Pre-Analysis Plan:
Rwanda Personality/Motivation Based Nudges for Extension Agents

Martin Abel (Middlebury College), Tomoko Harigaya (PxD),
Michael Kremer (University of Chicago), and Jessica Zhu (PxD)
This version: December, 2021

This document details the primary research questions in a randomized controlled trial of a digital motivation experiment in Rwanda, and analytical methods to be used to answer them. This pre-analysis plan was written and will be registered prior to collection/access to any outcome data. Our work is in the spirit of Banerjee et al. (2020), and conducted for this evaluation may extend beyond what is specified here.

Table of Content

Abstract

This study examines how customizing motivational messages for volunteer agents by personality traits influences performance and motivation in the context of agricultural extension in Rwanda. We administered a phone survey with 10,187 village agricultural volunteers to gather personality traits using Susceptibility to Persuasion Scale (STPS) questions and designed four sets of SMS task reminders that appeal to four different STPS personality traits. We then randomly assigned volunteers to receive (i) no SMS messages, (ii) randomly assigned SMS messages, (iii) personality-matched messages, or (iv) personality-mismatched messages. In addition we cross-randomize whether FPs receive a reminder of the performance goals they set. We will compare self-reported performance and motivation among agents across experimental groups using a combination of phone survey data and administrative data.

1. Introduction

In Rwanda, there is a network of volunteer village-level frontline agricultural extension agents, farmer promoters (FPs). Farmer promoters receive training and in-kind benefits, and are tasked to teach and promote the adoption of recommended inputs (e.g., hybrid seeds and fertilizer) and practices (e.g., good farming practice and tree planting) to their villagers. This study aims to examine how customizing motivational messages by personality traits and goal reminder messages can influence farmer promoters' performance and motivation.

2. Research design

2.1. Study sample and recruitment

The sample of this study is farmer promoters in Rwanda. In Rwanda, there are approximately 14,500 farmer promoters, of which OAF has phone numbers for approximately 13,000. To recruit participants, enumerators from OAF contacted farmer promoters via phone calls and asked if they consented to responding to a baseline survey. Farmers promoters who completed the baseline survey, 10,187 farmer promoters, become this study sample.

2.2. Study settings and intervention messages

The intervention took place in three settings, which are described below. Farmer Promoters received SMS messages right before and during the 2022A season (August 2021 - November 2021). We may continue to send a small number of messages during 2022.

(1) Subsidized input promotion campaign: In May – June 2021, OAF trained FPs on marketing techniques along with a radio campaign to accelerate farmer registration and purchase for a nationwide government subsidized inputs program, such as for hybrid maize seeds, macronutrient fertilizer, and compound blends. During the peak purchase/registration time in August 2021, we sent digital reinforcement messages to encourage farmer promoters

to help farmers register and purchase subsidized inputs, such as hybrid maize seeds and fertilizer.

(2) Demonstration plot and farmer training campaign: One of the key responsibilities of FPs is to maintain a demonstration plot and use the demonstration plot to train farmers on good agricultural practices. In July and August of 2021, OAF provided in-person agricultural training to FPs, so that FPs could train their fellow farmers on their demonstration plots starting in September. From mid-September to early-October, we sent digital reinforcement messages to encourage farmer promoters to maintain a good demonstration plot and train farmers on agricultural practices. In one message, we specifically remind FPs to use the correct fertilizer type and remember to weed, as these were practices OAF wished to emphasize.

(3) Agroforestry trees campaign: In October and November of 2021, OAF distributed 20 million agroforestry trees to smallholders in 27 districts in Rwanda for free. Tree planting mitigates soil erosion in notoriously hilly Rwanda and helps mitigate climate change. Farmer promoters are responsible for mobilizing farmers to pick up those trees on the distribution days and training farmers on how to plant trees in order to maximize the tree survival rate. There are two subprograms under this agroforestry tree campaign. One is called the consolidated land planting program, in which farmers come to a designated government land (i.e., a site) and plant trees together. The other is called the individual land planting program, in which farmers pick up trees from a distribution location and plant trees at locations of their own preference. Farmers promoters are tasked to help with both programs and have received training and guidelines from OAF since August 2021. From mid-October through November 2021, we sent digital reinforcement messages to encourage farmer promoters to mobilize farmers for individual tree pick-up and consolidated land planting day, to dig holes before consolidated land planting day, and to train farmers on tree planting and management for tree survival. Moreover, given the correct post-planting management is crucial for tree survival, we may send some messages on tree management over the year of 2022.

2.3. Baseline survey

From June – July 2021, OAF and PxD conducted a 25-minute baseline phone survey to collect data on FPs' demographic characteristics, motivations, personalities, performance in the previous agricultural season, and goals for the upcoming agricultural season.

