by each worker to complete a household interview and its component survey modules using timestamps that we programmed into the digital interview questionnaire.[15] The value calculated by these timestamps computes the duration the interview (or module) form was open on a worker's screen. This excludes, for example, the time when the device was powered off or another app was in-use prior to the finalization of the questionnaire. We aggregate across "priority modules" used to

[14]The median number of days per worker is 27 due to absences, often as a result of illness, for example, malaria.

[15]The questionnaire was designed and hosted by the widely used Survey CTO platform.

measure primary outcomes for the underlying study to construct a summary measure of time spent on important modules from the firm's perspective. Panel B of Table 2 summarizes worker performance in the first week of the project cycle. We see that the average worker spent 147 minutes on each interview and 81 minutes on the priority modules within an interview. The data also shows significant variation in interview times with a standard deviation of 34 minutes *across* workers.

3.3.2. Quality: Outcomes

The measurement of survey-data quality is challenging for two important reasons. First, it is typically not possible to verify the accuracy of information gathered against some independent benchmark of the truth (De Weerdt et al. 2020). Second, the difficulty of eliciting accurate information can vary with the traits of respondents, which compounds the challenge of attributing differences in output quality to *worker* performance. The first concern is specific to data-collection tasks, whereas the second may be encountered in a variety of customer-facing production settings, for example, call centers, healthcare, or sales. In this study, we leverage *design* choices regarding the interview questionnaire and implementation of fieldwork to limit these measurement concerns.[16]

Plot-and-Crop Observations. Our principal measure of data quality focuses on behavioral choices made by workers and captures the likelihood that they *cut corners* to hasten the completion of an interview (recall that workers were accountable for daily interview quantity targets). In practice, if worker misconduct such as shirking is identified, it may require the firm to conduct expensive re-interviews. In extreme cases, clients are advised to exclude *all* data collected by the errant interviewer from their analysis.

We use an important design feature of the questionnaire that contributed to interview length to construct a measure of shirking: the use of rosters to record household farm output at the plot-and-crop level (see Figure 1 for an illustration). We construct a continuous measure based on the number of plot-and-crop observations recorded by a worker during each interview. For example, if a worker reports that a household grew maize on two distinct plots of land, the number of plot-and-crop observations (or "crop count" in shorthand) is set to two. In our data, the average worker reported that a household had 6.4 plots-and-crops during the first week of the project. We also observe a standard deviation of 1.2 plots-and-crops across workers, with a range of

[16] Dillon et al. (2020) review how survey and fieldwork design choices in a variety of contexts influence the data quality production function.

values spanning 5.4 to 7.1 plots-and-crops between the 25th and 75th percentile. To construct a measure of *shirking*, we use the negative value of the crop count (in logs) in all our regression specifications. Under this rescaling, an interview with a higher number of plot-and-crop observations reported by a worker is tantamount to a lower likelihood of shirking.

We use the crop count recorded by a worker as a measure of shirking behavior for a number of reasons. First, as detailed in Appendix Table A1, a one-standard-deviation (1 SD) reduction in the crop count reported by a worker corresponds to an 11.4 percentage point (or 34.5% increase off the baseline) increase in the likelihood that a backcheck reveals a greater crop count. This correlation holds up even when controlling for fixed worker and back checker characteristics (Columns 1-3). Second, we establish that a 1 SD drop in the crop count reported also results in time savings of approximately 11 minutes, a 15.2% reduction, in priority survey modules. This correlation is robust to adjustments for time-invariant worker characteristics (Columns 4-5).

Using the crop count as a measure of a worker's output quality requires assuming households are randomly assigned to workers, and so any increase in the recorded number of plots or crops as a result of supervision is attributed to workers providing more comprehensive information about a household's agricultural output. In all our regression models using this outcome measure, we include controls for baseline household characteristics, such as the number of plot-and-crops cultivated in a prior agricultural season and the household's size. This is done to address concerns that our measurements might simply reflect differences between the households assigned to workers for interviewing. Notably, our interview allocation protocol ensured that workers interviewed the same (randomly assigned) distribution of households.[17] In Appendix Table A2 we verify that there were no significant departures from this random assignment procedure in practice.

Divergence. Our second measure of data quality is an index of Divergence which measures the dissimilarity in primary outcome data gathered by each worker, relative to all other workers. Under the maintained hypothesis that workers all draw from the same distribution of households (see Appendix Table A2), high quality information from any one worker should be a good match for the rest of the sample. Any systematic deviations, for example a high probability mass attributed by one worker to an otherwise

[17]The randomization was carried out at the level of geographically clustered groups of 2-3 households within a village. Supervisors oversaw the randomization process and neither they nor workers had any information regarding the households, besides location and contact information, at this stage.

unlikely response, represents the idiosyncrasies of the data-collector rather than the respondents whom they interview.

To construct the index we calculate the relative entropy or Kullback Leibler (KL) divergence for all primary outcomes of the underlying study.[18] We construct these measures separately for every worker in three-day bins, allowing us to compare a *distribution* of responses collected by the worker from multiple interviews. By construction, a measure of KL-Divergence is equal to 0 if there are no discernible differences between the main and the reference distribution. A higher (and always positive) value implies a larger degree of dissimilarity. We standardize the divergence measures corresponding to each outcome, and use an equally weighted average across z-scores to construct an aggregate KL Divergence Index. The Index is standardized, in turn, implying that our measure of KL Divergence has mean 0 and a standard deviation of 1. The mean represents the average mismatch in primary outcome distributions across workers and higher values correspond to a larger than average divergence. It is important to note that this outcome measure is not monitored by the firm and was only computed by us (researchers) upon the conclusion of the household survey.

3.3.3. Quality: Process

An important feature of our setting is that we can measure worker behavior on a range of production processes that are typically associated with data quality. These measures gauge the extent to which workers do things *by the book*, rather than if they do things correctly. Our evaluation of quality process indicators, in conjunction with the quality outcomes discussed earlier, allows us to determine whether supervision primarily encourages adherence to protocol or results in improved data quality outcomes. We construct the the following indicators of quality processes:

Deviate from Interview Order. We construct a binary variable to record any deviations from the prescribed order of modules while administering the interview questionnaire. The digital mode of data collection allows us to reconstruct the order in which workers administered the questionnaire using timestamps. The average worker deviated from the prescribed interview order in 15% of their interviews during the first week of the project, with a range spanning 0% at the 25th percentile and 22% at the 75th percentile (see Table 2).

[18] See https://www.socialscienceregistry.org/trials/10198here for a pre-registered list of outcomes.

Substitute for the Household Head. Next, we measure if a worker decided to interview a family member other than the household head. During pre-training, workers were strongly advised to interview the head of their assigned household(s). In the event that the head was unavailable, they were advised to interview the most knowledgeable household member available, such as the spouse. Accordingly, we define a dummy variable which is set to one if a worker decided to interview a substitute for the head of the household, and zero otherwise. The average worker interviewed a substitute for the household head in 25% of their interviews during the first week of the project.

Use Missing Data Codes. The third measure is a binary variable that is coded to one if the incidence of missing data codes, to indicate the extent of unobtainable information in an interview, exceeds the distribution median. This includes codes to convey if a respondent did not know the answer to a question or refused to answer. The average worker used above median codes in 51% of their interviews during the first week, with values ranging from 30% at the 25th percentile to 71% at the 75th percentile of the distribution of workers.

3.4. Empirical Patterns

In this subsection, we document three key empirical patterns pertaining to the speed and quality of data collection. In the next section, we will discuss the potential effects of supervision in light of these empirical patterns.

Pattern 1: It Takes Time to Be More Attentive. Figure 3A illustrates the correlation between interview duration (on the x-axis) and the recorded number of plots-and-crops (on the y-axis), while taking into account household characteristics and worker fixed effects. The binned scatter plot clearly shows a strong positive correlation between the number of crops recorded and the time invested in conducting the interview. Notably, for each additional plot-crop reported, there is an associated increase of approximately 4 minutes in the interview duration. We include controls for household traits, such as the plots-and-crops cultivated in a previous agricultural season, household wealth, and family size. Consequently, this measurement reflects changes in the expected crop count (and related information) for a given household type, rather than the additional time required to interview larger or wealthier households with more land. This pattern aligns with a critical decision made by data production workers in this context – the trade-off between interview speed and the depth of probing questions they ask farmers, such as those related to the number of crops grown on each plot or the verification of crop harvest estimates.

Pattern 2: Different Workers have Different Weaknesses. In Figure 3B we study the dispersion in the median number of plots-and-crops recorded (on the y-axis) and median interview completion times (on the x-axis) *across* workers. The binned scatter plot shows two important features. First, we see a notable dispersion in performance across workers - especially on the time dimension of performance. Second, we observe a significant positive correlation between median interview times and plot-crops recorded, implying that some workers prioritize interview completion speeds over data quality (and vice-versa). This suggests that workers may have varying skills when it comes to interview speed and data quality. On the one hand, some workers may hurry through the interview script, without asking too many follow-up questions. On the other hand, other workers are more attentive to details, asking farmers several probing questions which results in a longer interview length.

Pattern 3: Worker Production Speeds Tend to Converge. Figures 3C and 3D show two stark patterns in our data. We plot the dispersion in worker performance over the project cycle, separately for median interview times and the number of plots-and-crops recorded. The light grey, dark grey and red density plots show the distribution of performance across workers in week 1, weeks 2 and 3, and weeks 4 and 5, respectively. In Figure 3C we observe a notable convergence in worker interview times over the project cycle, with most of the change occurring in weeks 2-3 itself (and continued shrinkage through weeks 4-5). We also observe a leftward shift over time, indicating that workers become faster at conducting interviews with on-the-job experience. In Figure 3D we also observe a pattern of convergence in the number of plot-and-crop observations, though modest. Notably, the left tail of the week 1 distribution of plots-and-crops disappears over time as workers become more familiar with the interview task and structure. In the remainder of this paper, we examine whether (and how) worker supervision contributes to these stark patterns of performance improvements and convergence.

4. A Stylized Model of Workplace Supervision

In this section we provide a conceptual framework to explore how supervision affects the performance and effort allocation choices of workers. Our framework is linked to the canonical model of multitasking in Holmström and Milgrom (1991). The contribution of this paper is to include instruments available to a firm *beyond incentives*, specifically investments of time in coaching workers, to address the multitask agency problem. This section concludes with four empirical predictions in Section 4.5 that describe how supervision, consisting of monitoring and coaching, affects the effort (and effort margins) chosen by worker.

4.1. Model Setup

Actors. A supervisor assigns a worker to perform a production task. The worker ("she") chooses how much effort $e = (s, q)$ to exert to complete the task at speed s and quality q. The worker is endowed with task-specific production skills that determine how fast ($\underline{\mathrm{f}}$) and how attentive ($\underline{\mathrm{a}}$) she is. A faster worker produces at a higher speed, whereas a more attentive worker produces better quality output. We model this feature through a worker's cost function in Equation 1, where $\underline{\mathrm{f}}$ and $\underline{\mathrm{a}}$ reduce the marginal cost of producing higher speed and quality respectively.

$$C(s, q) = \frac{1}{2}\left((s - \underline{\mathrm{f}})^2 + (q - \underline{\mathrm{a}})^2 + 2\gamma sq\right) \tag{1}$$

The cost function in Equation 1 is strictly convex, and attains an interior minimum at a finite, strictly positive vector as in Holmström and Milgrom (1991).[19] The parameter γ modulates the impact of speed on the marginal cost of quality and vice-versa. We assume that $\gamma > 0$, which implies that it is harder for the worker to be *more attentive* at higher speeds, or to work *faster* while maintaining a higher output quality. This captures the key trade-off that workers face in our production setting - whether to work faster or ask farmers more probing questions while conducting an interview.

Production. The worker is assigned to work for $H > 0$ hours, which we normalize to 1 for simplicity. A worker's choice of speed and quality effort produces two *measurable* output dimensions that are always observed in our data. From the supervisor's perspective, the workers' output is valued as a geometric average of their speed and quality:

$$V(s, q) = s^{\phi} q^{1-\phi} \tag{2}$$

[19] We assume that $0 < \gamma < 1$ and $\gamma < \frac{\underline{\mathrm{f}}}{\underline{\mathrm{a}}} < \frac{1}{\gamma}$ throughout, which is required for an interior solution.

The parameter $\phi \in [0, 1]$ modulates the emphasis on speed over quality in the objective function. Three substantive assumptions regarding the choice of this objective function include: (i) there are diminishing returns to effort ($V_{ss} < 0$, $V_{qq} < 0$), (ii) distinct effort dimensions are complements in the data production function ($V_{sq} > 0$) and, (iii) the Inada conditions are satisfied ($\lim_{s \to 0} V_s = \infty$ and $\lim_{q \to 0} V_q = \infty$). Whereas assumptions one and three are standard, the second assumption is reasonable for the specific production context. It implies that there is little demand for low quality data and a lot of demand for high quality data.

Supervisor Roles. A supervisor has a limited set of instruments available to impact worker performance. First, they can motivate the worker to exert effort by reviewing their performance on speed and quality margins through spot checks. In line with our setting, we assume that a supervisor is always able to monitor a worker's production speed, but must conduct a check to assess worker performance on the quality margin. As a result, supervisors have instantaneous (or contemporaneous) effects on worker behavior by altering the visibility of their actions. The second instrument available to a supervisor is a time budget T to coach the worker on production *skills*. The supervisor allocates their scarce time to improving how fast or attentive the worker is while conducting interviews. Consequently, coaching can achieve persistent changes in the worker's *default* action choices, that is, their choice of effort when they are not monitored or otherwise incentivized.

Timing. This is a static model, where the sequence of actions is as follows. A supervisor (always) observes a worker's speed and gathers information on output quality through spot checks. She chooses how to allocate her limited coaching time between improvements to the worker's production speed and quality. The worker observes the supervisor's choices and decides how much effort (s^*, q^*) to exert. Output is realized, the worker is paid a fixed wage, and the supervisor is the residual claimant.

4.2. The Worker's Problem

We begin with a description of a worker's utility maximization problem, highlighting the contemporaneous and persistent effects of monitoring on their effort choices. A worker chooses effort $e = (s, q)$ to maximize their utility. This consists of a fixed wage ($w > 0$) and rehiring prospects linked to a worker's production speed ($\alpha > 0$) and output quality ($\beta > 0$) net of effort costs. Notice that a worker's output quality is only under review when a worker is monitored, i.e. $\mathbb{I}_M = 1$. The utility maximization problem is as follows:

$$(3) \qquad \max_{s,q} U(s, q; \theta) = w + \alpha s + \beta q * \mathbb{I}_m - \frac{1}{2}\left((s - F)^2 + (q - A)^2 + 2\gamma sq\right)$$

where $F = \underline{\mathrm{f}} + \hat{f}$ and $A = \underline{\mathrm{a}} + \hat{a}$ are a worker's *effective* skills. Here, $\hat{f}$ and $\hat{a}$ represent the optimal coaching inputs chosen by a supervisor, which impacts a worker's marginal costs of effort linked to production speed and quality. The above functional form helps to clarify that the contemporaneous effects of monitoring are linked to a worker's incentives (β), whereas the persistent effects operate through adjustments to a worker's cost function $(\hat{f}, \hat{a})$. As before, $\theta = (\underline{\mathrm{f}}, \underline{\mathrm{a}})$ represents a worker's innate skills, and γ modulates the extent of crowding out.

4.3. The Supervisor's Problem

In this paper we do not consider how the incentives for rehiring workers (α, β) are determined, and assume that this is exogenous from the supervisor's perspective (they merely gather and report on performance information to senior management, who make the decisions in our setting). Instead, the contemporaneous effects of monitoring on worker behavior are derived solely from the worker's utility maximization problem in Equation 3. We study how a supervisor's optimization problem results in more persistent changes in worker behavior through changes in their production skills. A supervisor must decide how to allocate the time available per worker T to make them faster $\hat{f}$ or more attentive $\hat{a}$. We define $\hat{\theta} = (\hat{f}, \hat{a})$ as the bundle of adjustments chosen by the supervisor. This decision is subject to two constraints. First, the cumulative lesson time must not exceed the available time budget T. Second, a worker must choose their actions $s(\hat{\theta})$ and $q(\hat{\theta})$ optimally after observing the supervisor's choices (see Equation 3). Recognizing this, the coach determines their optimal actions to influence a worker's best response.

$$
\begin{aligned}
&\max_{\hat{\theta}} \quad V(s,q) = s(\hat{\theta})^{\phi} \cdot q(\hat{\theta})^{1-\phi} \\
&\text{s.t.} \\
&\hat{f} + \hat{a} \leq T \\
&s(\hat{\theta}), q(\hat{\theta}) = \arg\max_{s,q} U(s,q,\hat{\theta};\theta) \\
&\hat{f} \geq 0, \quad \hat{a} \geq 0
\end{aligned}
\tag{4}
$$

The solution to this problem yields $\hat{\theta}^* = \hat{\theta}^*(T;\theta)$. This permits us to define the maximized value of a worker's output to the firm as: $V^*(T;\theta) = V^*(s(\hat{\theta}^*), q(\hat{\theta}^*))$. In what follows, we will set $\phi = \frac{1}{2}$, implying that a supervisor assigns equal weight to the speed and quality of production. This is merely a simplifying assumption, and does not affect our empirical predictions.

4.4. Characterization

From the worker's problem in Equation 3, we use first-order conditions to solve for their optimal choice of speed and quality:

$$
s^*(T;\theta) = \frac{1}{1-\gamma^2}(\alpha - \gamma\beta * \mathbb{1}_M) + \underbrace{\frac{1}{1-\gamma^2}(\underline{f} - \gamma\underline{a})}_{s_0(\theta)} + \frac{1}{1-\gamma^2}(\hat{f} - \gamma\hat{a}) * \mathbb{1}_c \tag{5}
$$

$$
q^*(T;\theta) = \frac{1}{1-\gamma^2}(\beta * \mathbb{1}_M - \gamma\alpha) + \underbrace{\frac{1}{1-\gamma^2}(\underline{a} - \gamma\underline{f})}_{q_0(\theta)} + \frac{1}{1-\gamma^2}(\hat{a} - \gamma\hat{f}) * \mathbb{1}_c \tag{6}
$$

Equations 5 and 6 show that the optimal production speed and quality can be described as a function of three components. These are: i) the worker's initial skills ($\underline{f}, \underline{a}$) (ii) the contemporaneous effects of monitoring ($\mathbb{1}_M = 1$); and (iii) the persistent effects of monitoring through coaching ($\mathbb{1}_c = 1$).[20] We denote by $s_0(\theta)$ and $q_0(\theta)$ the *potential outcomes* that would be produced by a worker of type θ in the absence of any supervisory inputs. Consequently, components (ii) and (iii) describe the average

[20] A supervisor must monitor worker performance in order to coach them, i.e. $\mathbb{1}_c = 1 \implies \mathbb{1}_M = 1$.

treatment effects through contemporaneous and persistent channels that are estimated by our experiment.