Personality modules included a selection of questions about Susceptibility to Persuasion Scale (STPS) (following Kaptein et al., 2012) and 10 of the 15 questions about the Big Five personality traits from the Skills Towards Employability and Productivity (STEP) World Bank surveys (following Laajaj et al., 2019). These questions were taken directly from the corresponding literature and then adapted to our context. The STPS questions aim to measure the susceptibility of people to the six fundamental principles of persuasion identified by Cialdini (2001), more explanation in the following subsection.

Performance and goal setting modules asked farmer promoters about the number of farmers they helped/ plan to help register for subsidized inputs, the number of training meetings they held/ plan to hold on their demonstration plots, and the number of farmers they trained/ plan to train on their demonstration plots. For the performance goals, we randomized the order of questions, whether we first ask about last year's performance or first ask about next year's goals. (This variation will be used in the analysis as explained below.)

2.4. Customization of motivational messages

We customized messages based on responses to the STPS module. Kaptein et al (2012) describe how messages can be tailored to STPS types: "**Reciprocity**, people feel obligated to return a favor, **scarcity**, people will value scarce products, **authority**, people value the opinion of experts, **consistency**, people do as they said they would, **consensus**, people do as other people do, and **liking**, we say yes to people we like." Based on the baseline variation within and correlations across six constructs, we excluded "liking" and "authority".

We identify the "matching" and "mismatching" personality traits using the following process. We order the 4 personality traits according to their baseline response variations among FPs. The trait that has the largest standard deviation is prioritized first, the specific order as follows - scarcity, consensus, reciprocity, and commitment. For the matching treatment arm, we first assign the farmer promoter who has the highest score of scarcity trait among FPs of this arm to the scarcity messages, then assign the remaining FP who has highest score of consensus trait to the consensus messages, then assign the remaining FP who has highest score of reciprocity trait to the reciprocity messages, and last assign the remaining FP who has highest score of commitment trait to the commitment messages. These 4 assignments make up 1 iteration. We repeat this iteration until every FP is assigned to one trait. For the mismatching treatment arm, the process is similar. The difference is that we assign the farmer promoter who has the lowest score of a certain trait among FPs of this arm to messages of that trait.

2.5. Randomization design

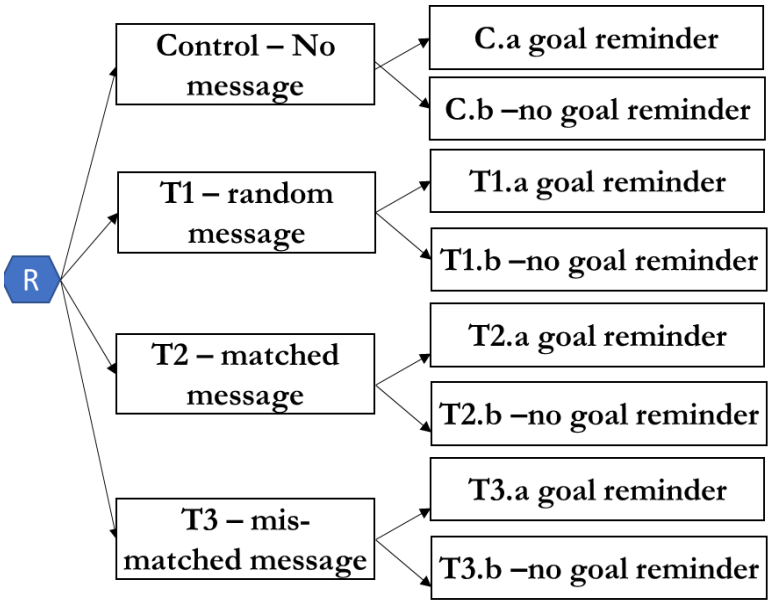
The objectives of the intervention are to examine whether motivational messages can improve FPs' performance and moreover to test whether motivational messages designed to appeal to the specific personality can lead to better FP performances and farmer outcomes. To do so, as depicted in Figure 1, we randomized participants into one of four main groups with stratification by the district and the quartile level of personality traits variation in baseline responses:

- Control: a group of FPs that don't receive any motivational messages.
- Random Message (T1): a group of FPs that receive motivational messages with personality traits that are randomly assigned to them.
- Matching Message (T2): a group of FPs that receive motivational messages with personality traits that match their personality traits.

- Mismatching Message (T3): a group of FPs that receive motivational messages with personality traits that mismatch their personality traits.

Additionally, farmer promoters in each of the four groups (C, T1, T2, and T3), were further subdivided to either (a) receiving a goal reminder based on their response in the baseline survey or (b) not receiving a goal reminder. The goal reminders were implemented for two settings, (i) the subsidized input promotion campaign and (ii) the demonstration plot and farmer training campaign. An example of the goal reminder messages is as follows: *"Remember you said your goal for this season was to help XXX farmers register for subsidized inputs! Keep up with your goal!"*

Figure 1. Randomization Design



We conducted the above (re-)randomization 100 times, following Banerjee et al. (2020). For each randomization, we performed the following 6 balancing tests:

$$Y_{ij} = \alpha + X'_i\gamma + \lambda_s + e_i \; if \; sample = j, (1)$$

where outcome Y_{ij} stands for assigned to received any SMS messages (including both motivational messages and goal reminder messages), assigned to the treatment group (i.e., receiving motivational messages), assigned to the T3 treatment group (i.e., receiving mismatched personality trait motivational messages), assigned to the T2 treatment group (i.e., receiving matched personality trait motivational messages), assigned to the T2 treatment group (i.e., receiving matched personality trait motivational messages), and assigned to receive goal reminder messages, respectively; X'_i stands for 4 groups of balancing variables: a) FPs' characteristics, b) FPs' performance in the 2021A season, c) FPs' goals for the 2022A season, and d) FPs' personality trait score;[1] $sample = j$ refers to full sample, full sample, sample of T1

[1] The specific balancing variables included in *X* are: a) FPs' characteristics - age, gender, and number of years lived in the village; b) FPs' performance in the 2021A season - number of farmers FPs helped register for subsidized inputs, number of training meetings FPs held on their demonstration plots, and number of

and T3, sample of T1 and T2, sample of T1 and T2, and full sample, respectively. We use Huber-White robust standard errors and include stratification block fixed effects λ_s. Within each balancing test, we saved the p-value of the joint F-test of all variables of each balancing group, 4 p-values corresponding to 4 balancing variable groups. Given 6 balancing tests and 4 p-values per test, we had 24 p-values per randomization. We first recorded the lowest p-value (among the 24 p-values) for each randomization, and then selected the randomization (among the 100 randomizations) that had the highest lowest p-value as our final randomization.

3. Variable Definitions

3.1. Key data sources

We will implement our intervention and collect outcomes in 3 settings, as described in section 2.2. We will conduct an endline phone survey with all farmer promoters in the study sample in February and March 2022. The phone survey will gather FP's self-reported behavior outcomes and also FPs' knowledge and attitude, and feedback toward the intervention messages for insights on potential mechanisms. In addition, we will use several administrative datasets collected through OAF's regular operations, such as SEDO survey as quality evaluations of the demonstration plots and tree registration and pick up records.

3.2. Outcomes

3.1.1 Subsidized inputs campaign

The main outcomes for this setting are 1) the numbers of farmers that FPs helped with the registration of subsidized inputs, 2) the share of farmers within the village that FPs helped with the registration of subsidized inputs, 1) the numbers of farmers that purchased subsidized inputs for the 2022A season, and 2) the share of farmers within the village that purchased subsidized inputs for the 2022A season. We will collect this information via the endline FP phone survey.

3.1.2 Demonstration plots and farmers training campaign

The outcomes of interest for this setting are 1) whether FPs have a demonstration plot, 2) the number of training sessions that FPs have held on their demonstration plot, 3) the number of farmers and the share of farmers that FPs have trained on their demonstration plot, and 4) the quality of FPs' demonstration plot. For outcomes (1)-(3), we will collect them in the endline FP phone survey. For outcome (1) and (4), we will use data from OAF's monitoring USSD surveys filled by sector economic development officers (SEDOs), the manager of farmer promoters.

farmers FPs trained on their demonstration plots; c) FPs' goals for the 2022A season - number of farmers FPs plan to help register for subsidized inputs, number of training meetings FPs plan to hold on their demonstration plots, and number of farmers FPs plan to train on their demonstration plots; and d) FPs' personality trait score - scores for reciprocity, scarcity, commitment, and consensus.

3.1.3 Agroforestry tree campaign

For the individual land program, our main outcomes of interest are a) the number and b) the share of households within the village that picked up trees during the distribution day. We will have this information from both the OAF's administrative distribution records and the endline FP phone survey. In addition, we may collect the specific activities that FPs took to ensure that farmers pick up trees in the phone survey.

For the consolidated land program, we will have data on 1) the number of farmers who came to the consolidated land to plant the trees, and 2) the number of trees that were planted on the consolidated land from OAF's administrative distribution records. These data are at the consolidated land (site) level, which usually includes 5 villages. Because the outcomes of the consolidated land program are the results of work by multiple (i.e., 5) FPs, of whom have different treatment statuses, we will only analyze the treatment impact of this setting as an exploratory analysis.