PROPOSITION 1. *The contemporaneous effect of monitoring is to decrease a worker's production speed and increase output quality today, where:*

$$\triangle s^*(T;\theta) = -\frac{\gamma\beta}{1-\gamma^2} < 0$$
$$\triangle q^*(T;\theta) = +\frac{\beta}{1-\gamma^2} > 0$$

The first proposition characterizes the effects of being monitored on a worker's performance today. Monitoring makes a worker's output quality apparent to their supervisor (whereas speed is always observable), inducing the worker to prioritize effort on the quality margin. In the presence of crowding out, with $\gamma > 0$, this is accompanied by a slow down in production speeds.

Next, we utilize first-order conditions from the supervisor's time allocation problem in Equation 4 to characterize their optimal coaching inputs as a function of worker's innate skills.

$$\hat{f}^*(T,\theta) = \max\left\{\frac{1}{2}[T + \beta - \alpha + \underline{a} - \underline{f}], 0\right\} \tag{7}$$
$$\hat{a}^*(T,\theta) = \max\left\{\frac{1}{2}[T + \alpha - \beta + \underline{f} - \underline{a}], 0\right\} \tag{8}$$

We observe two distinct patterns. First, the optimal investment in improving a worker's speed $\hat{f}$ is declining in their natural skills in this domain ($\underline{f}$). Second, the optimal investment in a worker's speed is increasing in their natural attentiveness ($\underline{a}$). Optimal investments in a worker's attentiveness ($\hat{a}$) follows a symmetric pattern. Intuitively, this occurs because the supervisor's objective function is concave in each individual component (speed and quality). Consequently, the returns to improving a worker's speed decrease (increase) in any parameters that raise (lower) a worker's optimal speed choice. However, it is not possible to directly test these comparative statics as we do not measure the content of supervisors' inputs into worker performance. Instead, we utilize the supervisor's optimal choice to derive predicted changes in *measurable* quantities, specifically worker speed and output quality.

PROPOSITION 2. *Coaching exerts a persistent effect on a worker's production speed and* *quality. The effects are contingent upon a worker's innate skills, as follows:*

$$\triangle s^*(T;\theta) = \begin{cases} = \frac{T}{1-\gamma^2} & iff \leq C_1 \\ = \frac{1}{2(1-\gamma)}[(\beta-\alpha)+\underline{a}-\underline{f}] + \frac{T}{2(1+\gamma)} & iff \in (C_1, C_2) \\ = \frac{-\gamma T}{1-\gamma^2} & iff \geq C_2 \end{cases}$$

$$\triangle q^*(T;\theta) = \begin{cases} = \frac{-\gamma T}{1-\gamma^2} & iff \leq C_1 \\ = \frac{1}{2(1-\gamma)}[(\alpha-\beta)+\underline{f}-\underline{a}] + \frac{T}{2(1+\gamma)} & iff \in (C_1, C_2) \\ = \frac{T}{1-\gamma^2} & iff \geq C_2 \end{cases}$$

C_1 and C_2 denote threshold values of a worker's innate speed $\underline{\text{f}}$, beyond which a supervisor's time is allocated entirely to improvements in their speed ($\hat{f} = T$ for $\underline{\text{f}} \leq C_1$) or quality ($\hat{f} = 0$ for $\underline{\text{f}} \geq C_2$). These values are derived from corner solutions corresponding to the supervisor's problem (see Equation 7) and can be ordered on the real line with $C_1 < C_2$, as shown on the x-axis of Figure 4.

Proposition 2 demonstrates three important features. First, we can describe the changes in worker speed and quality as piece-wise linear functions depending on whether the supervisor's problem yields a corner solution - with investments in only speed or only quality - or an interior solution with positive investments in both sets of skills. Second, in the presence of crowding out ($\gamma > 0$), a corner solution with supervisor investments in only speed skills (or only quality skills) is accompanied by a deterioration in worker performance on the other margin. The net effect of an interior solution depends on the relative magnitude of adjustments in a worker's speed and quality. Third, in the case of an interior solution we observe that the change in a worker's production speed is monotonically *decreasing* in their natural speed skills ($\underline{\text{f}}$). Conversely, the changes to a worker's output quality is monotonically *increasing* in a worker's innate speed skills ($\underline{\text{f}}$). This pattern illustrates that a supervisor invests their scarce time budget to address a worker's relative weakness in the speed or quality dimension, given the supervisor's objective to *jointly* maximize production speed and quality. Intuitively, a higher ($\underline{f}$) reduces the marginal benefit of allocating time to improvements in a worker's speed and increases the marginal benefit of time allocated to quality improvements. We observe similar properties corresponding to a worker's innate attentiveness $\underline{\text{a}}$.

Given the monotone relationship of treatment effects with ($\underline{f}$), we derive thresholds to delineate when we expect to see (i) corner solutions; and (ii) positive (or negative) changes on the two performance outcomes. This is summarized in Figure 4 which plots the persistent effects of coaching on a worker's optimal production speed and output quality as a function of their innate speed skills: $\underline{f}$. The downward sloping black line depicts the expected change in production speed $\triangle s^*(T;\theta)$, whereas the upward sloping solid gray line depicts the expected change in output quality $\triangle q^*(T;\theta)$.

COROLLARY 1. *$\exists$ thresholds T_1, $T_2 \in \mathbb{R}$ such that $\triangle q^*(T;\theta) \geq 0$ only if $f \geq T_1$ and $\triangle s^*(T;\theta) \leq 0$ only if $f \geq T_2$. Moreover, $T_2 > T_1$.*

Thresholds T_1 and T_2 are labelled along the x-axis in Figure 4. Their specific ordering on the real line, with $T_2 > T_1$, permits the division of the x-axis ($x \geq 0$) into intervals featuring: (i) an increase in speed and a deterioration in quality; (ii) improvements in both speed and quality; and, (iii) a decrease in speed and an improvement in quality. This allocation highlights that a supervisor's decision on what information to share is customized to a worker's innate skills (and weaknesses).

$$
\begin{aligned}
&\triangle s(T,\theta) \geq 0 \;\&\; \triangle q^*(T,\theta) \leq 0 \text{ if } \underline{f} \in [0, T_1) \\
&\triangle s(T,\theta) \geq 0 \;\&\; \triangle q^*(T,\theta) \geq 0 \text{ if } \underline{f} \in [T_1, T_2] \qquad (9) \\
&\triangle s(T,\theta) \leq 0 \;\&\; \triangle q^*(T;\theta) \geq 0 \text{ if } \underline{f} > T_2
\end{aligned}
$$

These intervals are marked as regions I, II, and III respectively in Figure 4.

COROLLARY 2. *$\triangle s^*(T;\theta) \geq 0$ and $\triangle q^*(T;\theta) \geq 0$ iff $f \in I = [T_1, T_2]$. The length of interval I is increasing in T.*

Corollary 2 illustrates that a supervisor's decision on how to coach a worker varies with the time available to coach them, T. Conditional on requiring an improvement in speed (with $\underline{f} \leq T_2$), a worker also experience *improvements* in quality when a supervisor has a sufficient amount of time available to coach them. Conversely, a time-constrained supervisor only does one thing and invests in improving a worker's speed. This may be accompanied by a decline in output quality, depending on the strength of the crowding out parameter (γ).

4.5. Summary of Empirical Predictions

Our model provides insight into the impact of supervision on worker performance along two key dimensions: production speed and quality. We show that supervision has a contemporaneous effect through performance monitoring - enhancing worker performance in the newly observable dimension of quality. However, this improvement comes at the cost of reduced speed in the presence of crowding out. The effects of coaching are both persistent and targeted to a worker's innate skills, denoted as θ.

For workers with low-speed skills ($\underline{f} < T_1$), coaching leads to a direct increase in their production speed. Conversely, workers with high-speed skills ($\underline{f} > T_2$) experience a direct increase in their output quality. Whether these direct effects are accompanied by an indirect increase or decrease in the alternate performance margin depends on a supervisor's time constraints (T) and the extent of crowding out (γ).

In the subsequent sections, we empirically examine several predictions derived from our model. The contemporaneous effects of monitoring are as follows:

1. Lower production speeds today (with crowding out).
2. Higher output quality today.

The persistent effects of coaching manifest as:

3. Increased production speeds for low-speed workers (and perhaps a decrease in their output quality with crowding out).
4. Increased output quality for high-speed workers (and perhaps a decrease in their production speeds with crowding out).

5. Empirical Strategy

5.1. Experiment Design

This section describes how our experiment is designed to estimate the causal effects of supervision on worker performance. The desired treatment(s) must overcome two important identification challenges in a workplace (field) setting. First, there is a challenge of selection bias inherent in the allocation of a supervisor's time across workers. Second, we need a strategy to disentangle the contemporaneous and persistent effects of supervision on performance. Specifically, we aim to understand if exposure to supervision in the past has any persistent effects on worker performance. We then examine if effects conform to the empirical predictions outlined in Section 4.5. We elaborate on our approach to each of these challenges in turn.

To identify the causal effects of supervision on worker performance, we leverage a common organizational practice in survey implementation, i.e. the use of spot checks. As the name suggests, this practice entails unannounced visits conducted at random by supervisors during household interviews. In normal circumstances, the "randomization" procedure is left to a supervisor's discretion which could of course involve some selection on worker characteristics. For purposes of our experiment, we conducted the spot check assignment process in two steps to induce random variation in the allocation of supervisor time across workers.

The first step was conducted at the worker-level. 68 workers were randomly assigned to one of 6 teams, stratifying by workers' gender and prior task experience. This allocation resulted in two teams of size 12, and four teams of size 11. To improve our statistical power and covariate balance with a small sample of workers, we constructed matched blocks of workers within each survey team to minimize the Mahalanobis distance of covariates that could predict worker performance on the survey (we did not have any baseline values of performance outcomes prior to the survey). The covariates included our index of cognitive skills, communication skills, and training quiz scores. Due to unequal team sizes, these blocks comprised of 14 matched quadruplets and 4 matched triplets.

In the second step, we randomly assigned *worker-days* ($N = 68 * 30$) to spot checks, stratifying by a worker's team, matched block, and the survey day. The randomization procedure ensured that *one* worker per team and block (and three workers per team) were assigned to the spot check treatment every day.[21] This randomization procedure was adopted to meet with the supervisors' time constraints in the field, which did not permit them to conduct any more than two to three spot checks in a day. Moreover, the procedure lowered our departures from business as usual at the firm by ensuring that our experiment maintained the average intensity of spot checks otherwise practiced at the parent organization. Our intervention simply formalized the randomization procedure observed for implementing spot checks.

As illustrated in Figure 5A this randomization procedure has a number of distinctive properties. The assignment of worker-days to treatment, instead of workers, implies that the a worker could be assigned to a spot check on day t and not on day $t + 1$. As we sample with replacement, a worker could also be assigned to treatment on consecutive days. The resulting independence in treatment assignment(s) over time enables us to

[21]Notice that the unequal size of team-blocks on each worker-day induces differences in the (known) propensity of assignment to treatment across workers. All our regression specifications adjust for this imbalance with inverse probability weights.

separately identify the impact of: (i) a spot check today, and (ii) cumulative exposure to spot checks in the past on contemporaneous performance outcomes.

Indeed, the randomization procedure generates variation in a worker's cumulative exposure to *prior* spot checks on any day of the 30-day project. In our key regression specification, we utilize the intensity of checks in the *first week* of the project to identify the persistent effects of supervision on worker performance from week 2 onward. This specification was motivated by evidence that supervisors mainly coach workers early on in the project cycle, by sharing production tips upon the completion of spot checks (see Figure 2B, for example).[22] We compute the following treatment variables for every worker i to distinguish the effects of contemporaneous (D_{it}) and past supervision (M_i^1), where D_{it} is a binary variable that is coded to 1 if worker i is assigned to a spot check on day t and 0 otherwise:

$$M_i^1 = \Sigma_{s=1}^{s=6} D_{is} \tag{10}$$

The above specification utilizes one treatment vector, D_{it}, and yet it measures the impact of two qualitatively distinct interventions - across workers and over time. D_{it} measures the contemporaneous effect of a spot check through weeks 2 to 5. That is, it measures the change in performance on days when a worker's effort on both speed and quality margins (as opposed to just speed) is observable to a supervisor. M_i^1 measures the persistent and cumulative effect of checks in the early stage (week 1) of the project. Intuitively, it is as if we randomly assigned workers to a different *history* of supervision exposure at the start of the second week, independently of their later treatment assignments through weeks 2-5. The variable M_i^1 is standardized in all our regression specifications, implying that average exposure to supervision in the first week is set to 0. Figure 5 shows the variation in intensity of treatment (i.e., cumulative supervision exposure) across workers throughout the project (in Panel B) and at the end of the first week (in Panel C). The average worker was exposed to 1.6 spot checks in the first week, with a standard deviation of 1.05 checks across workers. By the end of the project, the average worker encountered 7.9 checks with a standard deviation of 2.2 checks across workers.

[22] We demonstrate robustness to alternate time horizon choices for the effect window in Section 6.4. Note that a longer time horizon implies we have fewer observations to use for analysis.

5.2. Details of the Intervention

This section presents further information about the experimental intervention. During random checks, supervisors participated in household interviews as silent observers. This approach served as an effective quality control measure by enabling supervisors to assess the performance of workers in an otherwise unobserved aspect, namely, the quality of their work. Additionally, we provided supervisors with *checklists* and *performance reports* for specific survey modules. This served two purposes: (i) drawing supervisors' attention to the speed and quality of worker performance on these modules, and (ii) encouraging supervisors to provide feedback or share production tips with the workers they spot-checked after completing the household interviews. It's important to note that supervisors typically shared their feedback with workers in one-on-one interactions after the interview had concluded.

We developed checklists for a subset of survey modules, specifically 8 out of 20. The selection of these modules was based on workers' reports of the perceived difficulty of these questions before the survey began. These modules were areas where on-the-job learning and knowledge sharing were likely to be relevant. We also prioritized modules essential for measuring primary research outcomes, as it is common to closely monitor such questions through field-level and digital supervision. The checklists guided supervisors to observe how workers interpreted specific survey questions compared to their peers, the time taken to administer survey modules, and their ability to effectively gather additional information from respondents when needed. An example of a checklist, specifically for household agricultural output computation, is provided in Appendix Figure A2. These checklists ensured the collection of reliable data from households.

Half of these modules were further randomly selected for an additional intervention. In this case, supervisors received additional information on *one dimension* of worker performance, namely speed, through high-frequency checks of electronically collected data. This information included summaries of a worker's median time to administer these survey modules compared to the rest of the firm. This intervention provided quantitative evidence to supervisors to gauge whether they respond to informational inputs and provide customized coaching.

The checklist and performance report interventions were designed to increase the likelihood of supervision (including monitoring and coaching) on specific survey modules. We categorize survey modules with checklists only as Group M1 and those with both checklists and performance reports as Group M2. The remaining survey modules where active encouragement of supervision was not a part of the experimental setup is classified as Group M0. Analyzing worker performance patterns within the M0 modules allows us to investigate whether workers adapt by reducing their effort in modules subject to lower-intensity supervision, essentially testing if they learn to game the firm's supervision policy. Furthermore, any variations in worker performance across the M1 and M2 survey modules offer valuable insights into the responsiveness of supervisors to quantitative evidence of worker performance, given a specific level of supervision intensity. The module-level heterogeneity in supervision induced by the experiment design will be essential for an exploration of causal mechanisms.

5.3. Identifying Assumptions

The experiment design relies on two core identification assumptions related to the allocation of workers to spot checks each day (D_{it}). Firstly, for any given day, we assume that the allocation was independent of workers' characteristics. Secondly, over the entire duration of the project, we assume that the allocation process was independent across time periods. We test the validity of these assumptions in turn.

Figure 6 displays regression coefficients from two sets of balance tests. In panel A, we regress a vector of daily assignments to spot checks for the entire duration of the project, D_{it}, on a bundle of worker traits. This includes a worker's gender, experience, schooling, test scores, cognitive skills, communication skills, and non-cognitive skills (all the regressors are standardized for ease of interpretation). Panel B displays coefficients from a similar regression specification, but with the sample restricted to only the first week of the project. In both cases, none of the 9 coefficients are individually statistically significant. An aggregate test of all covariates fails to reject orthogonality of the allocation procedure, for either of the two samples (with a p-value of 0.46 in Panel A and 0.74 in Panel B).

In Figure 6C we evaluate the relationship between spot check assignments over time. Given the binary nature of daily treatment assignments (i.e., 0/1), we can calculate the probability of assignment to various combinations of treatment in any two consecutive days under the assumption of independence and an average assignment probability of 0.27, specifically 0-0, 0-1, 1-0, and 1-1. The expected probability of each treatment combination is shown using dotted red lines in the figure. Next, we plot the realized probability of treatment assignment using bar charts, along with 95% confidence intervals. The figure demonstrates a close link between the expected and realized assignment probabilities, suggesting that daily assignments to treatment were not systematically correlated over time.

5.4. Compliance

We verify the extent of compliance with our experiment design using worker reports of spot checks conducted each day. In Table 3, we examine the effects of: (i) daily assignments to spot checks during weeks 2-5 (D_{it}), and (ii) the assigned intensity of monitoring in Week 1 (M_i^1) on a range of outcomes associated with *actual* worker-supervisor interactions.

Column 1 demonstrates a significant relationship: a one-standard-deviation increase in the intensity of assigned monitoring in week 1 corresponds to a 0.46 standard-deviation increase in actual coaching during that same week, supported by an F-statistic of 22.92. Encouragingly, we find that daily assignments to spot checks in weeks 2-5 did not influence the intensity of (actual) coaching in the preceding week.

Column 2 highlights another noteworthy result: assignment to a spot check throughout Weeks 2-5 increased the probability of an actual spot check by 61.3 percentage points, relative to a control mean of 4%, with a robust F-statistic of 384.7. It is worth noting that our spot check assignments were generally followed by supervisors, albeit with constraints due to the competing demands on their time during fieldwork.

Columns 3 and 4 provide further evidence of the positive impact of spot check assignments. Column 3 indicates that assignment to a spot check enhanced the probability of a supervisor's presence throughout most of the household interview. Meanwhile, Column 4 underscores that such assignments increased the likelihood that a worker received feedback from their supervisor upon completing a spot check. These findings collectively reaffirm the effectiveness of our experiment design in fostering one-on-one interactions between workers and their supervisors.

In Appendix Table A3 we further verify that the intensity of checks for prioritized survey modules, M1 and M2, exceeded supervisor checks on M0 modules by 26 and 18 percentage points, respectively (relative to a control mean of 50%). The supervisor reports used to compile the statistics in this table are only available in instances where spot checks were conducted.

6. Main Results

This section presents our main results regarding the influence of supervision on worker performance. In Section 6.1, we present our primary regression specification and its connection to the underlying conceptual framework. In Section 6.2, we test whether our findings are consistent with empirical predictions from our model. We discuss the magnitude of estimated effects at the firm-level in Section 6.3. Further, we conduct a series of robustness tests which are presented in Section 6.4.

6.1. Regression Specifications

Our core regression specification for the analysis of experimental results is provided in Equation 11. The unit of observation is a household interview h, conducted by worker i on day t of the survey, with $t \in \{8, 9, \ldots 30\}$. As specified in Section 5.1 we split the sample into two parts, corresponding to week 1 and weeks 2-5, respectively. This approach allows us to separately estimate the contemporaneous effects of supervision on day t (β_1), and the persistent effects of week 1 supervision (β_2).[23] The dependent variable y_{hit} measures the speed or quality of a household (h) interview administered by worker i on day t.

$$y_{hit} = \beta_1 D_{it} + \beta_2 M_i^1 + \gamma_0 M_{i,t-1}^{2+} + \gamma_1 X_{it} + \gamma_2 X_h + \nu_{\tau t} + \epsilon_{hit} \tag{11}$$

The regression equation includes a number of control variables. γ_0 adjusts for any imbalances across workers from (randomly assigned) differences in exposure to spot checks between week 2 and day $t-1$. $M_{i,t-1}^{2+}$ represents the sum of spot check assignments D_{it} over this period, which is standardized by day to construct a stationary time-series.

In all regression specifications we control for nonlinear time trends interacted with several worker characteristics that we measured through a worker survey prior to the commencement of the project. In particular, we include day fixed effects interacted with dummy variables that identify workers' gender, and those with above median: schooling, sector experience, locus of control and self esteem (X_{it}).

We also control for a rich set of household characteristics X_h, using data from a prior household survey wave, which are predictive of the expected length of interview time and the number of plots-and-crops cultivated by households in the current survey wave. These include controls for the plots-and-crops cultivated by the household in the prior agricultural season, household size, the gender of the household head and an index of household wealth.

The coefficient $\nu_{\tau t}$ represents the interaction of team (τ) and day fixed (t) effects, implying that we estimate a first difference across workers within each team-and-day cell. This empirical strategy enables us to estimate the effects of supervision on worker

[23]The construction of the treatment variables, D_{it} and M_i^1 is discussed in detail in Section 5.1.

performance for a *given* supervisor.

Given our focus on *tacit* production knowledge, we augment the regression framework in equation 11 to flexibly allow for qualitative differences across workers with and without prior experience of conducting a similar survey ("task experience"). To do this, we include interactions between a dummy variable for workers' prior task experience and each of our control variables: $M^{2+}_{i,t-1}, X_{it}, X_h, \nu_{\tau t}$. We also consider heterogeneity in the estimates of treatment effects (β_1, β_2) by workers' prior task experience to test empirical predictions from the model outlined in Section 4.5 (as discussed below). We estimate intent-to-treat effects and cluster standard errors by worker throughout.

Estimation of Contemporaneous Effects. To estimate the contemporaneous effects of supervision on worker performance we include an additional interaction term $D_{it} * M^1_i$ in equation 11. This modification ensures that the coefficient β_1 represents the effect of being supervised today for the *average* history of week 1 supervision ($M^1_i = 0$). The associated empirical predictions are: $\beta^S_1 \leq 0$ and $\beta^Q_1 \geq 0$, respectively, where S and Q represent speed and quality outcomes respectively. This regression specification is estimated for the full sample of workers over weeks 2 to 5.

Estimation of Persistent Effects. To test for patterns in persistent effects we estimate Equation 11 with the sample restriction $D_{it} = 0$. This ensures that we estimate the persistent effects of week 1 supervision on worker performance in weeks 2 to 5 on days when they are <u>not</u> directly supervised. The corresponding empirical predictions are: (i) $\beta^Q_{2,H} \geq 0$ and perhaps $\beta^S_{2,H} \leq 0$ (with crowd-out) for high-speed workers (H) and, (ii) $\beta^S_{2,L} \geq 0$ and perhaps $\beta^Q_{2,L} \leq 0$ (with crowd-out) for low-speed workers (L). Consequently, this regression specification is estimated separately for the sample of high and low-speed workers over unsupervised worker-days in weeks 2 to 5.

In order to test these empirical predictions we need to categorize workers based on their innate production speed for a given level of attentiveness (recall the definition of worker skills in Section 4). To accomplish this, we utilize data from the worker survey and production outcomes from the first week of the project, which is excluded from the analysis sample. Since a worker's initial performance might be influenced by supervisory guidance during on-boarding, we analyze which worker traits are the most robust predictors of their interview time (per plots-and-crops recorded) at the beginning of the project.[24]

[24]In this context, we gauge worker attentiveness based on the number of plots and crops recorded,

Figure 7 illustrates the significance of a worker's prior task experience as a predictor of median interview time per plot-and-crop during the first week. The Figure shows the cumulative distribution of median interview times (per plot-crop) for workers with prior experience (in solid black markers) and those without (in hollow gray markers). The series is standardized so that the value zero represents the performance of the average worker. We find that approximately 86% of experienced workers exhibit interview times (per plot-crop) below average, indicating above-average production speeds. In contrast, only 45% of workers without prior task experience demonstrated below-average interview times (per plot-crop) in the first week. This pattern highlights that experienced workers maintain greater speeds, even after accounting for potential variations in output quality. Accordingly, in the remainder of this paper, we will refer to workers with prior task experience as "high-speed" workers and workers without prior task experience as "low-speed workers".

Appendix Figure A3 highlights that previous task experience emerges as the most reliable predictor of interview time per plot-crop compared to other worker attributes (Panel A). We also find that workers with prior task experience primarily differ from those without such experience in terms of sector experience rather than cognitive or non-cognitive skills (Panel B).

6.2. Test of Empirical Predictions

In this section, we explore patterns in the contemporaneous and persistent effects of supervision. This analysis is guided by empirical predictions from the model in Section 4.5 and the estimation strategy described above.

Contemporaneous Effects. We begin with a test of empirical predictions 1 and 2, which suggest that workers who are supervised today will exhibit lower production speeds and higher output quality. In Table 4, we estimate equation 11 for the full sample of worker-days through weeks 2 to 5 to estimate the effects of supervision today. We augment the regression equation to include the interaction between supervision today and week 1 supervision (std.). As a result, the coefficient on 'supervise today' (β_1) estimates the impact corresponding to the average history of week 1 supervision.

We find that, consistent with empirical prediction 1, supervision today leads to a 12% or 12-minute increase in interview time (column 1, p-value: 0.000). Moreover, on days when workers are supervised, we observe a significant improve in the quality of

while controlling for specific household traits (for a comprehensive description, refer to Section 3).

data collected. In line with empirical prediction 2, supervision today results in a 4.1% increase in the number of plot-and-crop combinations recorded by a worker (column 3, p-value: 0.006). Supervised workers are also 5.5 percentage points (42.3%, p-value: 0.000) less likely to deviate from the prescribed interview order. Put together, these findings suggest that supervisors closely monitor data quality outcomes and practices while conducting their spot check rounds. However, the improvements in output quality are accompanied by a significant slowdown in production speeds.

Persistent Effects. In Table 5 we study the persistent effects of week 1 supervision (std.) on worker performance through weeks 2 to 5. To calculate these estimates, we apply equation 11 with an additional sample constraint. Specifically, we estimate the effects of past supervision on days when workers are not presently supervised ($D_{it} = 0$). This sample restriction is essential to isolate the persistent effects of prior supervision exposure on the current output of workers.

We find that week 1 supervision does not have a significant impact on the production speed of the average worker, whether it pertains to the entire interview or the priority survey modules (Columns 1 and 2). Instead, we observe a persistent improvement in the quality of data collected by workers. Following 1 standard deviation (henceforth 1 SD) increase in week 1 supervision, or approximately one spot check, workers record 2.9% (Column 3, p-value: 0.082) additional plot-crop observations during the interviews they conduct in weeks 2 to 5. This is accompanied by a 0.11 SD (Column 4, p-value: 0.045) reduction in the divergence of outcome data that they collect. Workers are also 4.1 percentage points (31.5%) more likely to adhere to the prescribed survey order through weeks 2 to 5 (Column 5, p-value: 0.06). Of these effects, however, only the average effects on the divergence index is robust to the computation of bootstrapped standard errors clustered by worker at the 10% level of statistical significance (row 6).

If supervisors provide targeted on-the-job coaching, we anticipate heterogeneity as the persistent effects of supervision will be tailored to individual performance weaknesses. Empirical predictions 3 and 4 outline the expected pattern of treatment effects by workers' innate production speeds - coaching increases the interview speed of low-speed workers and increases the output quality of high-speed workers. Additionally, coaching may also have crowd-out effects - with a decrease in the output quality of low-speed workers and an increase in the interview time of high-speed workers when supervisors are time-constrained and do only one thing.

In Table 6 we study the effects of week 1 supervision on the performance of low-speed workers. We find that a 1 SD increase in week 1 supervision intensity results in a 6.8% reduction in interview times for these workers (Column 1, p-value: 0.000). This decrease amounts to a time-saving of approximately 7 minutes per interview conducted in weeks 2 to 5 and is consistent with Prediction 3. This treatment effect is both quantitatively and qualitatively distinct from the contemporaneous effects of supervision today, which also increases the interview time of low-speed workers (see: Figure 8A).

Week 1 supervision did not result in a significant change in the quality outcomes of low-speed workers (Columns 3, 4). Moreover, these workers were 3.7 percentage points (14.8%) more likely to interview someone other than the household head (Column 6, p-value: 0.017). While this is consistent with a deviation from recommended quality practices, it does not necessarily indicate a decrease in quality outcomes. This is especially true when considering that the alternative to interviewing a substitute for the household head might have been to skip interviewing the household entirely.

In Table 7, we assess the influence of week 1 supervision on the performance of high-speed workers. We find that a 1 SD increase in week 1 supervision intensity leads to a 3.9% *increase* in interview times for these workers (Column 1, p-value: 0.03). This translates to approximately 3.6 additional minutes per interview in weeks 2 to 5. The effects on the time spent on priority interview modules is qualitatively similar (see: Column 2). These findings are consistent with Prediction 4, which posits a possible decrease in the *speed* of high-speed workers.[25]

We examine the effects of week 1 supervision on the quality outcomes of high-speed workers in columns 3 and 4. We find that, following a 1 SD increase in week 1 supervision intensity, the number of plots and crops recorded by these workers increases by 5.3%,. This is equivalent to 0.36 unit increase per interview in weeks 2 to 5 (Column 3, p-value: 0.012). Additionally, the divergence in outcome data recorded by these workers decreases by 0.28 SD units (Column 4, p-value: 0.012). These findings are consistent with the direct effects on output quality outlined in Prediction 4.

The reported improvements in quality outcomes are accompanied by an increase in non-adherence to quality process recommendations. High-speed workers were 6.6 percentage points (14.7%) more likely to use missing data codes with an above-median frequency (Column 7, p-value: 0.059). Nevertheless, it is reassuring that these process deviations coexist with overall improvements in quality outcomes.

Taking stock of our results so far, it is evident that our experimental data reveals a

[25] The reported slowdown in interview speeds is not robust to small sample corrections using bootstrapped standard errors clustered at the worker-level (p-value: 0.183). However. the slowdown in priority survey modules is robust to this correction with a p-value of 0.069.

significant level of heterogeneity in the effects of week 1 supervision on worker speed and quality outcomes (see Figure 8 for a summary). Our findings suggest that past supervision has persistent effects on worker performance, even on days when they are not supervised. Moreover, these effects are concentrated on dimensions of performance where workers initially exhibited performance weaknesses. These observations align with empirical predictions 3 and 4 of our model, alongside predictions 1 and 2, which pertain to the contemporaneous effects of supervision. This collective evidence lends support to the hypothesis that past supervision exerts a *persistent* influence on workers' effort allocation choices through targeted coaching inputs.

Appendix Figure A4 further highlights that the improvements in speed for low-speed workers and quality for high-speed workers persist in weeks 2-3 and weeks 4-5 until the conclusion of the survey project. The figure also shows heterogeneity in average treatment effects over the project cycle. We observe a significant decrease in average interview times in weeks 2-3 of the survey, but not in weeks 4-5 when the slow down among high-speed workers is more pronounced. In Section 7 we consider further evidence to characterize the causal mechanisms driving these persistent effects.

6.3. Magnitudes

To gauge the scale of coaching effects, we consider a hypothetical scenario in which all workers receive an additional coaching session with their supervisors during the first week. Our findings and back-of-the-envelope calculations can be summarized as follows: In terms of cost to the firm, this intervention would incur approximately 204 supervisor hours, under the assumption that each supervisor devotes 3 hours to each of 68 workers. Recall that a supervisor-hour is valued at only 15% more than an hour worked by a production worker.

Moving on to net savings in interview hours for the firm, our analysis suggests a significant reduction, estimated at 264 worker-hours. This estimate is calculated by aggregating changes in interview duration in weeks 2 to 5 for high-speed and low-speed workers, as inferred from our experimental data. We adjust our experimental estimates to account for partial compliance, since the scenario under consideration specifies that all workers *receive* the coaching intervention.

Furthermore, the firm stands to gain in terms of data quality, with an estimated (net) increase of 879 plot-and-crop observations. Using data on the number of plots-and-crops recorded per worker-hour, we estimate that this change in data quality is equivalent to 216 worker hours. This computation entails summing up the increase in plot-and-crop records by high-speed workers during weeks 2-5. As for low-speed workers, the net effect is presumed to be zero, given the lack of statistically significant changes within this subgroup (on average).

Summing up the estimated gains in both speed and data quality, assuming uniform weights, the firm is estimated to accrue a total gain of 480 worker-hours. This implies that for every supervisor-hour invested in coaching workers, the firm earns the equivalent of 2.27 worker-hours. Even when the analysis is limited to intent-to-treat estimates, the ratio remains significant at 1.04. In summary, directing supervisor-hours toward coaching workers in the first week is a cost-effective strategy for the firm.

It is important to note that this analysis does not compute the returns to investing in an additional hour of supervisor time, as, in practice, supervisors would optimally allocate their time across workers. The estimates here assume a uniform time allocation across workers, in line with our experiment. Consequently, our calculations provide a lower-bound on the returns to additional supervisor-time for the firm (assuming that supervisors' interests are aligned with the firm's).

6.4. Robustness Checks

In this section, we test the robustness of our findings through a series of empirical assessments. First, we employ an alternative definition of worker-speed to ensure that our observed effects are indeed driven by the workers' initial speed rather than being influenced by other factors tied to their experience (for example, workers without prior firm experience may be treated differently by supervisors). To do this, we restrict our attention to a subset of 19 low-speed workers, i.e. those *with* prior experience in the firm. This controls for the possibility that our empirical proxy for worker speed (prior task experience) may conflate two forces - their innate speed at the production task and the role of experience at the firm. Consequently, Appendix Figure A5 examines the effects of week 1 supervision on the performance of low-speed but experienced workers. The Figure confirms our predictions for the direct effects of targeted coaching on low-speed workers: they experience a significant decrease in interview times (9.2%, p-value: 0.098 with bootstrap clustered standard errors). For this subset of workers we observe a crowding-out effect with a decrease in the number of plots-and-crops recorded (by 8.4%). However, this effect is not robust to small sample corrections with a p-value of 0.238 computed with bootstrap standard errors clustered at the worker-level.

Second, we investigate the robustness of our findings to the choice of the past supervision window of one week. This choice has implications for the estimation sample, as it affects the period over which we measure the effects of past supervision on outcomes. In Appendix Figure A6, we vary the initial exposure period from the first three days up to the first ten days. Reassuringly, the estimates, especially for the first 5-8 days, closely resemble those reported in Tables 7 and 6.

Third, we assess the linear specification of the week 1 supervision treatment variable (M_i^1). In Appendix Figure A7, we visually inspect the treatment effect through scatter plots of residual outcome measures (y-axis) against the residual week 1 supervision stock (x-axis), after controlling for household and worker traits and team-day fixed effects as specified in equation 11. These binned-scatter plots demonstrate the non-parametric conditional expectation function. Reassuringly, the linear best fit line (in red) closely approximates the data points in each instance.

In a final set of tests, we examine the robustness of our findings to attrition from the worker panel. In our setting, absenteeism from the worker-day panel is relatively common, with an average attrition rate of 12.9% – typically due to illness or worker assignment to tracking households that could not be located on scheduled interview dates. To ensure that this attrition does not bias our results, we apply various sample restrictions. We re-estimate the effects of week 1 supervision on high-speed and low-speed workers for those who were absent for no more than 15, 10, or 5 days during the 30-day project; these thresholds correspond to the 95th, 90th, and 75th percentiles of the worker attrition distribution, respectively. Additionally, we consider a sample restriction where we limit the estimation to workers present during all five weeks of the survey project. In each of these cases, we find that the coefficient estimates remain consistent with our main results.

7. Mechanisms

In this section we explore the mechanisms by which supervision can most plausibly lead to persistent changes in worker performance. Exposure to supervision during the first week can influence worker performance along three distinct dimensions. First, supervision may have deterrence or incentive effects on worker behavior in line with the literature on government audits (Finan et al. 2017). This could transpire if, for example, prior supervision alters workers' expectations or knowledge of the firm's supervision practices. Section 7.1 examines the significance of incentive effects. Second, supervisors may affect worker performance by transmitting information regarding the firm's multi-faceted objectives to workers, rather than imparting "know how". This is explored in Section 7.2, which examines the role of improved clarity in driving performance improvements. Finally, supervisors might enhance worker performance by imparting tacit production knowledge to workers on-the-job. The potential role of knowledge flows is discussed in Section 7.3.

7.1. Incentives

We explore the influence of incentives or deterrence on worker performance using various features of our experiment design. First, we utilize the variation in supervision intensity across survey modules. This allows us to examine the performance of workers on M0 modules which, by design, were supervised at a lower intensity compared to M1 and M2 modules where supervisors were equipped with checklists. In Appendix Table A4, we present the impact of week 1 supervision on two key aspects: the total time required to complete these survey modules (Columns 1 and 3) and the deviation in primary outcome data derived from these modules (Columns 2 and 4). Columns 1 and 2 report the effects on high-speed workers, while Columns 3 and 4 report the effects on the performance of low-speed workers.

We find that, with a 1 SD increase in week 1 supervision, high-speed workers tend to produce higher-quality data resulting in a 0.27 SD decrease in data divergence. This improvement is not accompanied by a significant slowdown in the time required to complete these modules. Conversely, low-speed workers tend to administer these modules 6.4% faster, but do not record significantly lower quality data. If workers were motivated by the fear of sanctions and aimed to circumvent the supervision policy, we would anticipate a significant decrease in data quality in M0 modules as a result of gaming. However, our results indicate otherwise, suggesting that other forces may be driving the observed changes in worker performance. Instead, we observe that persistent *improvements* in worker performance spillover to the less intensely supervised M0 survey modules.