Furthermore, at the end of 2022, OAF plans to conduct an in-person monitoring evaluation with farmers who have picked up trees in 2021, to learn how many trees that farmers picked, planted, and are still alive after 1-year. If we will gain access to this survey data, we will include them in our analysis as well.

3.1.4 Additional outcomes from the endline FP phone survey

In order to understand the mechanism of this intervention, in addition to behavior outcomes described above, we will collect 3 others sets outcomes in the endline phone surveys with farmers promoters, as follows:

- Agricultural knowledge outcomes: These questions aim to provide a measure on FPs' engagement with their responsibilities through examining their agriculture knowledge, which is an important component of FPs tasks.
- Attitude outcomes: We will ask questions about FPs' attitudes towards meaningfulness of work, role ambiguity, recognition from work, and satisfaction with work, which we expect to be affected by the SMS intervention and be correlated with FPs' productivity.
- Feedback on the SMS messages: These questions aim to understand whether FPs received, remembered, learned from, and had time to act on the treatment SMS messages.

3.1.5 Summary index

To avoid bias in standard errors due to multiple hypothesis testings, we will construct summary indices using variables listed above (see section XXX for more details), for both downstream behavior outcomes and mechanism outcomes. Our main outcomes of interest will be those summy indices, rather than individual outcomes.

3.3. Procedures for addressing missing data and outliers

3.2.1. Missing data from item non-responses

For demographic variables in the baseline survey data, we will substitute the missing values with the imputed median values of that variable by the stratification block. For personality trait and outcome variables, we do not perform imputation and substitution for missing values. To note, we will conduct imputation for balance checks and the treatment analysis, but we did not do it during the (re-)randomization stage.

3.2.2. Outliers

Because of the noisy nature of self-reported measures, we will winsorize our continuous outcome variables at the 99 percentile as the main specification, while conducting robustness checks with the unwinsorized values and values that are winsorized at 95 percentile.

4. Hypotheses

4.1. Impact on primary outcomes

We aim to answer the following specific research questions:

R1: What is the effect of motivational messages?
We will examine this question through comparing treatment FPs and control FPs. We expect that motivational messages will improve FPs' performance, as those messages serve both as a direct encouragement and a reminder.

R2: What is the effect of personalizing messages?
We will examine this question through comparing T1, T2, and T3 with control, and among T1, T2, and T3. We expect that motivational messages that match FPs' personality traits (T2) will have the largest impact on improving FPs' performance, and motivational messages that mismatch FPs' personality traits (T3) will have the least impact on improving FPs' performance. It is also possible that mismatched messages in T3 may have adverse impacts on FPs' performance compared to no messages in the control group.

R3: What is the effect of receiving goal reminders?
We will examine this question through comparing control FPs who received goal reminders with those who did not. We expect goal reminders to have a positive effect on farmer promoters' performance.

R4: What is the additional effect of adding goal reminders to customized motivational messages?
We will examine the interaction impact between goal reminder messages and customized motivational messages.

4.2. Impact mechanisms

As part of our endline survey, we will collect data on farmers promoters' knowledge and attitude, and feedback toward the intervention messages. Responses from those questions could provide insights on potential mechanisms for the observed behavior changes.

Attitude change: We measure the following attitudes, which have been shown to be important predictors of worker productivity (Rubenstein et al. 2018, Cassar and Meier 2018).

- How *meaningful* do people find their work?
  We measure this by asking "Sometimes my work as FP does not feel very meaningful."
- How much *role ambiguity* do workers face?
  We measure this by asking "I sometimes don't know what is expected of me in my role as a FP."
- How much *recognition* do workers receive?
  We measure this by asking "Other farmers understand how difficult my role as a FP is."
- How *satisfied* are workers?
  We measure this by asking "I hope that some other farmer in my village will take over as the FP in the next 2 years."

Behavior change: We collect the hours worked and the farmer reach in the role as a FP:

- In a typical week in the 2022A season, how many hours did you work in your role as a FP?
- In a typical week in the 2022A season, how many farmers did you visit in your role as a FP?

Knowledge change: One mechanism for why the messages lead to better outcomes is because they may have a positive effect on how much relevant knowledge FPs have. This could be for two reasons:

- FPs are more engaged in their work and thus gain more knowledge. We measure this by asking at endline:
  - "What is the correct planting density you should use for planting maize seeds (e.g., the correct spacing between rows)?"
  - "How many agroforestry species should each farmer receive at least during the individual tree distribution day?"
- FPs may learn directly from the content of messages. We measure this by asking: "When you received a message, how often did you learn new information from it?"

Memorizing messages: One mechanism for why (customized) messages are more impactful is because recipients are more likely to retain the messages.

- We measure this by asking FPs in the endline survey whether they recall any specific content from the messages they received:
  - "Can you recall the content of those messages?"
  - "What was your goal regarding the number of meetings to hold on your demonstration plots during the 2022A season?"
- We then compare the share that memorizes messages across treatment arms.