Second, as noted above, we also find evidence of improvements in the divergence index of data quality. Improvements on this index are indicative of enhanced uniformity in the survey data recorded by workers due to a shared interpretation of survey questions, data coding protocols, or interviewing techniques. Improvements on this index solely due to a deterrence motive is unlikely as the firm does *not* monitor performance on this outcome. Moreover, inputs related to this metric, such as whether a worker's adherence to the questionnaire script, are only observable when a supervisor is physically present to oversee an interview (and workers are aware of this).

Finally, we delve deeper into the possible role of incentives by investigating the interaction between contemporaneous and past supervision effects. Our aim is to discern whether workers who have been supervised more intensely in the first week internalize the firm's supervision policy. This would suggest a pattern of diminishing returns to *current* supervision in weeks 2 to 5. Furthermore, we would expect this effect to be particularly pronounced in interview modules that are monitored more intensely, such as the M1 (checklist) and M2 (checklist and report card) modules. Table A5 shows the interaction between supervision today and week 1 supervision separately for total interview time (column 1) and the time taken to administer survey module groups M0, M1 and M2 in Columns 2-4, where M0 modules were monitored at a lower intensity compared to M1 and M2 modules. We do not find any evidence of interactions between past and current supervision exposure in any of these cases.

7.2. Clarity

Moving forward, we delve into the impact of supervision on worker performance, with a specific focus on its role in enhancing clarity regarding the firm's priorities. In this context, we investigate how supervision influences the allocation of time by workers to survey modules that are classified as high priority for computing the crucial outcome variables of our study. In Table A6, we present the effects of current supervision, week 1 supervision, and their interaction on the distribution of time allocated to priority modules. We observe that both low and high-speed workers, as indicated in columns 1 and 2, allocate between 1.7% and 2.7% more time to priority survey modules when they are being supervised on the same day. This outcome is unsurprising, given that the modules subject to intense monitoring (M1 and M2) by supervisors are, in fact, a subset of the firm's high-priority list.

Of particular note are the persistent effects of past supervision on the time allocation by workers across various interview modules. We find that the reallocation effects are more pronounced among low-speed workers, who are relatively new to the firm's project priorities due to their limited task experience. Low-speed workers consistently allocate a higher share of their time (1.4%) to priority modules while maintaining an overall increase in efficiency. On the other hand, high-speed workers do not significantly alter their time allocation to priority modules in response to week 1 supervision, while slowing down their overall interview speed. This suggests that the patterns of slower interview speeds observed among high-speed workers are not primarily driven by a reallocation of time between high and low-priority segments of the questionnaire.

7.3. Knowledge Flows

We investigate the role of supervisors in imparting tacit production knowledge to workers using three pieces of additional evidence. First, we use a feature of our experiment to verify that supervisors *process information* on worker performance and *tailor* their advice. To examine this, we consider the persistent effect of supervision in two cases: M1 (checklists) and M2 (checklists and performance reports) survey modules, as depicted in Appendix Figure A9. This comparison enables us to discern whether *workers* respond differently when reports on their speed performance are provided to their *supervisors*.

Our findings confirm that supervisors tailor their guidance to workers who underperform on the speed margin when equipped with quantitative evidence concerning their speed. Indeed, the effects of supervision on low-speed workers' production speeds are primarily driven by the M2 survey modules in which supervisors were equipped with performance reports. Moreover, supervisors do *not* appear to advise high-speed workers any differently across the M1 and M2 survey modules. In both cases, high speed workers exhibit a persistent slowdown in module times in response to past supervision (involving checklists to encourage a focus on output quality).

Second, we present descriptive evidence from worker reports detailing the topics of conversation with supervisors during the first week of the project in Appendix Figure A10. Each bar represents the likelihood of a worker selecting a particular option from a multiple-choice question about conversation topics. Instances where a worker did not receive any advice are coded as 0. We find that low-speed workers with limited task experience were more likely to receive advice on all topics, including time management. Moreover, a substantial portion of the advice-given centered on topics related to essential data collection skills such as interpreting survey questions or effectively probing respondents for accurate information.

Third, we simply ask workers whom they mainly turn to for advice on the problems they encounter at work. Appendix Figure A11 shows the distribution of workers' responses (which we collected prior to the experiment). Workers consistently report that supervisors are their main source of advice on how to solve problems related to interview respondents, field situations and survey questions. This evidence is consistent with the importance of supervisor-worker interactions for knowledge sharing.

8. Conclusion

Real organizations routinely accumulate and deploy knowledge generated in their daily operations (Gibbons and Prusak 2020). Nevertheless, organizational knowledge is tricky to study empirically as it is often *tacit*, making it hard to describe and measure. In this paper, we study the role of frontline supervisors in transmitting tacit production knowledge to workers who conduct household surveys for a research organization in Uganda. We show that the effects of supervision support the notion that supervisors coach workers on-the-job.

We uncover three important features of the causal effects of supervision on worker performance. The effects are persistent, tailored to individual worker attributes, and economically meaningful. Workers tend to excel most in performance dimensions where they were initially predicted to be weakest, and these improvements persist even on unsupervised days, highlighting the cost-effectiveness of supervision for firms. These patterns shed light on *what supervisors do* to improve worker performance.

The view that supervisors are conduits for the transmission of production knowledge in organizations has clear implications for personnel policies. First, effective supervisors require a deep understanding of the industry, the specific firm, and the tasks they oversee. Tacit knowledge is a cornerstone for delivering tailored coaching. Second, organizations should design career paths that nurture supervisors' coaching skills. It may be a mistake to reward supervisors who foster talent with promotions to strategy roles that do not harness these skills. This is a specific example of the Peter Principle, highlighting a mismatch in the skills set required to be an effective frontline supervisor and a higher-level manager (Benson et al. 2019). Finally, our results also underscore the significance of early coaching for new employees. Initiating on-the-job coaching from the onset of their tenure may yield substantial improvements in worker performance.

9. Figures

FIGURE 1. Sample Interview Process

A. An Agricultural Plot with Intercropping

B. Sample Plot-and-Crop Roster to Record Output

Harvest > Plot (1) — **Repeat Group Over Plots**

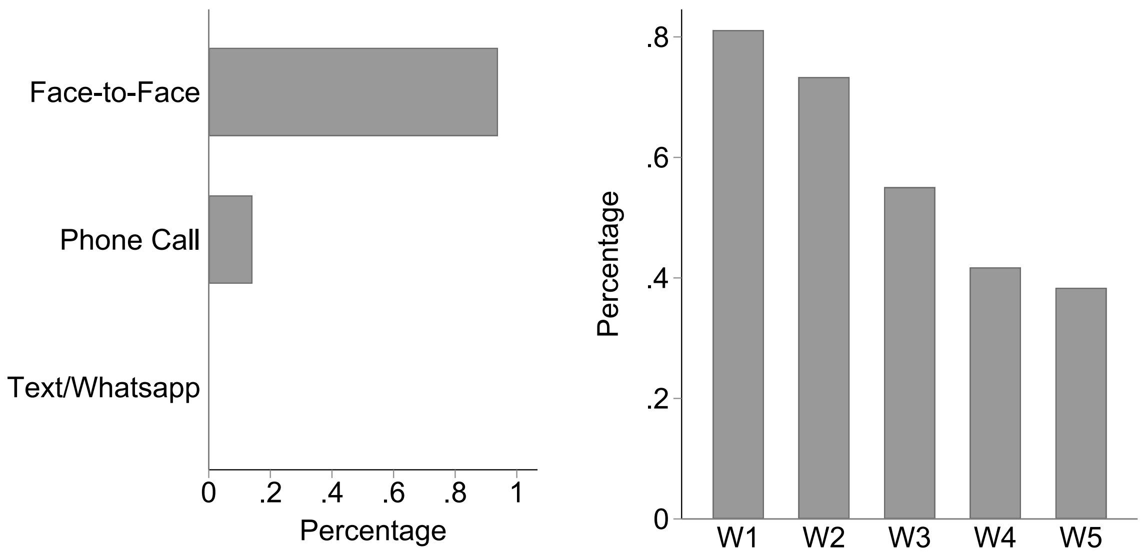
Harvest > Plot (1) > Crop (1) — **Repeat Group Over Crops on Plot**

Sample Question Text	**Record Type**
How many [units] of [crop name] in total did your household harvest from [plot name] during the Jan-June 2022 Agricultural Season?	Numeric Entry
While still in the field, was any of [farmer name]'s [crop name] from [plot name] damaged by pests or disease?	Select yes/no
How many [units] of [crop name] were lost?	Numeric Entry

Notes: Panel A shows a picture of an agricultural plot in Uganda which was intercropped with cassava (top), maize (middle) and beans (bottom). Panel B shows that each interviewer was required to record crop output at the plot-and-crop level, using 'repeat groups', commonly known as survey rosters.

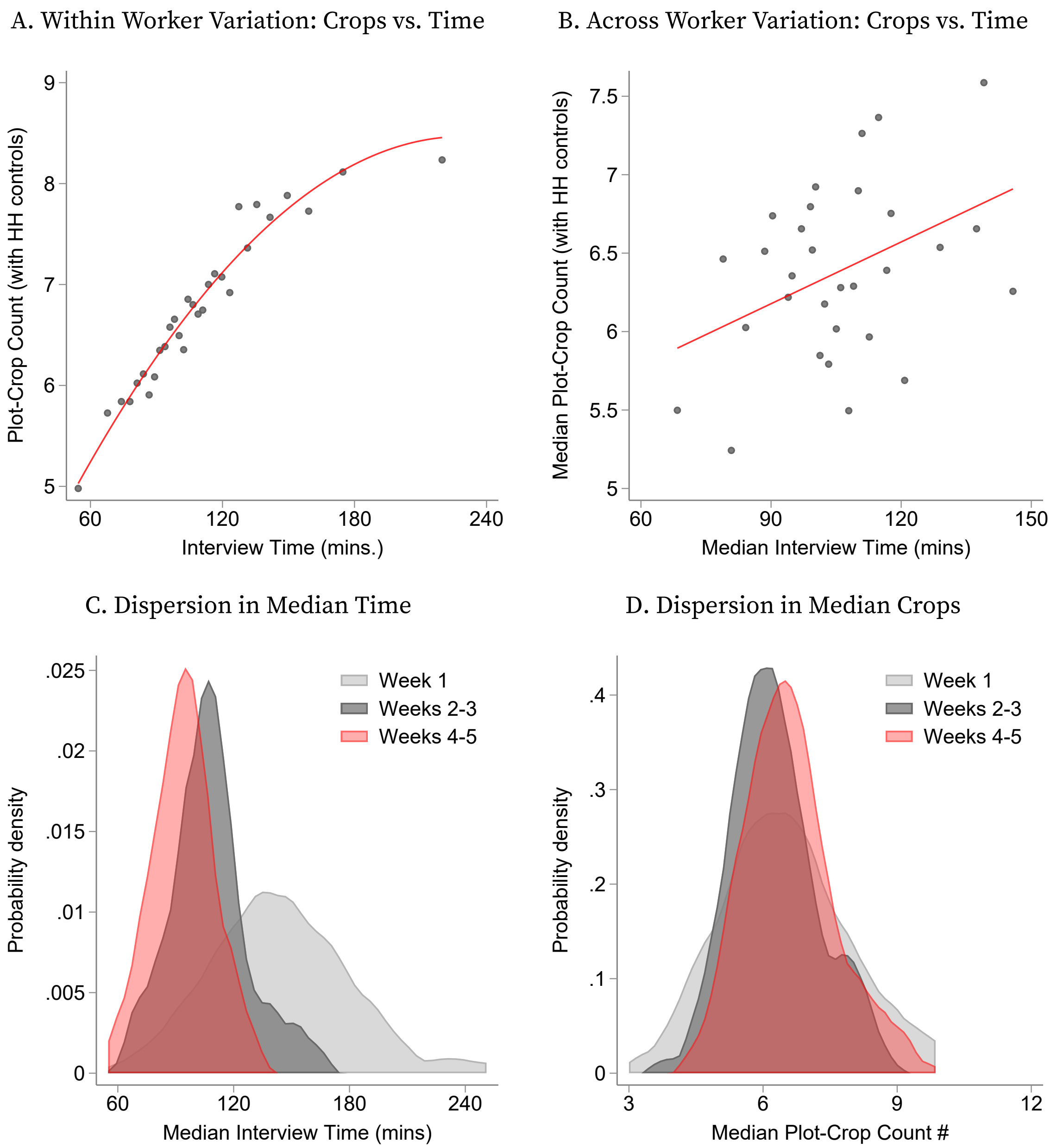
FIGURE 2. Patterns in Workplace Communication

A. Modes of Communication B. Production Tips by Week



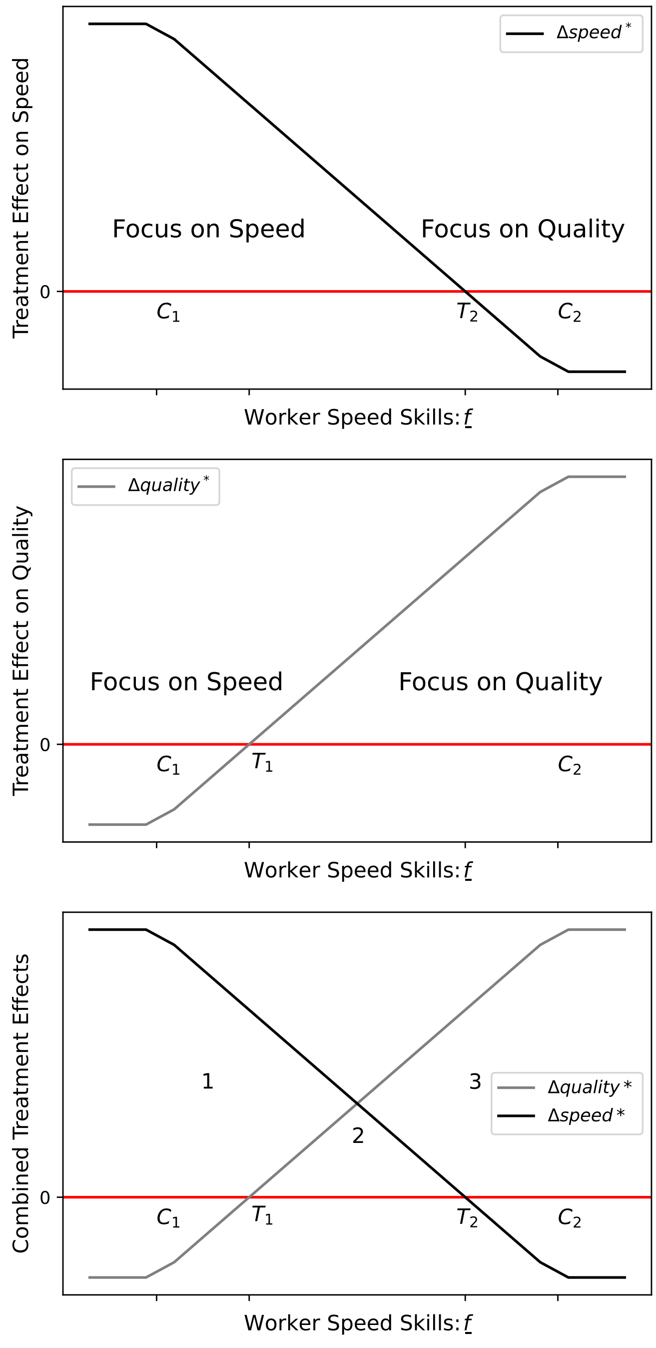
Notes: Panel A shows patterns in the mode of communication between supervisors and workers upon the completion of a spot check. This data is based on brief post-interview reports submitted by workers. Panel B shows supervisors' self-reported assessments of whether they could share any tips, following a spot check, to help workers improve on their performance. The probability that a supervisor could share production tips/tricks is plotted by project week. The data reported in both Panels A and B are conditional on the conduct of a spot check.

FIGURE 3. Patterns in Speed and Quality Outcomes



Notes: These figures show three empirical patterns observed in our data. Panel A shows displays a binned scatter plot, showing the positive correlation between interview time (y-axis) and the plot-and-crop count recorded by a worker (y-axis), after controlling for worker fixed effects and household traits (e.g. plot-crops cultivated last season, household wealth). Panel B shows the correlation between the median plots-and-crops (x-axis) and interview time (y-xis) across workers in our sample. Panels C and D plot the density of worker-level median interview times and plots-and-crops, respectively. The density plots are distinguished by weeks 1 (light gray), weeks 2-3 (dark dray) and weeks 4-5 (red) of the project cycle. See Section 3.4 in the main paper for further details.

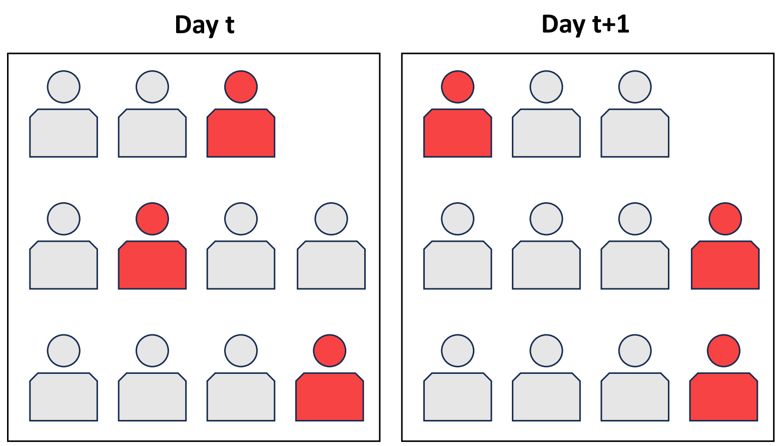
FIGURE 4. Summary of Empirical Predictions



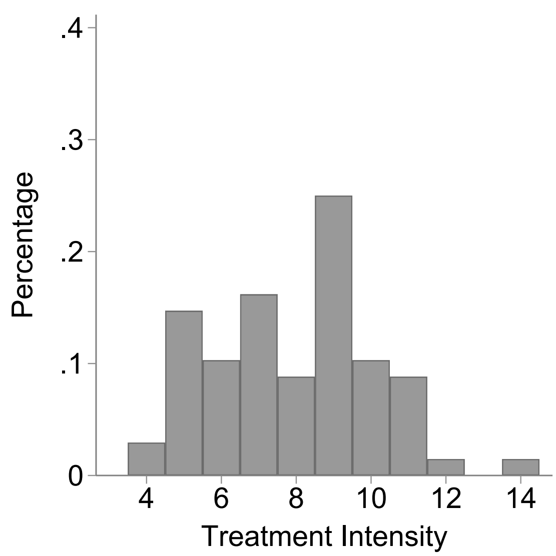
Notes: These figures depict how coaching impacts worker speed and quality based on their innate speed skills (x-axis). Panel A shows that as innate speed increases, the predicted change in worker speed decreases. Panel B shows that the predicted change in output quality increases as innate speed goes up. Panel C combines these predictions and divides workers into three regions: Region 1 - low-speed workers improve speed but reduce quality, Region 2 - low-speed workers improve both speed and quality, and Region 3 - relatively high-speed workers focus on quality but slow down.

FIGURE 5. Overview of Experiment Design

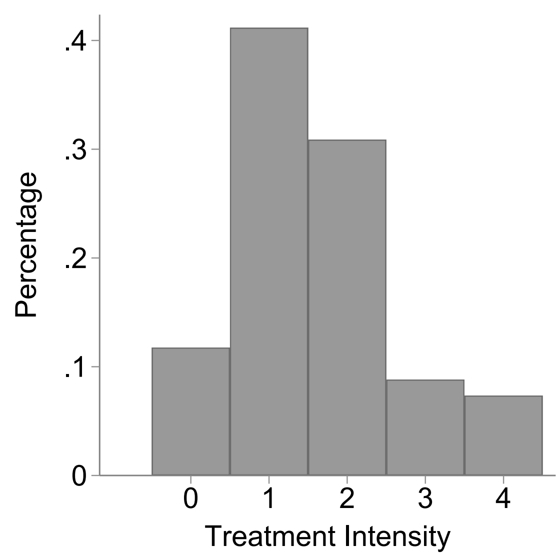
A. Assignment of Spot Checks



B. Exposure to Spot Checks: All Weeks



C. Exposure to Spot Checks: 1st Week

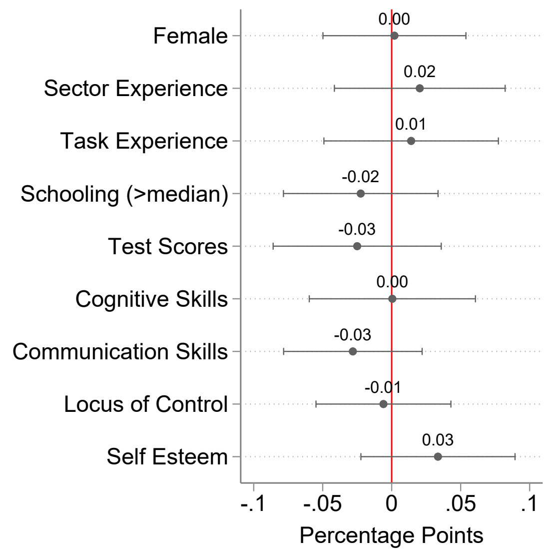


Notes: Panel A illustrates key features of our experimental design. Each box represents a worker team, where each row corresponds to a block of workers who were matched based on covariates compiled from a worker survey (cognitive skills, communication skills, recruitment test scores). The color red is used to highlight that workers were assigned to treatment on a daily basis, where we sampled with replacement. That is, a worker could be assigned to treatment on day t, but not on day $t+1$. They could also have been assigned to treatment on consecutive days. Panels B and C show the resulting variation in cumulative (i.e. total) monitoring intensity over the project cycle, and in the first week of the project, respectively. Further details on the experimental design are provided in Section 5.1 in the main text.

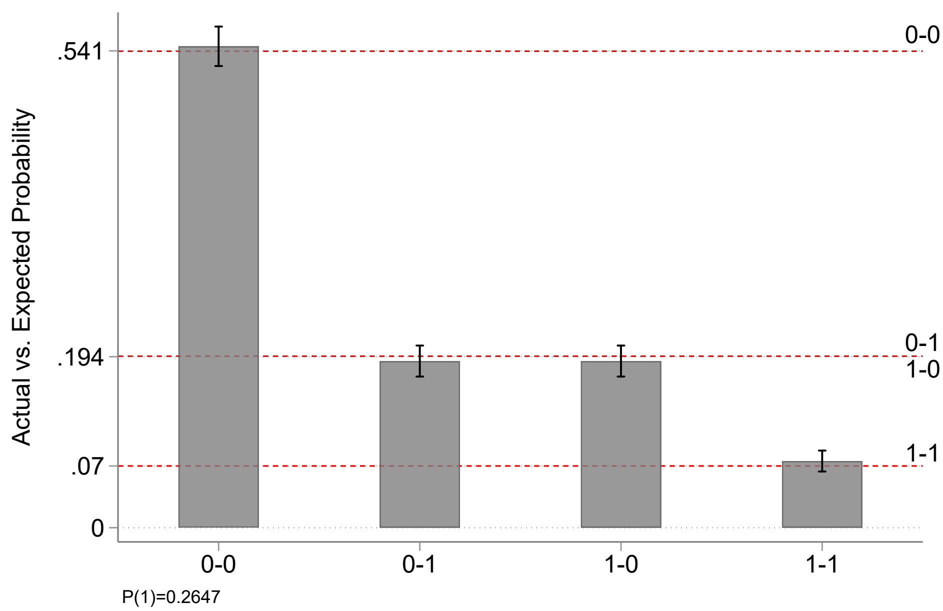
FIGURE 6. Experiment Implementation Checks

A. Balance: All Weeks

B. Balance: Week 1

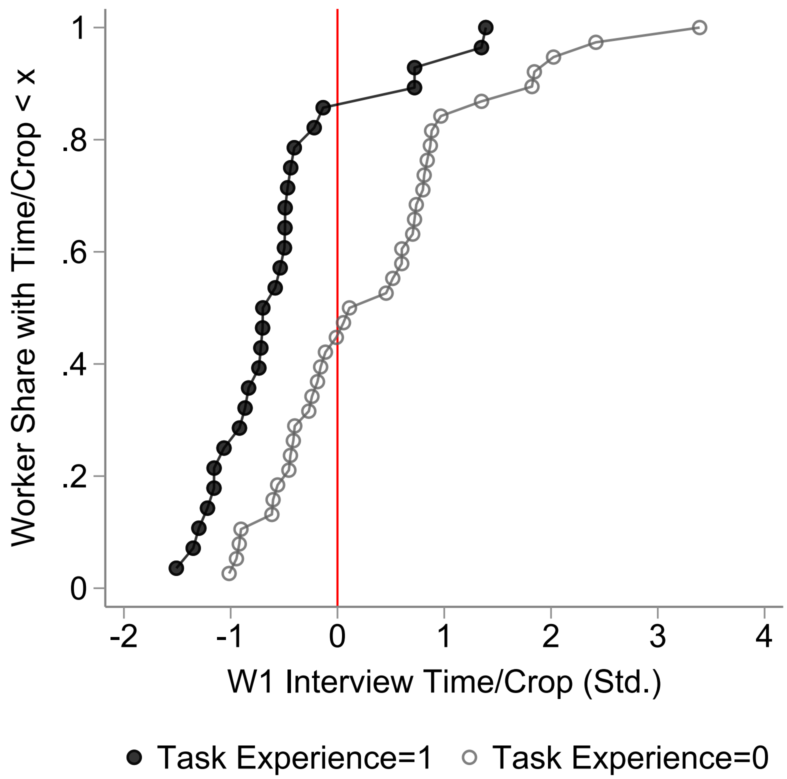


C. Independence of Treatment Assignments Over Time



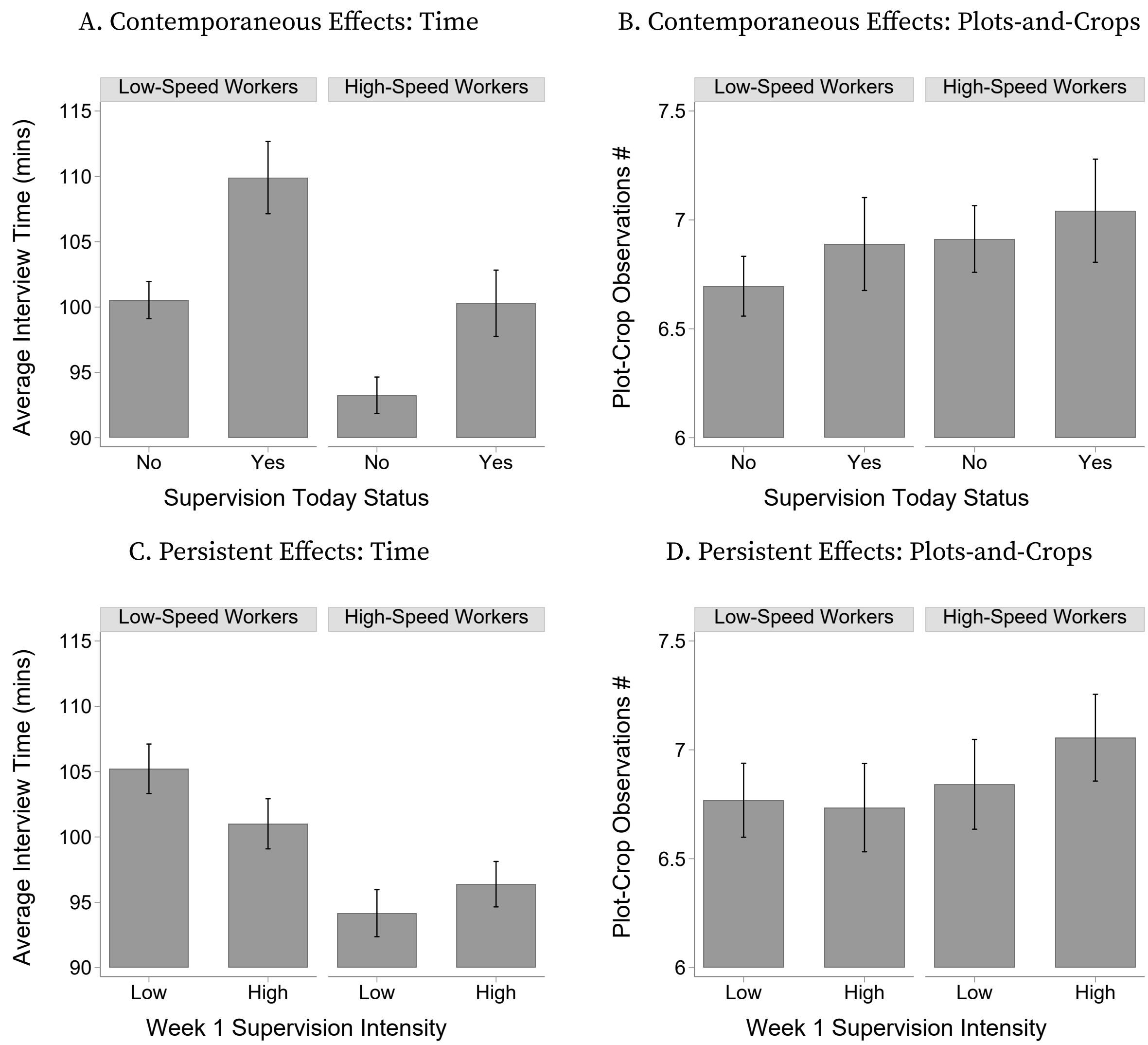
Notes: Panels A and B display coefficients from balance tests corresponding to spot check assignments in all weeks and the first week. Regressions are of the form $D_{it} = \alpha + \beta X_i + \epsilon_{it}$ where observations are at the worker i and day t level, D_{it} denotes assignment to treatment and X_i denotes a vector of standardized worker characteristics. Standard errors are clustered at the worker level. P-values corresponding to F-tests for the significance of all worker characteristics are 0.46 (Panel A) and 0.740 (Panel B). All columns include strata fixed effects. Panel C plots the expected probability of various binary treatment assignment combinations in two successive days under independent assignment (0-0, 0-1, 1-0, 1-1) in red lines. The bars indicate the realized probability of assignment in our data.

FIGURE 7. Patterns in Interview Time per Plot-and-Crop, by Task Experience



Notes: This Figure plots the cumulative distribution of median interview time per plot-and-crop across workers during the first week of the project. The distributions are plotted separately for workers with and without prior task experience. Interview time per plot-crop shown on the x-axis has been standardized. Consequently, 0 corresponds to the average time per plot-crop across workers during the first week.

FIGURE 8. Summary of Supervision Effects, by Initial Production Speed



Panel A reports the average interview time (in minutes) by worker speed and supervision today status. Panel B reports the average number of plot and crop observations by worker speed and supervision today status. Panel C reports the average interview time (in minutes) by worker speed and week 1 supervision intensity. Panel D reports the average number of plot and crop observations by worker speed and week 1 supervision intensity. Observations correspond to Weeks 2-5 of the project cycle. The reported interview time and plot-crop measures include controls for household and worker traits, and team-day fixed effects. See Equation 11 for details regarding each of these controls. Contemporaneous Effects reported in Panels A and B include controls for the intensity of prior supervision. Persistent Effects in Panels C and D are computed over non-supervised worker days in Weeks 2-5.

10. Tables

TABLE 1. Employee Characteristics

	Mean	SD
Demographics		
Age (Years)	30.3	4.8
Female (0/1)	0.65	0.48
Experience		
Experience in Surveyor Roles (Months)	39.0	34.6
Experience at the Parent Organization (Months)	23.8	32.7
Prior Task Experience (0/1)	0.43	0.50
Education		
Years of Schooling	16.5	2.1
Recruitment Test Scores (%)	64.5	9.3
Cognitive Skills		
Forward Digit-Span Recall	4.8	1.3
Backward Digit-Span Recall	4.3	1.2
Arithmetic Score (%)	62.1	20.9
10-Item Raven's Matrices Score (%)	59.5	26.6
Communication Skills		
Uniformity (1-5 scale)	4.2	0.51
Intonation (1-5 scale)	2.8	0.71
Presence (1-5 scale)	4.0	0.61
Effectiveness (1-5 scale)	4.2	0.64
Personality		
Locus of Control (1-5 scale)	3.3	0.70
Self Esteem (1-5 scale)	4.3	0.50

Notes: This table presents descriptive statistics of workers' cognitive and non-cognitive skills. We collected this data through a worker survey during a pre-training (onboarding) period prior to the commencement of household interviews.

TABLE 2. Overview of Performance Outcomes

A. Measurement Frequency	N				
No. of households interviewed	3,330				
No. of survey modules per interview	20				
Median interviews per worker	50				
Median interview days per worker	27				

B. Negative Outcome Variables	Observation Level	Worker-Level Statistics: W1			
		Mean	SD	p25	p75
Speed					
Interview Time (mins.)	interview	146.7	34.0	127.5	173.6
Priority Modules Time (mins.)	interview	81.3	15.2	71.5	91.1
Quality: Outcomes					
Plot-and-Crop Observations # *	interview	6.4	1.2	5.4	7.1
Divergence Index (Std.)	3-day bins	0.00	1.00	-0.66	0.70
Quality: Process					
Deviate from Interview Order (0/1)	interview	0.15	0.16	0.00	0.22
Substitute for the Household Head (0/1)	interview	0.25	0.16	0.14	0.38
Use Missing Data Codes: Above Median (0/1)	interview	0.51	0.27	0.30	0.71

Notes: Panel A describes the measurement frequency of outcomes over the project cycle (30 days). Panel B lists the key outcome variables (in Column 1), and the level of observation pertaining to each outcome (in Column 2). The distribution of performance outcomes across workers during Week 1 of the project is shown in Columns 3 to 6, which report the mean, standard deviation, the 25th percentile and the 75th percentile of the worker-level outcome distribution, respectively. A description of each outcome measure is provided in the main text (see Section 3). Notice that the outcome measures are 'negative outcomes', e.g. a higher interview time implies lower speed. *We consider the *negative* value of the plot-and-crop observations reported by a worker (in logs) in all our regression specifications, as a proxy for the incidence of shirking.

TABLE 3. Compliance with Assigned Supervision

	Actual Supervision			
	(1)	(2)	(3)	(4)
VARIABLES	w1 supervision (std.)	supervise today=1	presence=1	advice=1
Supervise Today=1	0.006	0.613***	0.491***	0.296***
	(0.043)	(0.031)	(0.031)	(0.033)
W1 Supervision (Std.)	0.459***	0.004	-0.003	0.003
	(0.096)	(0.009)	(0.009)	(0.011)
Observations	1,632	1,632	1,632	1,632
Adjusted R^2	0.212	0.492	0.387	0.214
Control Mean	-0.03	0.04	0.02	0.05
F-Stat: Supervise Today	0.02	384.67	253.87	80.36
F-Stat: W1 Supervision	22.92	0.16	0.15	0.06

Notes: Observations are at the worker and day level. This table shows the association between assigned and actual supervision through weeks 2 to 5 of the survey project. The dependent variables are: The actual stock of week 1 supervision (column 1), actual supervision today (column 2), a dummy variable coded to 1 if a supervisor was present for most of the time for at least one interview on that day, and 0 otherwise (column 3), a dummy variable coded to 1 if a supervisor was present for at least one interview that day and provided advice to the worker, and 0 otherwise. The independent variables are the assigned supervision status today (row 1) and the assigned stock of week 1 supervision (row 2). All columns incorporate team and day fixed effects, with standard errors clustered by worker shown in parentheses.

TABLE 4. The Effect of Supervision Today: All Workers

	Speed		*Quality: Outcomes*		*Quality: Process*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	log time	log time(p)	-log crop	divergence	deviate	substitute	use-code
Supervise Today (0/1)	0.113***	0.141***	-0.042***	-0.102	-0.055***	-0.015	0.020
	(0.014)	(0.016)	(0.015)	(0.129)	(0.014)	(0.021)	(0.022)
Observations	2,632	2,632	2,625	400	2,632	2,632	2,632
Adjusted R^2	0.333	0.292	0.390	0.133	0.073	0.124	0.096
Control Mean	100.16	64.12	6.77	-0.07	0.13	0.28	0.48
Control SD	30.24	20.40	3.23	1.01	0.34	0.45	0.50
Bootstrap p-value	0.000	0.000	0.010	0.430	0.000	0.509	0.394

Notes: Observations are at the interview level (except column 4 as the Divergence Index is computed at 3-day bin level). Estimation sample: Weeks 2 to 5, all workers and days. The regression specification shown in Equation 11 is modified to include the interaction of Supervision Today (0/1) and Week 1 Supervision (Std.). Hence, the reported effect of Supervision Today (0/1) corresponds to the average exposure to Week 1 Supervision (Std.). The dependent variables are: log interview time (column 1), log time on priority survey modules (column 2), - log of the number of plot-crop combinations recorded (column 3), a standardized index of divergence across outcome data (column 4) and dummy variables to capture deviations from interview order (column 5), interview of a substitute respondent other than the household head (column 6), and above median usage of missing data codes (column 7). All columns include controls for household characteristics, worker characteristics (dummies for above median traits) interacted with day fixed effects, and team-day fixed effects. In addition, all control variables and fixed-effects are interacted with a dummy for workers' prior task experience. Standard errors clustered by worker are reported in parentheses. We also report p-values computed using cluster bootstrap standard errors at the worker level in Row 6.