4.3. Heterogeneous effects

The heterogeneous effect that we are particularly interested in is the differential treatment effects across FPs that have different levels of variation in their personality traits. Specifically, we will estimate results separately for each quartile of FP's personality measurements (which was also used as part of a stratification block). We expect the impact of personality matched messages is stronger among individuals that have larger variation in their personality traits, in other words, individuals with more salient specific traits.

5. Empirical strategy

5.1. Balance checks

We examine 3 groups of variables for (re-)randomization balance checks:

- Characteristics of farmer promoters: FPs' age, gender, number of years being a FP, number of years lived in the village, and the version of baseline survey was used.
- Farmers promoters' self-reported previous performance in 2021A season and future performance goals for the 2022A season:
    - The number of farmers FPs helped/plan to help register for subsidized inputs,
    - The number of training meetings FPs held/plan to hold on their demonstration plots,
    - The number of farmers FPs trained/plan to train on their demonstration plots.
- Farmer promoters' personality traits: the trait scores for reciprocity, scarcity, commitment, and consensus.

We use the following specification to check balance for each variable listed above:

$$X_i = \alpha + \beta_1 CG_i \quad + \beta_2 \quad T1N_i + \beta_3 T1G_i + \beta_4 T2N_i + \beta_5 T2G_i + \beta_6 T3N_i + \beta_7 T3G_i + \lambda_s + e_i, \quad (2)$$

where X_i stands for the balancing variable, CG, $T1N$, $T1G$, $T2N$, $T2G$, $T3N$, and $T3G$ are indicators for whether FP i was assigned to control with goal reminder group, T1 without goal reminder group, T1 with goal reminder group, T2 without goal reminder group, T2 with goal reminder group, T3 without goal reminder group, and T3 with goal reminder group, respectively. We use Huber-White robust standard errors and include stratification block fixed effects λ_s (of which there are 126 blocks in the sample). Furthermore, we will check the significance for a joint test of those individual tests of 7 treatment groups.

Results of balance checks are presented in Table 1-6 in Appendix 1. As expected from (re-)randomization, we observe balance on most of those variables across various experimental arms. The three exceptions are the i) number farmers FPs helped register for subsidized inputs in 2021A season, ii) number of training meetings FPs held on their demonstration plots in 2021A season, and iii) number of farmers FPs plan to train on their demonstration plots in 2022A season.

We observed 10 imbalanced comparisons out of 150 tests we ran, this is in line with the number of comparisons one would expect to be imbalanced due to chance given the number of tests conducted.

5.2. Treatment effects: Intention-to-treat effect (ITT)

To test the effect of receiving *any* motivational message on FPs' performance **(***#R1***)**, we will estimate the following OLS specification:

$$y_i = \alpha + \beta_1 T_i + X'_i\gamma + \lambda_s + e_i, \qquad (3)$$

where y_i stands for outcome y for FP i, and T is an indicator for whether FP i was assigned to any of the treatment groups. We control for FP specific covariates and dummy variables indicating whether those values are imputed or not due to missing observations in vector X, and the stratification block fixed effects λ_s.[2] Robust standard errors are computed at the individual level.

To test the effects of different treatment messages on FPs' performance **(***#R2***)**, we will estimate the following OLS specification:

$$y_i = \alpha + \beta_1 T1_i + \beta_2 T2_i + \beta_3 T3_i + X'_i\gamma + \lambda_s + e_i, \quad (4)$$

where $T1$, $T2$, and $T3$ are indicators for whether FP i was assigned to treatment group 1, 2, and 3, respectively, and other notations are the same as above. Robust standard errors are computed at the individual level. Coefficient β_i measures the effect of treatment group i relative to the control group (pooled across the goal reminder messages). We will also test whether coefficients of various treatment groups are significantly different from each other**.**

To measure the effect of goal reminders across all groups, we estimate the following OLS specification:

$$y_i = \alpha + \beta_1 G_i + X'_i\gamma + \lambda_s + e_i, \qquad (5)$$

where G indicates whether FP i received goal reminder messages or not, and other notations are the same as above. Robust standard errors are computed at the individual level.

To test the additional effect of goal reminders on different types of treatment message, we estimate the following OLS specification:

$$y_i = \alpha + \beta_1 T1_i + \beta_2 T2_i + \beta_3 T3_i + \beta_4 G_i + \beta_5 T1_i \times G_i + \beta_6 T2_i \times G_i + \beta_7 T3_i \times G_i + X'_i\gamma + \lambda_s + e_i. \quad (6)$$

Notations are the same as above. Robust standard errors are computed at the individual level. β_1, β_2, and $\beta_3$measure the effect of treatment T1, T2, T3 *without* additional goal reminders. β_4 estimates the effect of only receiving the goal reminder. β_5, β_6, and β_7 measure the *additional* effect of goal reminders for treatment T1, T2, T3, respectively.