TABLE 5. Persistent Effects of Week 1 Supervision: All Workers

	Speed		*Quality: Outcomes*		*Quality: Process*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	log time	log time(p)	-log crop	divergence	deviate	substitute	use-code
W1 Supervision (Std.)	-0.024	-0.018	-0.030*	-0.108**	-0.041*	0.013	0.030
	(0.017)	(0.017)	(0.017)	(0.053)	(0.022)	(0.014)	(0.023)
Observations	1,853	1,853	1,847	400	1,853	1,853	1,853
Adjusted R^2	0.393	0.297	0.408	0.136	0.112	0.121	0.107
Control Mean	100.16	64.12	6.77	-0.07	0.13	0.28	0.48
Control SD	30.24	20.40	3.23	1.01	0.34	0.45	0.50
Bootstrap p-value	0.323	0.415	0.169	0.052	0.113	0.439	0.318

Notes: Observations are at the interview level (except column 4). Estimation sample: Weeks 2 to 5, all workers, and non-supervised days only (i.e. $D_{it} = 0$). The estimates in Column 4 are computed over all periods regardless of current supervision status, as the Divergence Index is measured at a 3-day bin level. The dependent variables are: log interview time (column 1), log time on priority survey modules (column 2), - log of the number of plot-crop combinations recorded (column 3), a standardized index of divergence across outcome data (column 4) and dummy variables to capture deviations from interview order (column 5), interview of a substitute respondent other than the household head (column 6), and above median usage of missing data codes (column 7). All columns include controls for household characteristics, worker characteristics (dummies for above median traits) interacted with day fixed effects, and team-day fixed effects. In addition, all control variables and fixed-effects are interacted with a dummy for workers' prior task experience. Standard errors clustered by worker are reported in parentheses. We also report p-values computed using cluster bootstrap standard errors at the worker level in Row 6.

Table 6. Persistent Effects of Week 1 Supervision: Low-Speed Workers

	Speed		*Quality: Outcomes*		*Quality: Process*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	log time	log time(p)	-log crop	divergence	deviate	substitute	use-code
W1 Supervision (Std.)	-0.070***	-0.056**	-0.012	-0.015	-0.041	0.037**	0.003
	(0.018)	(0.024)	(0.025)	(0.078)	(0.029)	(0.015)	(0.030)
Observations	1,091	1,091	1,087	234	1,091	1,091	1,091
Adjusted R^2	0.359	0.263	0.409	0.142	0.068	0.125	0.058
Control Mean	104.58	66.15	6.70	-0.12	0.11	0.25	0.50
Control SD	32.55	21.88	3.27	0.99	0.31	0.44	0.50
Bootstrap p-value	0.010	0.067	0.708	0.823	0.329	0.058	0.924

Notes: Observations are at the interview level (except Column 4). Estimation sample: weeks 2 to 5, low-speed workers, and non-supervised days only (i.e. $D_{it} = 0$). The estimates in Column 4 are computed over all periods regardless of current supervision status, as the Divergence Index is measured at a 3-day bin level. The dependent variables are: log interview time (column 1), log time on priority survey modules (column 2), - log of the number of plot-crop combinations recorded (column 3), a standardized index of divergence across outcome data (column 4) and dummy variables to capture deviations from interview order (column 5), interview of a substitute respondent other than the household head (column 6), and above median usage of missing data codes (column 7). All columns include controls for household characteristics, worker characteristics (dummies for above median traits) interacted with day fixed effects and team-day fixed effects. Standard errors clustered by worker shown in parentheses. We also report p-values computed using cluster bootstrap standard errors at the worker level in Row 6.

TABLE 7. Persistent Effects of Week 1 Supervision: High-Speed Workers

	Speed		*Quality: Outcomes*		*Quality: Process*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	log time	log time(p)	-log crop	divergence	deviate	substitute	use-code
W1 Supervision (Std.)	0.038**	0.032**	-0.054**	-0.277**	-0.042	-0.019	0.066*
	(0.017)	(0.013)	(0.020)	(0.103)	(0.032)	(0.020)	(0.034)
Observations	761	761	759	166	761	761	761
Adjusted R^2	0.478	0.374	0.409	0.076	0.163	0.112	0.200
Control Mean	93.07	60.87	6.87	0.02	0.17	0.31	0.45
Control SD	24.54	17.32	3.18	1.07	0.37	0.46	0.50
Bootstrap p-value	0.183	0.069	0.085	0.018	0.250	0.441	0.215

Notes: Observations are at the interview level (except Column 4). Estimation sample: weeks 2 to 5, high-speed workers, and non-supervised days only (i.e. $D_{it} = 0$). The estimates in Column 4 are computed over all periods regardless of current supervision status, as the Divergence Index is measured at a 3-day bin level. The dependent variables are: log interview time (column 1), log time on priority survey modules (column 2), - log of the number of plot-crop combinations recorded (column 3), a standardized index of divergence across outcome data (column 4) and dummy variables to capture deviations from interview order (column 5), interview of a substitute respondent other than the household head (column 6), and above median usage of missing data codes (column 7). All columns include controls for household characteristics, worker characteristics (dummies for above median traits) interacted with day fixed effects and team-day fixed effects. Standard errors clustered by worker shown in parentheses. We also report p-values computed using cluster bootstrap standard errors at the worker level in Row 6.

References

Adhvaryu, A., A. Nyshadham, and J. Tamayo (2022): "An anatomy of performance monitoring," *mimeo*.

——— (2023): "Managerial quality and productivity dynamics," *The Review of Economic Studies*, 90, 1569–1607.

Alchian, A. A. and H. Demsetz (1972): "Production, information costs, and economic organization," *The American economic review*, 62, 777–795.

Bandiera, O., I. Barankay, and I. Rasul (2007): "Incentives for managers and inequality among workers: Evidence from a firm-level experiment," *The Quarterly Journal of Economics*, 122, 729–773.

Becker, G. S. (1962): "Investment in human capital: A theoretical analysis," *Journal of political economy*, 70, 9–49.

Benkard, C. L. (2000): "Learning and forgetting: The dynamics of aircraft production," *American Economic Review*, 90, 1034–1054.

Benson, A., D. Li, and K. Shue (2019): "Promotions and the peter principle," *The Quarterly Journal of Economics*, 134, 2085–2134.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013): "Does management matter? Evidence from India," *The Quarterly Journal of Economics*, 128, 1–51.

Bloom, N. and J. Van Reenen (2007): "Measuring and explaining management practices across firms and countries," *The Quarterly Journal of Economics*, 122, 1351–1408.

BLS (2022): "Occupation and Employment Wage Statistics, US Bureau of Labor Statistics," https://www.bls.gov/oes/tables.htm, accessed: 2023-10-30.

Bruhn, M., D. Karlan, and A. Schoar (2018): "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico," *Journal of Political Economy*, 126, 635–687.

Carletto, C., A. Dillon, and A. Zezza (2021): "Chapter 81 - Agricultural data collection to minimize measurement error and maximize coverage," in *Handbook of Agricultural Economics*, ed. by C. B. Barrett and D. R. Just, Elsevier, vol. 5 of *Handbook of Agricultural Economics*, 4407–4480.

Chan, T. Y., J. Li, and L. Pierce (2014): "Learning from peers: Knowledge transfer and sales force productivity growth," *Marketing Science*, 33, 463–484.

Cullen, Z. and R. Perez-Truglia (2023): "The Old Boys' Club: Schmoozing and the Gender Gap," *American Economic Review*, 113, 1703–40.

Czura, K., F. Englmaier, H. Ho, and L. Spantig (2023): "Employee Performance and Mental Well-Being: The Mitigating Effects of Transformational Leadership during Crisis," *CESifo Working Paper*.

De Philippis, M. (2021): "Multi-task agents and incentives: The case of teaching and research for university professors," *The Economic Journal*, 131, 1643–1681.

De Weerdt, J., J. Gibson, and K. Beegle (2020): "What can we learn from experimenting with survey methods?" *Annual Review of Resource Economics*, 12, 431–447.

Dillon, A., D. Karlan, C. Udry, and J. Zinman (2020): "Good identification, meet good data," *World Development*, 127, 104796.

Emanuel, N., E. Harrington, and A. Pallais (2023): "The Power of Proximity to Coworkers: Training for Tomorrow or Productivity Today?" *Working Paper*.

Finan, F., B. Olken, and R. Pande (2017): "The Personnel Economics of the Developing State," in *Handbook of Economic Field Experiments*, ed. by A. V. Banerjee and E. Duflo, North-Holland, vol. 2 of *Handbook of Economic Field Experiments*, 467–514.

Frederiksen, A., L. B. Kahn, and F. Lange (2020): "Supervisors and Performance Management Systems," *Journal of Political Economy*, 128, 2123–2187.

Friebel, G., M. Heinz, and N. Zubanov (2022): "Middle managers, personnel turnover, and performance: A long-term field experiment in a retail chain," *Management Science*, 68, 211–229.

Friebel, G. and M. Raith (2022): "Talent management: The role of bosses," *Available at SSRN 4162631*.

Fryer, R. and R. Holden (2013): "Multitasking, dynamic complementarities, and incentives: A cautionary tale," *Cambridge, MA: Harvard University*.

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

Garicano, L. and E. Rossi-Hansberg (2006): "Organization and inequality in a knowledge economy," *The Quarterly journal of economics*, 121, 1383–1435.

Gibbons, R. and L. Prusak (2020): "Knowledge, stories, and culture in organizations," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 110, 187–192.

Gibbons, R. and M. Waldman (2004): "Task-specific human capital," *American Economic Review*, 94, 203–207.

Gosnell, G. K., J. A. List, and R. D. Metcalfe (2020): "The impact of management practices on employee productivity: A field experiment with airline captains," *Journal of Political Economy*, 128, 1195–1233.

Haegele, I. (2024): "Talent hoarding in organizations," *arXiv preprint:2206.15098v2*.

Hoffman, M. and S. Tadelis (2021): "People management skills, employee attrition, and manager rewards: An empirical analysis," *Journal of Political Economy*, 129, 243–285.

Holmström, B. and P. Milgrom (1991): "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design," *The Journal of Law, Economics, and Organization*, 7, 24–52.

Hong, F., T. Hossain, J. A. List, and M. Tanaka (2018): "Testing the theory of multitasking: Evidence from a natural field experiment in Chinese factories," *International Economic Review*, 59, 511–536.

Hubbard, T. N. (2000): "The demand for monitoring technologies: The case of trucking," *The Quarterly Journal of Economics*, 115, 533–560.

——— (2003): "Information, decisions, and productivity: On-board computers and capacity utilization in trucking," *American Economic Review*, 93, 1328–1353.

Jensen, N., E. Lyons, E. Chebelyon, R. Le Bras, and C. Gomes (2020): "Conspicuous monitoring and remote work," *Journal of Economic Behavior & Organization*, 176, 489–511.

Kelley, E. M., G. Lane, and D. Schönholzer (2023): "Monitoring in Small Firms: Experimental Evidence from Kenyan Public Transit," Tech. rep., Working paper.

Laajaj, R. and K. Macours (2021): "Measuring skills in developing countries," *Journal of Human resources*, 56, 1254–1295.

Lazear, E. P., K. L. Shaw, and C. T. Stanton (2015): "The value of bosses," *Journal of Labor Economics*, 33, 823–861.

Levitt, S. D., J. A. List, and C. Syverson (2013): "Toward an understanding of learning by doing: Evidence from an automobile assembly plant," *Journal of Political Economy*, 121, 643–681.

Menzel, A. (2021): "Knowledge exchange and productivity spill-overs in Bangladeshi garment factories," *Journal of Economic Behavior & Organization*, 185, 721–746.

Metcalfe, R. D., A. B. Sollaci, and C. Syverson (2023): "Managers and Productivity in Retail," Working Paper 31192, National Bureau of Economic Research.

Minni, V. (2023): "Making the invisible hand visible: Managers and the allocation of workers to jobs," Tech. rep., Centre for Economic Performance, LSE.

Mookherjee, D. (2013): "Incentives in hierarchies," *The handbook of organizational economics*, 764–798.

Nagin, D. S., J. B. Rebitzer, S. Sanders, and L. J. Taylor (2002): "Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment," *American Economic Review*, 92, 850–873.

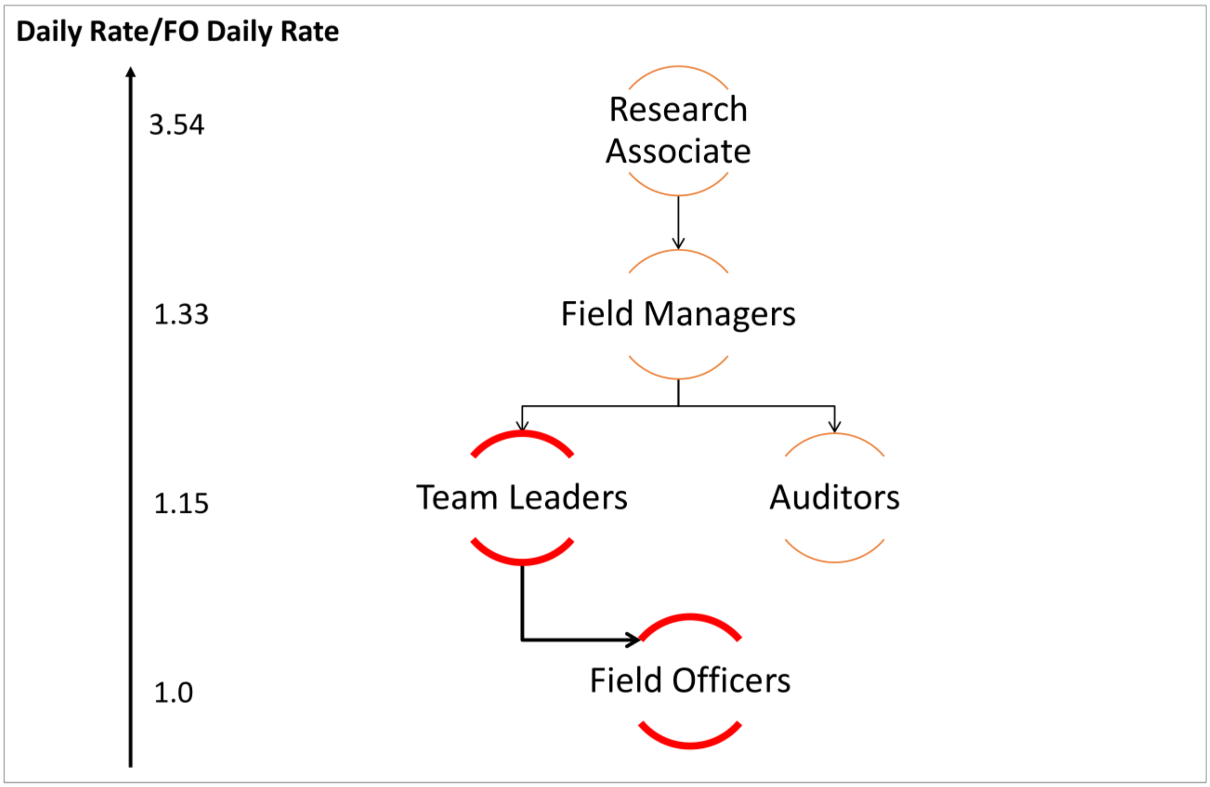
Polanyi, M. (1962): "Personal Knowledge: Towards a Post-Critical Philosophy," *University of Chicago Press*.

Sandvik, J. J., R. E. Saouma, N. T. Seegert, and C. T. Stanton (2020): "Workplace knowledge flows," *The Quarterly Journal of Economics*, 135, 1635–1680.

UBOS (2017): "The National Population and Housing Census 2014 – Education in the Thematic Report Series," *Uganda Bureau of Statistics, Kampala, Uganda, November 2017*.

Appendix A. Additional Figures

FIGURE A1. Organizational Structure



Notes: This Figure shows the organizational structure observed in the data collection firm. Our study focuses on the employment relationship between Team Leaders (supervisors) and Field Officers (workers). The figures on the left-hand side represent the pay-scale gradient for each organizational tier, relative to the earnings of Field Officers in Tier 1. For example, Team Leaders earn 15% more than the Field Officers whom they supervise.

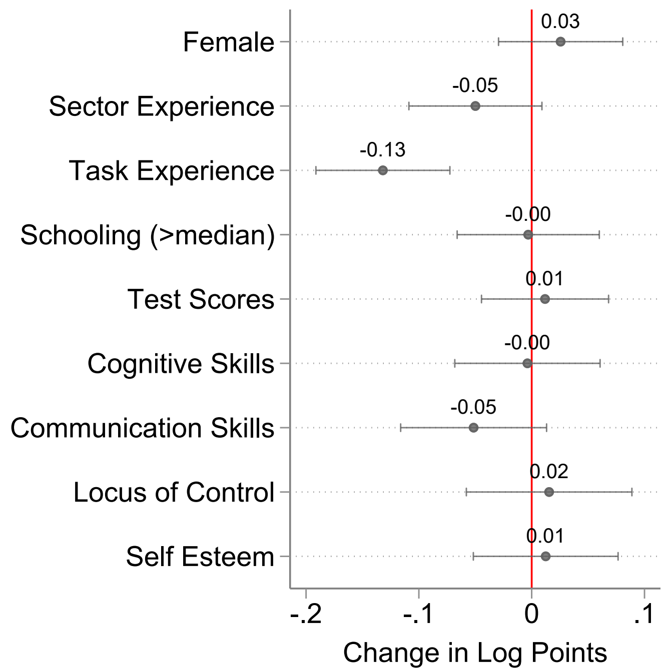
FIGURE A2. Exhibit - Checklists

Module: Crop Sales and Disposition
1. Does the FO rush through the backcheck (confirming total household output of each crop) or give the respondent time to think?
2. Do they make an effort to ensure that the numbers add up or make sense? E.g. quantity sold cannot exceed quantity produced
3. Are the price estimates per unit of crop sold reasonable based on what you have seen/heard so far?

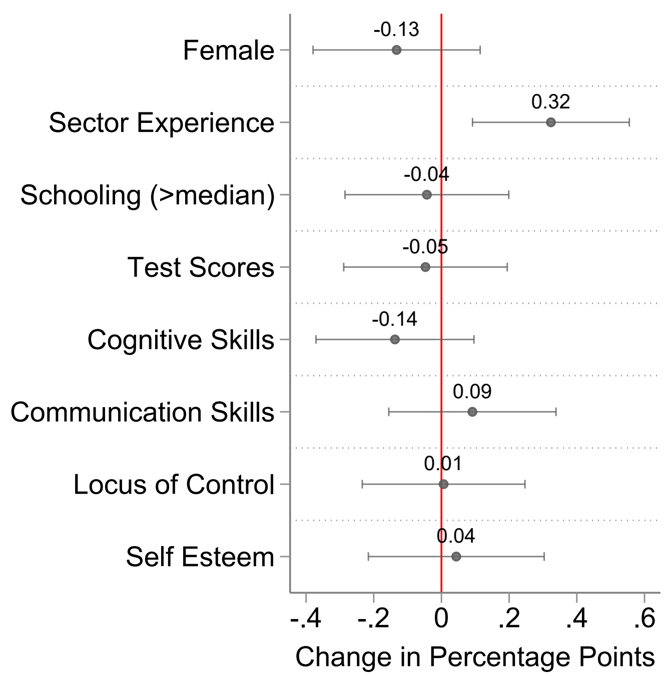
Notes: This is an example of checklists provided to supervisors for M1 and M2 survey modules. The checklists were used by supervisors during spot check rounds.

FIGURE A3. Further Evidence on Worker Skills Classification

A. Correlates of Interview Time per Crop (in Logs)



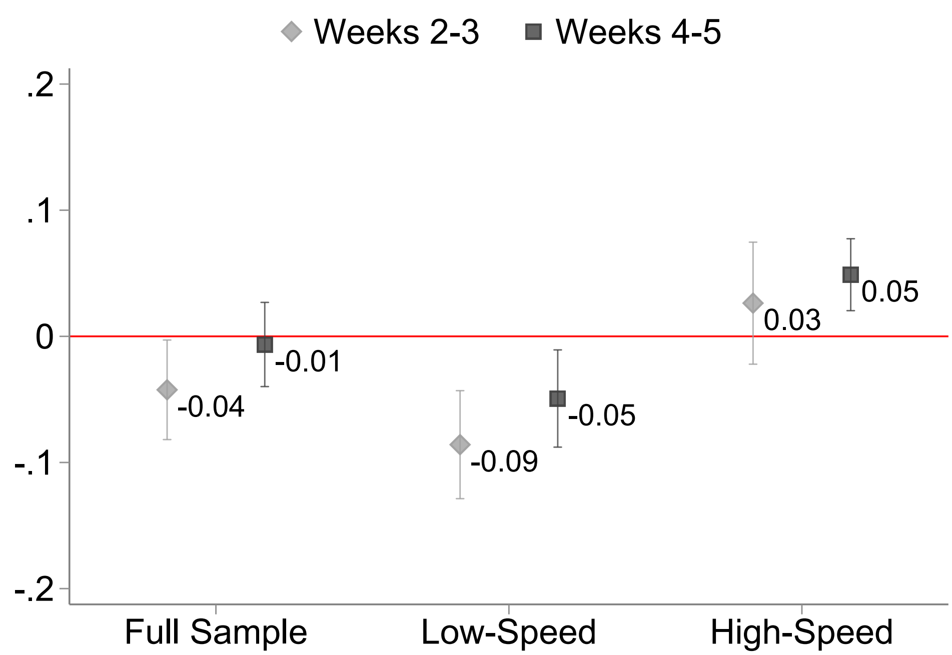
B. Correlates of Prior Task Experience (0/1)



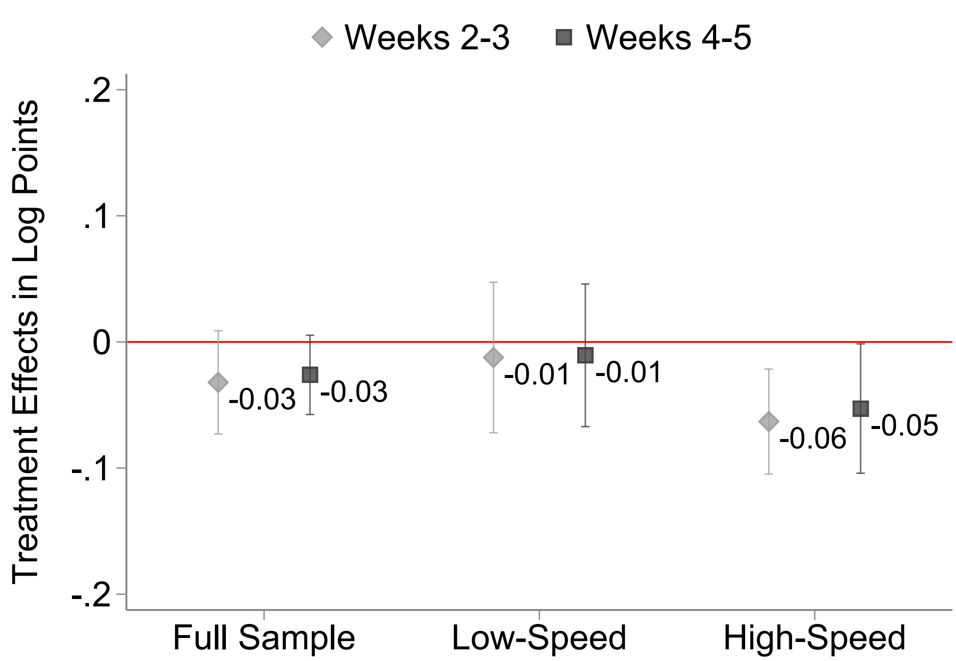
Panel A shows worker traits which predict their interview time (per plot-and-crop) in the first week of the project. The regression includes household controls and team-day fixed effects, and standard errors are clustered by worker. Panel B plots coefficients from 8 different worker-level regressions of the form: $X_i = \alpha + \beta \text{Task Experience}_i + \epsilon_i$, where X_i denotes each of the worker traits reported and Task Experience is a binary variable set to 1 of a worker has previously worked on a simlar task, and 0 otherwise.

FIGURE A4. Persistent Effects of Week 1 Supervision (by project phase)

A. Changes in Log Time (Mins)

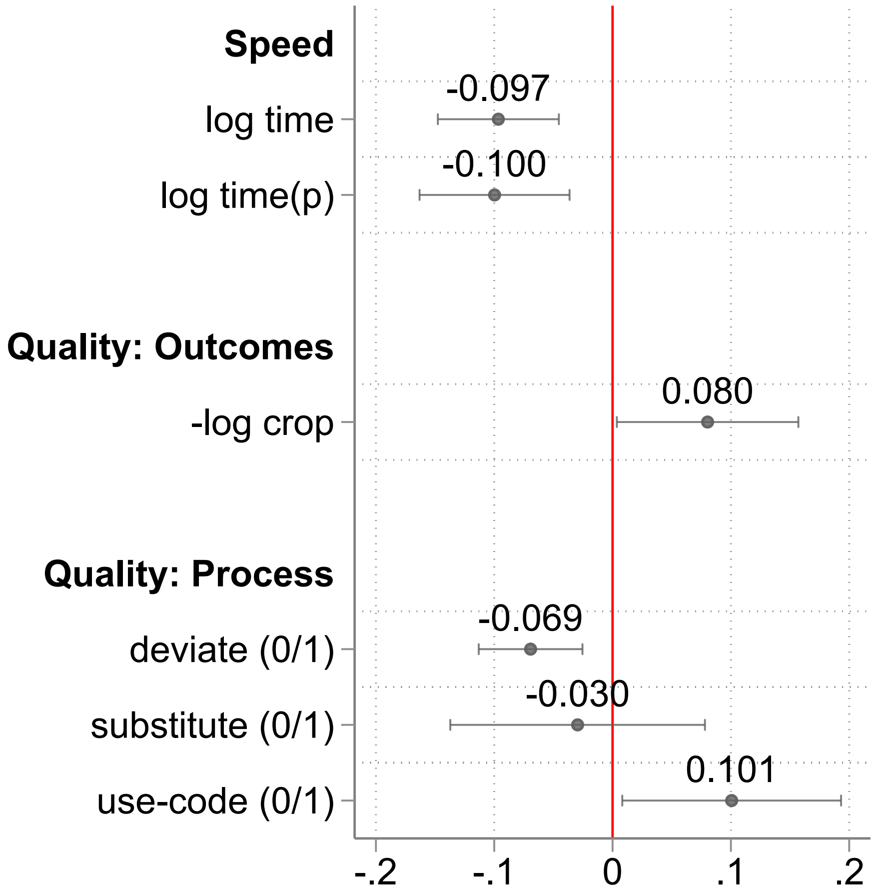


B. Changes in - Log Crop Observations



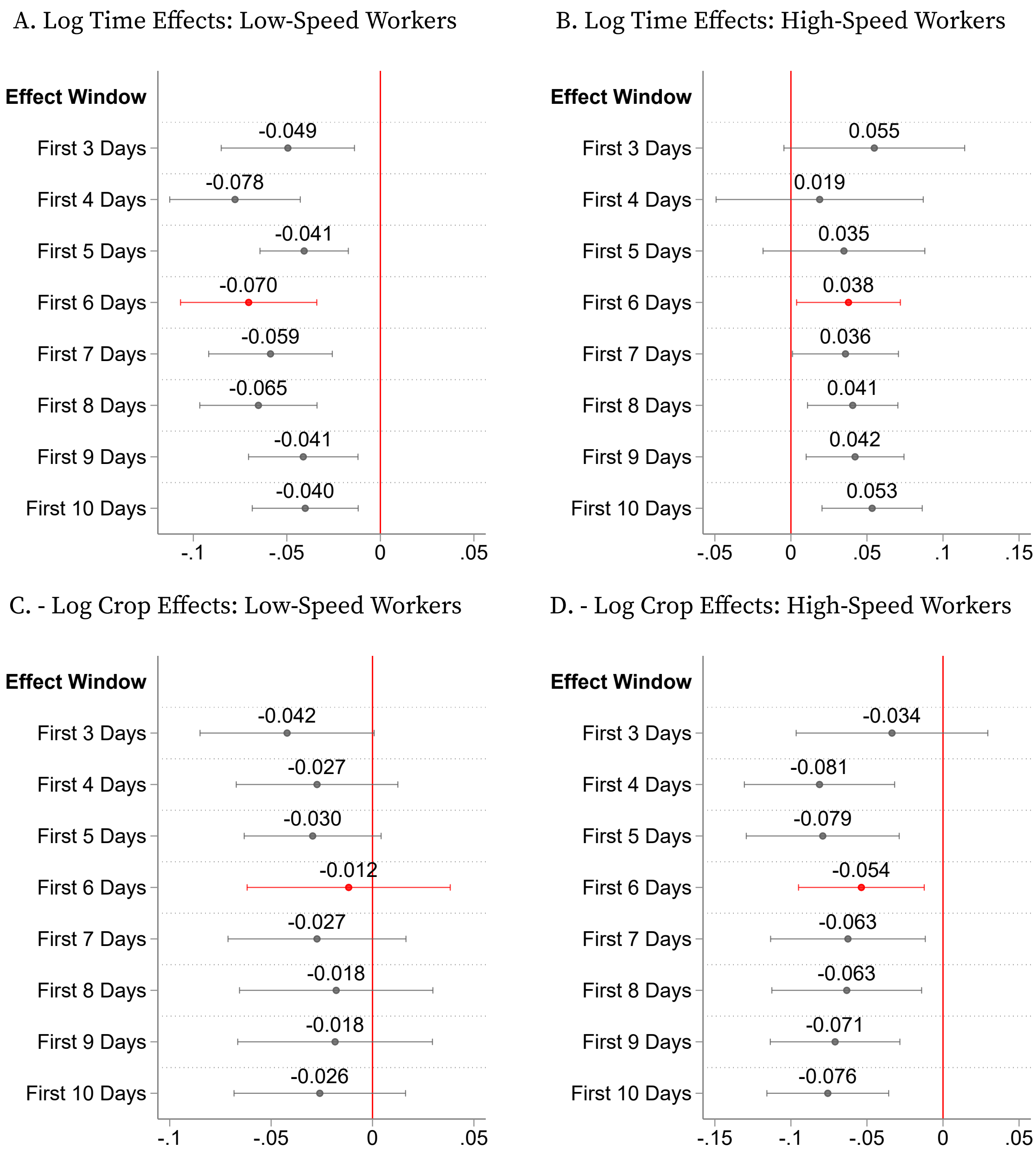
This Figure plots coefficients the persistent effects of week 1 supervision separately for the full sample of workers, low-speed workers and high-speed workers. The effects are computed separately over two phases of the project: weeks 2-3 (diamond markers) and weeks 4-5 (square markers). The sample is restricted to non-supervised worker days and standard errors are clustered at the worker level.

FIGURE A5. Robustness: Low-Speed Workers with Prior Firm Experience



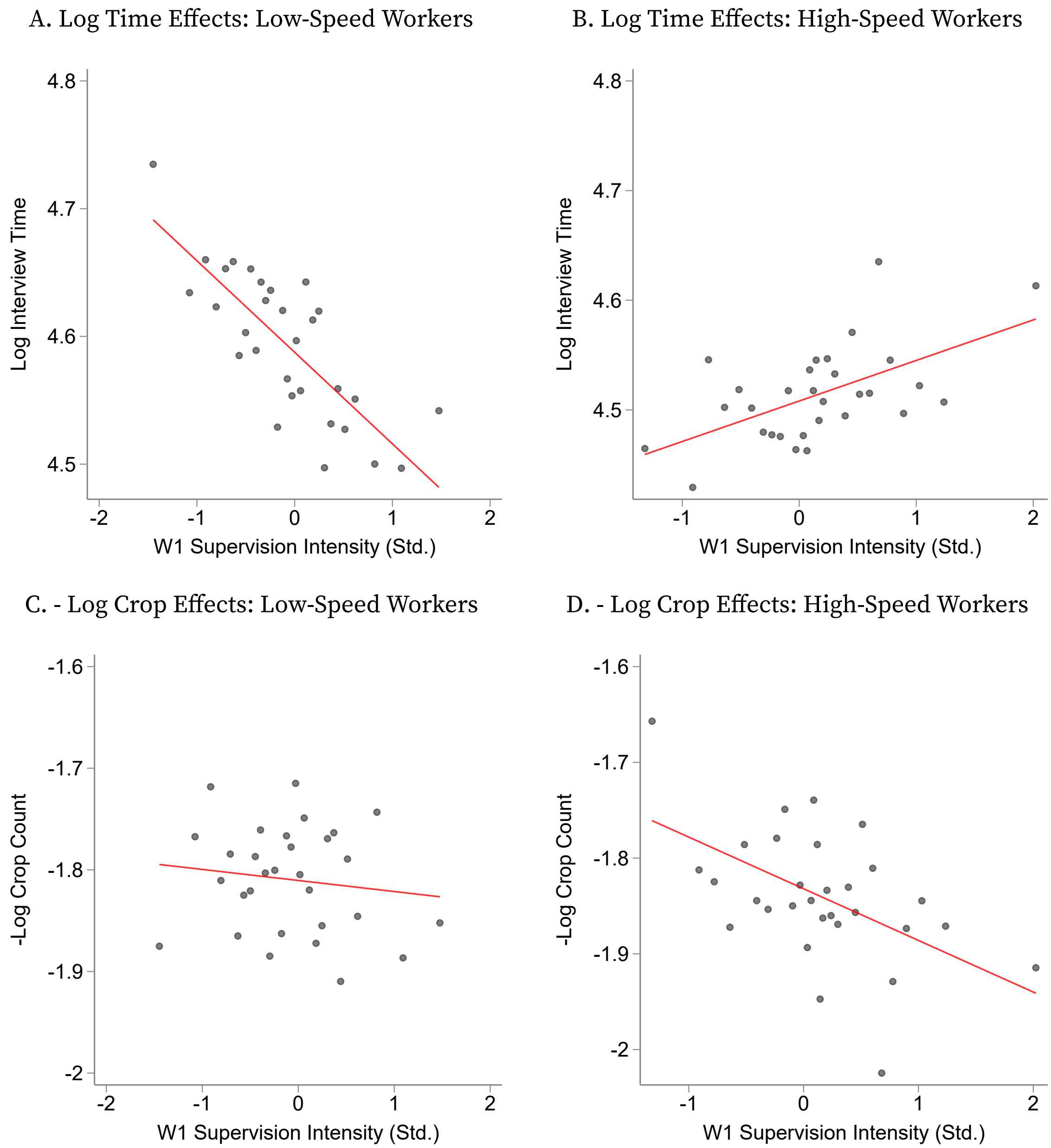
Sample: Not supervised worker-days, weeks 2-5. The sample is further restricted to 19 low-speed workers with prior firm experience. Standard errors are clustered at the worker level in the coefficient plot. We also report p-values computed using bootstrapped cluster standard errors at the worker level for small sample corrections. These are: log time (0.098), log time-p (0.183), -log crop (0.238), deviate (0.13), substitute (0.623) and use-code (0.25) respectively. This figure shows that we still observe a pattern of speeding up in response to week 1 supervision exposure when we focus on more experienced but low-speed workers. This is consistent with empirical predictions from our model.

FIGURE A6. Robustness to the Choice of Effect Window



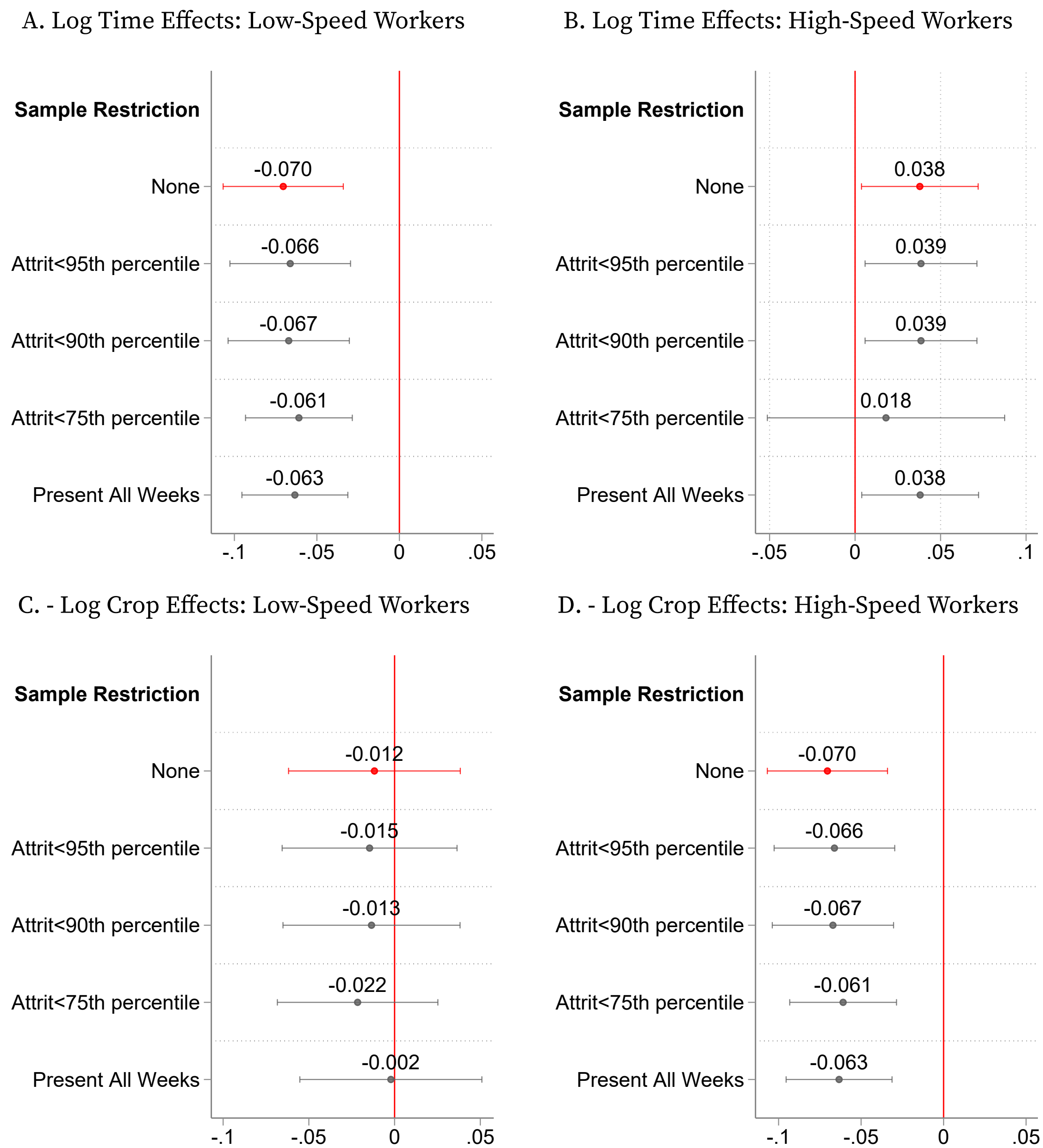
Sample Restricted to non-supervised worker days. The sample varies with each effect window. For example, the persistent effect of Day 3 is computed over Days 5 and onward (where the control for cumulative exposure since day 3 is non-missing). In red, we highlight the coefficient in our main specification.