As part of the goal elicitation, we randomized whether we first asked respondents about their last year's performance or first about their goal. This created exogenous variation in the goal magnitude. However, one limitation of the data is that differences in magnitude did not reach

[2] The stratification block is composed of two variables: 1) district and 2) quartile group of variance of FP's personality measurements in the baseline survey.

conventional significance levels, which raises concerns about estimation with weak instruments. We will therefore only use this variation in an exploratory analysis. Specifically, we will instrument for the goal magnitude with our random order indicator in the first stage and then estimate whether a higher goal has an effect on performance in the second stage.

To note, for all the analyses above, we will run both a basic specification which only controls for the treatment factors and design factors, and a specification that includes various covariates to improve the precision.

5.3. Secondary analysis

5.3.1 Heterogeneous effects

To explore hypothesized heterogeneous treatment effects (outlined in Section 4.3), we interact the treatment variable with baseline covariates of interest.

$$y_i = \alpha + \beta_1 T_i + \beta_2 Z_{i0} + \beta_3 (T_i \times Z_{i0}) + X'_i \gamma + \lambda_s + e_i. \quad (7)$$

5.3.2. Analysis on the mechanism

We will use the outcomes described in section 4.2. as dependent variables in regression specification (3) through (6) to test for the mechanisms described above.

5.4. Standard error adjustments

In our main specifications, we use Huber-White robust standard errors as the treatment is randomized at the individual level. Nevertheless, we will consider two adjustments due to (re-)randomization in the experiment design and multiple hypothesis testing.

5.4.1 Adjusting standard errors due to (re-)randomization

Banerjee et al. (2017) present simulation results showing that re-randomization has a very small effect on inference. Simulated treatment coefficients and standard errors from regression are “substantially the same” than those from randomization and re-randomization inference estimations (Figure C.5). However, we will also use randomization inference as an additional randomization test.

5.4.2 Adjusting standard errors due to multiple hypothesis testing

To address the multiple hypothesis testing issues, we adopt two approaches. First, we divide our outcomes of interests (as described in section 3.2 and 4.2) into 6 categories - inputs campaign related behaviors, demonstration plot and training campaign related behaviors, tree campaign related behaviors, attitude outcomes, general behavior outcomes, and knowledge outcomes. The two approaches that we are considering are creation of indices and adjustments of *p* values/*q* values.

Index:

- Among the 6 categories listed above, for each category that has 3 or more outcome variables, we will construct an inverse covariance weighted summary index, following Anderson (2008).
- Moreover, we will construct two overall summary indices using all downstream outcomes and all mechanism outcomes, respectively. To note, within these specifications, we will control the number of categories that are relevant for FPs.

Adjustments of *p* values/*q* values:

- When looking at the overall downstream outcome index and its corresponding indices of each category/setting, or looking at the overall mechanism outcome index and its corresponding indices, we control the family-wise error rate within these 2 sets of outcomes using the bootstrap resampling method (Romano and Wolf, 2016). We will report both the "naive" p values and Romano-Wolf step-down adjusted p values.
- When exploring the treatment effect on individual behavior/mechanism outcomes as an exploratory exercise, we will control for the false discovery rate following Anderson (2008), which is less conservative than adjusting for family-wise error rate, giving us better statistical power. We will report both the "naive" *p* values and sharpened *q*-values.

5.5. Survey attrition

We expect there will be attrition from the baseline survey to the endline survey. We will examine whether there is differential attrition related to the treatment status and baseline characteristics. We will conduct the following estimation,

$$A_i = \alpha + \beta_1 T1_i + \beta_2 T2_i + \beta_3 T3_i + \beta_4 (T1_i \times X_{i0}) + \beta_5 (T2_i \times X_{i0}) + \beta_6 (T3_i \times X_{i0}) + X'_{i0}\gamma + \lambda_s + \varepsilon_i , \quad (8)$$

where A_i is an indicator of whether FP i attrites from baseline to endline, T_i is the treatment status of FP i, and X_i is a vector of FPs' baseline characteristics, and λ_s represents the stratification block fixed effects. We will also perform a joint test of the treatment and baseline characteristics interaction terms.

If the attrition is not significantly correlated with the treatment status at the 5 percent significance level or is at a reasonable magnitude (i.e., ≤5%), then we will proceed our treatment analysis without any adjustment for attrition. If the attrition is found to be significantly correlated with the treatment status, we will follow Lee (2009) and compute the bounds on treatment effects. If the attrition is found to be significantly correlated with baseline characteristics, we will correct for attrition using inverse probability weights (Baulch and Quisumbing 2011).