FIGURE A7. Additional Evidence on Persistent Effects



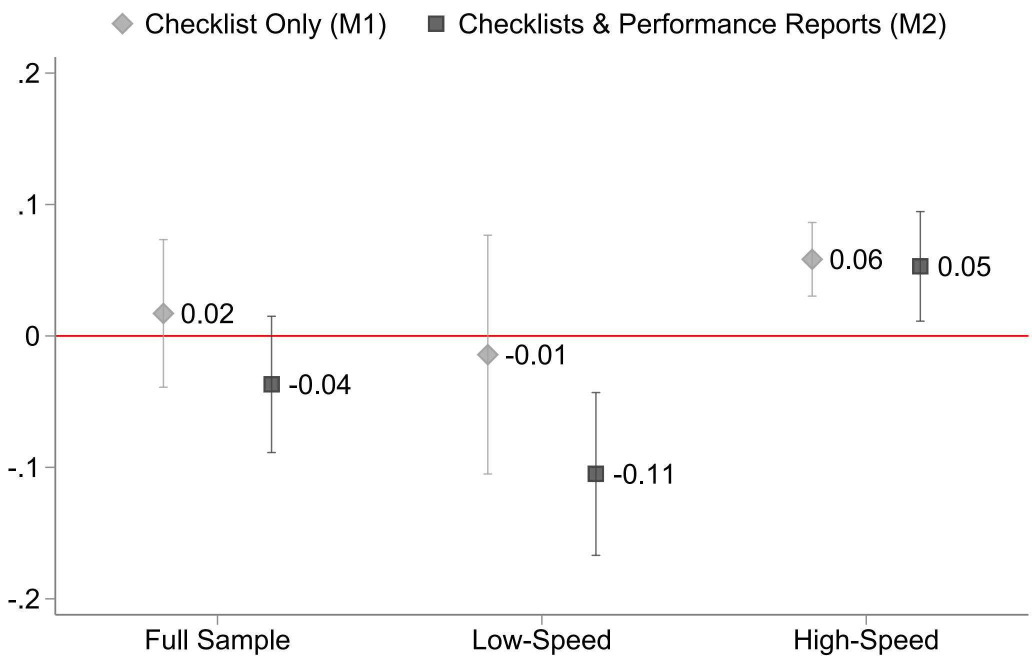
Notes: This figure shows binned scatter plots. Sample: Weeks 2-5 and restricted to non-supervised worker days. The x and y axes show residual series after correcting for the standard set of household, worker traits and team-day fixed effects.

FIGURE A8. Robustness to Attrition from the Sample



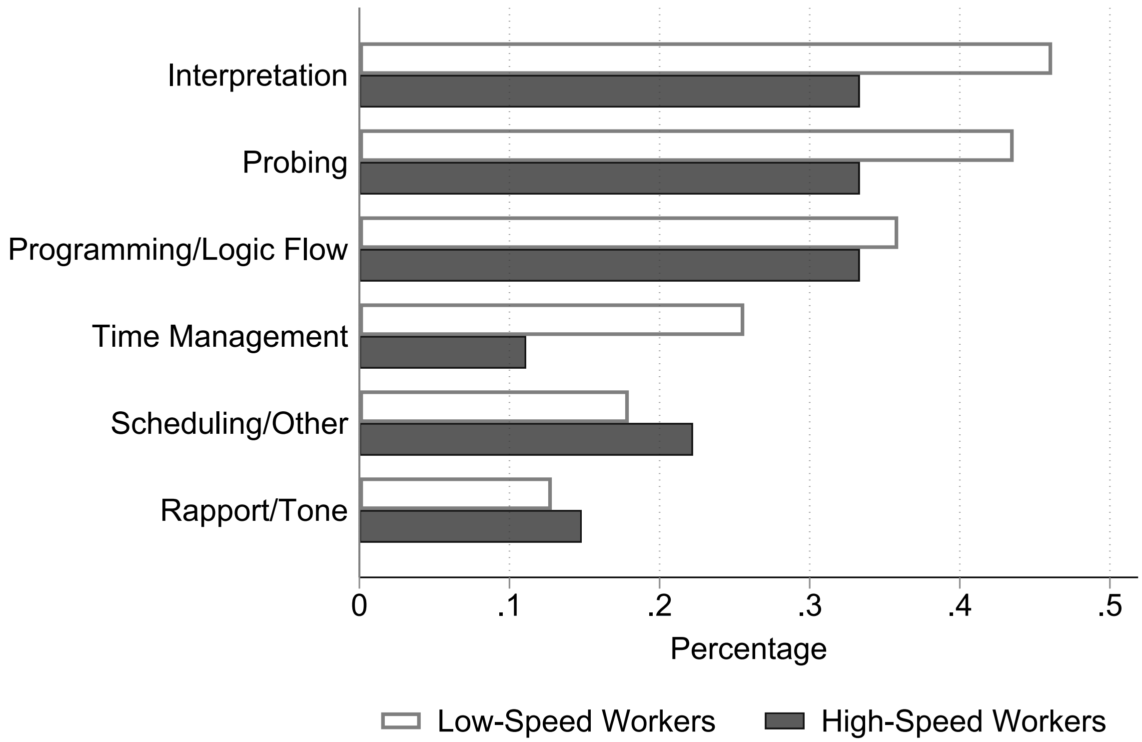
Sample Restricted to non-supervised worker days, from day 8 onward. Other sample restrictions vary with the specifications in each row. No sample restrictions (in red) denote our main estimates. Mean attrition: 12.89%. Less than 95th percentile attrition denotes employees who absent for no more than 15 of 30 days. Less than 90th percentile attrition denotes employees who were absent for no more than 10 days, Less than 75th percentile attrition denotes employees who were absent for no more than 5 days. The last model is restricted to employees who were present in all five weeks of the project.

FIGURE A9. Patterns in Persistent Effects, by information inputs



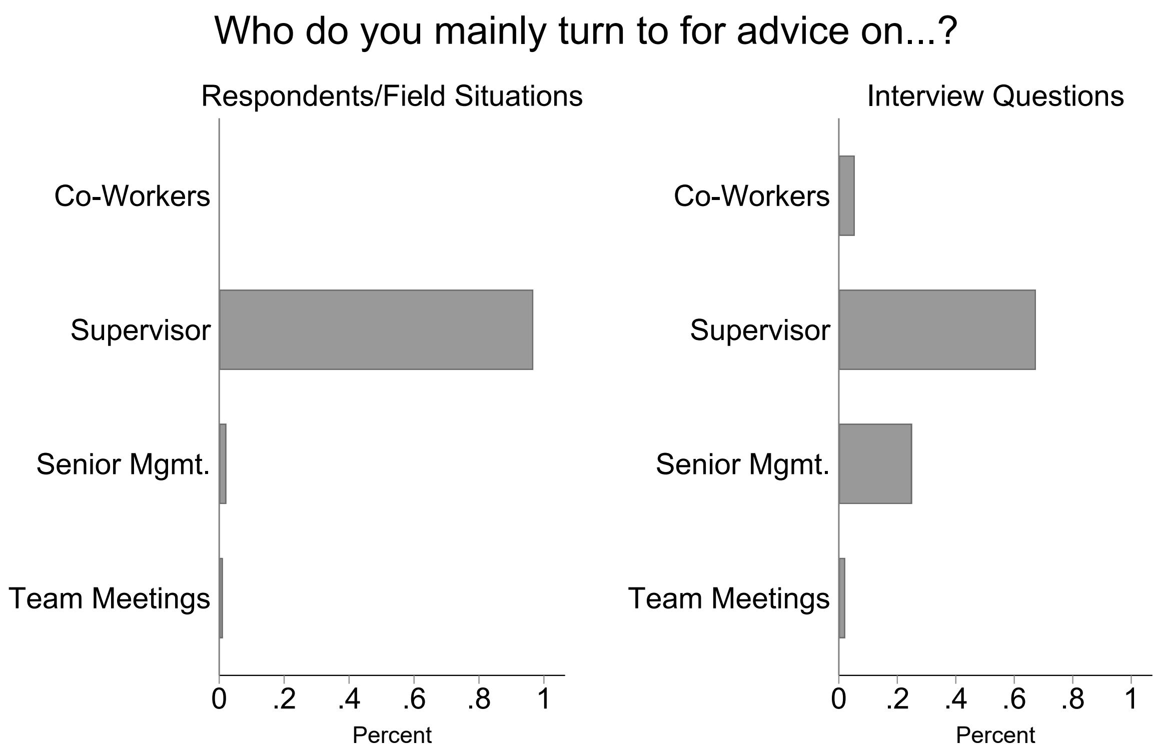
Sample restricted to non-supervised worker-days and weeks 2-5. Observations are at the interview-level. Dependent Variable: Module Times. Coefficients plotted: The persistent effect of a 1 SD unit change in Week 1 supervision on module times. Group M1 (diamond markers) consists of survey modules where supervisors were equipped with checklists only. Group M2 (square markers) consists of survey modules where supervisors were provided both checklists and performance reports.

FIGURE A10. Probability of Advice: by Topic and Worker Speed



Notes: This figure plots worker-reported probabilities of receiving advice from their supervisors (by topic) during the first week of the project. No advice is set to 0.

FIGURE A11. Cited Sources of Knowledge at Work



Notes: This figure plots workers' responses to the survey question: Who do you mainly turn to for advice on interviewees/field situations (Panel A) and interview questions (Panel B). This data was collected as part of the Worker Survey conducted before the onset of the household survey.

Appendix B. Additional Tables

TABLE A1. Correlates of Low Crops Recorded in an Interview

VARIABLES	(1) underreport	(2) underreport	(3) underreport	(4) log time(p)	(5) log time(p)
Crop Count (Std.)	-0.114***	-0.116***	-0.115***	0.143***	0.142***
	(0.034)	(0.037)	(0.036)	(0.010)	(0.007)
Observations	389	386	386	2,632	2,632
Adjusted R^2	0.061	0.102	0.111	0.234	0.404
Sample	Backchecks	Backchecks	Backchecks	All	All
Dep-Var Mean	0.33	0.33	0.33	72.60	72.60
Worker FE	No	Yes	Yes	No	Yes
Auditor FE	No	No	Yes	-	-

Notes: This table displays key correlations with our measure of shirking behavior. Observations are at the interview level. Sample: Interviews from Week 2 onward. Columns 1 and 2 are restricted to a random sample of interviews for which backchecks (or re-interviews) were conducted. The independent variable is the crop count recorded by a worker during a household interview (standardized). The independent variable in Columns 1-3 is a dummy variable coded to one if a backcheck revealed that a household had a higher number of plots and intercropped plots than what was reported in the initial interview, and zero otherwise. The number of observations vary across columns due to corrections for singleton observations implemented by the reghdfe package in STATA. The dependent variable in Columns 4 and 5 is the log of time spent on priority survey modules by a worker during the initial household interview (not the backcheck). All columns include controls for the survey day and household characteristics from a prior survey wave, including the gender of the household head, household size, an index of household wealth, and the log of plot-crop combinations cultivated (in the preceding season).

TABLE A2. Independence of Interview Assignments (Actuals)

VARIABLES	(1) female	(2) age	(3) task-exp	(4) firm-exp	(5) sector-exp	(6) cognitive	(7) comm-skills	(8) schooling	(9) test-score
HH Size (log)	0.021	0.021	-0.000	-0.030	-0.034*	0.005	0.013	-0.003	0.027
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.015)	(0.018)
Female head of HH (0/1)	0.037	-0.011	-0.009	-0.045*	0.029	-0.020	-0.001	-0.002	-0.032
	(0.023)	(0.023)	(0.024)	(0.024)	(0.023)	(0.023)	(0.023)	(0.017)	(0.022)
Wealth PCA score 1 (Std.)	0.011	-0.014	-0.010	-0.004	0.015	0.008	-0.016	0.003	0.010
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.009)	(0.012)
Baseline Plots-and-Crops (Log)	-0.041*	-0.010	-0.001	0.010	-0.005	0.017	0.017	-0.010	0.027
	(0.021)	(0.022)	(0.022)	(0.022)	(0.021)	(0.021)	(0.022)	(0.017)	(0.020)
Observations	2,632	2,632	2,632	2,632	2,632	2,632	2,632	2,632	2,632
Adjusted R^2	-0.030	0.014	-0.027	0.004	0.078	0.056	0.054	0.006	0.140
F-Test p-val	0.15	0.68	0.95	0.23	0.26	0.60	0.69	0.97	0.03

Notes: Observations are at the household interview level. The dependent variables consist of binary variables that represent traits of workers who interviewed the corresponding household. These include: gender (female=1, column 1) and above median age (column 2), prior task experience (column 3), prior firm experience (column 4), above median: sector experience (column 5), cognitive skills (column 6), communication skills (column 7), schooling (column 8) and recruitment test scores (column 9). The independent variables include key characteristics of the households interviewed, as measured during a prior baseline survey. These include: the household size (number of family members), a dummy variable for the household head's gender (female=1), an index of household wealth, and the number of plots-and-crops cultivated by the household in a prior agricultural season. Robust standard errors are reported in parenthesis. Overall, there is little evidence to suggest that supervisors systematically allocated certain types of households to specific workers (or worker types). Estimation sample: weeks 2 to 5.

Table A3. Compliance with Module Prioritization

VARIABLES	(1) module check=1	(2) module check=1	(3) module check=1
Checklists (M1)	0.257***	0.257***	0.257***
	(0.017)	(0.017)	(0.017)
Checklists & Reports (M2)	0.180***	0.180***	0.180***
	(0.018)	(0.018)	(0.018)
Observations	7,959	7,959	7,959
Adjusted R^2	0.074	0.105	0.209
Supervisor FE	Yes	Yes	No
Worker FE	No	Yes	No
Interview FE	No	No	Yes
M0 Mean	0.502		

Notes: Observations are at the interview and survey module level. Standard errors clustered by worker are in parentheses. Reports are based on supervisor reports and only available in instances where spot checks were conducted (statistics are conditional).

Table A4. Additional Evidence on Gaming

	Low-Speed Workers		*High-Speed Workers*	
	(1)	(2)	(3)	(4)
VARIABLES	log time	divergence	log time	divergence
W1 Supervision (Std.)	-0.067***	0.031	0.010	-0.267**
	(0.022)	(0.079)	(0.021)	(0.107)
Observations	1,091	234	761	165
Adjusted R^2	0.217	0.105	0.312	0.038
Control Mean	45.90	-0.12	45.90	0.01
Control SD	14.01	1.01	14.01	1.08

Notes: Observations are at the interview level in columns 1 and 3; and 3-day bin level in columns 2 and 4. Estimation sample: weeks 2-5. Standard errors clustered by worker are in parentheses. The interview level regressions reported in columns 1 and 3 are further restricted to non-supervised worker days. The dependant variables (log time and the divergence index) reported pertain to survey modules that were supervised less intensively (M0). All columns include controls for household characteristics, worker characteristics (dummies for above median traits) interacted with day fixed effects and team-day fixed effects. Standard errors clustered by worker shown in parentheses. See Equation 11 for further details.

Table A5. Additional Evidence on Incentives

VARIABLES	(1) log time (all)	(2) log time (M0)	(3) log time (M1)	(4) log time (M2)
Supervise Today (0/1)	0.113***	0.094***	0.130***	0.162***
	(0.014)	(0.017)	(0.020)	(0.022)
W1 Supervision (Std.)	-0.015	-0.030	0.031	-0.022
	(0.020)	(0.018)	(0.024)	(0.027)
W1 Supervision x Supervise Today	0.002	-0.012	-0.006	0.018
	(0.010)	(0.015)	(0.019)	(0.023)
Observations	2,631	2,631	2,624	2,630
Adjusted R^2	0.333	0.227	0.225	0.242
Control Mean	100.16	44.11	23.51	23.22
Control SD	30.24	13.32	8.70	10.35

Notes: Observations are at the interview level. Estimation sample: Weeks 2 to 5, all workers and days. The dependent variables are: log interview time (column 1), log M0 module time (column 2), log M1 module time (column 3) and log M2 module time (column 4). M1 and M2 survey modules were supervised more intensively as supervisors were provided checklists (for both M1 and M2) and performance reports (M2 only). All columns include controls for household characteristics, worker characteristics (dummies for above median traits) interacted with day fixed effects and team-day fixed effects. In addition, all control variables and fixed-effects are interacted with a dummy for workers' prior task experience. Standard errors clustered by worker shown in parentheses.

Table A6. Additional Evidence on Clarity

VARIABLES	(1) low-speed workers	(2) high-speed workers
Supervise Today (0/1)	0.012**	0.019***
	(0.005)	(0.005)
W1 Supervision (Std.)	0.010*	0.002
	(0.005)	(0.006)
W1 Supervision x Supervise Today	0.006	0.002
	(0.005)	(0.006)
Observations	1,557	1,074
Adjusted R^2	0.096	0.086
Control Mean	0.712	0.701

Notes: Observations are at the interview level. Estimation sample: weeks 2-5, all days. The dependent variable in both columns 1 and 2 represents the share of time allocated to high priority survey modules used to compute outcomes for the underlying study. Sample restrictions: low-speed workers only (column 1), high-speed workers only (column 2). All columns include controls for household characteristics, worker characteristics (dummies for above median traits) interacted with day fixed effects and team-day fixed effects. Standard errors clustered by worker shown in parentheses.

www.theigc.org