5.6. Robustness checks

We plan to explore several robustness checks to our analysis.

- For several outcome variables, we will have both self-reported values and values from administrative records. While we use the administrative data as our main outcomes, we will also analyze the self-reported data as robustness checks.

- For outcomes that are continuous variables, our preferred specification is winsorization at 99th percentile. We will also report results using unwinsorized values and winsorization at 95th percentile as robustness checks.
- In our preferred specification, we will select control variables based on economic literature. We will also run parallel regressions using controls as chosen by the post-double-selection LASSO procedure elucidated in Belloni et al. (2104).

6. References

Michael, Anderson. 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry preschool, and Early Training Projects.” *Journal of the American Statistical Association*, 103 (484): 1481-1495.

Banerjee, Abhijit, Esther Duflo, Amy Finkelstein, Lawrence F. Katz, Benjamin A. Olken and Anja Sautmann. 2020. “In Praise of Moderation: Suggestions for the Scope and Use of Pre-analysis Plan for RCTs in Economics.” *National Bureau of Economic Research Working Paper Series*, No. 26993.

Banerjee, Abhijit, Sylvain Chassang, Sergio Montero, and Erik Snowberg. 2017. “A Theory of Experimenters: Robustness, Randomization, and Balance.” *National Bureau of Economic Research Working Paper Series*, No. 23867.

Baulch, Baulch, and Agnes Quisumbing. 2011. “Testing and Adjusting for Attrition in Household Panel Data.” *CPRC Toolkit Note*.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. “Inference on Treatment Effects after Selection among High-Dimensional Controls”. *The Review of Economic Studies*, 81 (2): 608-650.

Cassar, Lea, and Stephan Meier. 2018. “Nonmonetary Incentives and the Implications of Work as a Source of Meaning.” *Journal of Economic Perspectives* 32 (3): 215–38. https://doi.org/10.1257/jep.32.3.215.

Cialdini, Robert B. 2001. "The science of persuasion." *Scientific American* 284, no. 2: 76-81.

Kaptein, Maurits, Boris De Ruyter, Panos Markopoulos, and Emile Aarts. 2012. “Adaptive Persuasive Systems: A Study of Tailored Persuasive Text Messages to Reduce Snacking.” *ACM Transactions on Interactive Intelligent Systems* 2 (2): 1–25. https://doi.org/10.1145/2209310.2209313.

Laajaj, Rachid, Karen Macours, Daniel Alejandro Pinzon Hernandez, Omar Arias, Samuel D. Gosling, Jeff Potter, Marta Rubio-Codina, and Renos Vakis. 2019. “Challenges to Capture the Big Five Personality Traits in Non-WEIRD Populations.” *Science Advances* 5 (7): eaaw5226. https://doi.org/10.1126/sciadv.aaw5226.

Lee, David S. 2009. “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects.” *The Review of Economic Studies* 76 (3): 1071–1102. https://doi.org/10.1111/j.1467-937X.2009.00536.x.

Romano, Joseph P., and Michael Wolf. 2016. “Efficient Computation of Adjusted P-Values for Resampling-Based Stepdown Multiple Testing.” *Statistics & Probability Letters* 113 (June): 38–40. https://doi.org/10.1016/j.spl.2016.02.012.

Rubenstein, Alex L., Marion B. Eberly, Thomas W. Lee, and Terence R. Mitchell. 2018. “Surveying the Forest: A Meta-Analysis, Moderator Investigation, and Future-Oriented Discussion of the Antecedents of Voluntary Employee Turnover.” Personnel Psychology 71 (1): 23–65. https://doi.org/10.1111/peps.12226.

Appendix 1. Balance Checks

Table 1: Balance checks on farmer promoters' demographics

	(1) Age	(2) Male	(3) Years as FP	(4) Years in Village	(5) Survey Version
Control group × Goal reminder	0.044	-0.003	-0.177	-0.371	-0.002
	(0.392)	(0.017)	(0.147)	(0.695)	(0.037)
Random group × No goal reminder	-0.133	-0.001	0.080	0.252	-0.015
	(0.389)	(0.018)	(0.150)	(0.730)	(0.042)
Random group × Goal reminder	0.435	-0.011	-0.060	-0.154	-0.034
	(0.356)	(0.016)	(0.145)	(0.744)	(0.038)
Match group × No goal reminder	0.308	-0.013	0.015	-0.468	-0.008
	(0.420)	(0.017)	(0.140)	(0.681)	(0.042)
Match group × Goal reminder	-0.314	-0.013	0.019	-0.666	0.009
	(0.347)	(0.019)	(0.150)	(0.685)	(0.042)
Mismatch group × No goal reminder	0.597	0.009	-0.144	0.200	0.027
	(0.365)	(0.018)	(0.132)	(0.700)	(0.041)
Mismatch group × Goal reminder	0.034	-0.014	-0.041	-0.609	0.032
	(0.406)	(0.018)	(0.136)	(0.709)	(0.039)
N	10187	10187	10187	10187	10187
Mean of control w/o goal group	46.863	0.784	6.694	34.946	1.866
p-value of joint F-test	0.171	0.838	0.364	0.716	0.603

Note: The sample includes all farmer promoters that responded to the baseline survey. All variables are measured in the baseline. All regressions include randomization strata fixed effects. Robust standard errors in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

Table 2: Balance checks on farmer promoters' past performance and future goal

	2021A Performance			2022A Goal		
	(1) Inputs	(2) Sessions	(3) Farmers	(4) Inputs	(5) Sessions	(6) Farmers
Control group × Goal reminder	-0.625	-3.641	-0.070	-0.050	-0.086	-1.419
	(2.664)	(3.538)	(0.059)	(0.069)	(1.312)	(2.187)
Random group × No goal reminder	6.088**	4.794	-0.019	0.005	1.947	1.698
	(2.632)	(3.470)	(0.064)	(0.079)	(1.361)	(2.016)
Random group × Goal reminder	2.867	1.071	-0.062	-0.109	-2.155	-5.838***
	(2.428)	(3.213)	(0.060)	(0.076)	(1.322)	(1.910)
Match group × No goal reminder	2.323	0.015	-0.076	-0.059	-1.042	-3.642*
	(2.754)	(3.610)	(0.065)	(0.077)	(1.323)	(2.025)
Match group × Goal reminder	3.412	3.963	-0.090	-0.033	-0.127	-0.557
	(2.635)	(3.603)	(0.060)	(0.079)	(1.399)	(2.114)
Mismatch group × No goal reminder	0.645	-4.279	-0.018	-0.049	-0.097	-2.038
	(2.760)	(3.429)	(0.063)	(0.082)	(1.452)	(2.057)
Mismatch group × Goal reminder	8.122***	8.570***	0.001	0.054	1.170	-1.179
	(2.443)	(3.224)	(0.062)	(0.078)	(1.272)	(1.903)
N	10181	10183	9967	9886	9918	10153
Mean of control w/o goal group	95.873	133.196	3.525	4.202	58.098	82.906
p-value of joint F-test	0.012	0.002	0.638	0.431	0.067	0.004

Note: The sample includes all farmer promoters that responded to the baseline survey. All variables are measured in the baseline. All regressions include randomization strata fixed effects. Robust standard errors in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

Table 3: Balance checks on farmer promoters' personality traits

	(1) Reciprocity	(2) Scarcity	(3) Commitment	(4) Consensus
Control group × Goal reminder	-0.018	0.024	0.011	0.006
	(0.021)	(0.028)	(0.018)	(0.028)
Random group × No goal reminder	-0.015	0.010	0.006	-0.008
	(0.024)	(0.030)	(0.017)	(0.025)
Random group × Goal reminder	-0.014	0.031	0.017	-0.017
	(0.020)	(0.031)	(0.017)	(0.022)
Match group × No goal reminder	-0.023	0.018	-0.005	-0.041
	(0.022)	(0.027)	(0.017)	(0.026)
Match group × Goal reminder	-0.031	0.028	0.006	-0.010
	(0.021)	(0.029)	(0.018)	(0.027)
Mismatch group × No goal reminder	0.003	0.031	0.018	-0.011
	(0.022)	(0.029)	(0.017)	(0.025)
Mismatch group × Goal reminder	-0.015	0.025	0.006	-0.028
	(0.022)	(0.031)	(0.016)	(0.025)
N	10187	10187	10187	10187
Mean of control w/o goal group	5.855	4.903	6.121	5.646
p-value of joint F-test	0.754	0.988	0.795	0.588

Note: The sample includes all farmer promoters that responded to the baseline survey. All variables are measured in the baseline. All regressions include randomization strata fixed effects. Robust standard errors in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

Table 4: Balance checks on farmer promoters' demographics

	(1) Age	(2) Male	(3) Years as FP	(4) Years in Village	(5) Survey Version
Random group	0.128	-0.005	0.099	0.235	-0.023
	(0.255)	(0.012)	(0.097)	(0.508)	(0.027)
Match group	-0.025	-0.012	0.106	-0.381	0.002
	(0.294)	(0.013)	(0.101)	(0.488)	(0.030)
Mismatch group	0.293	-0.001	-0.004	-0.019	0.031
	(0.284)	(0.012)	(0.094)	(0.504)	(0.030)
N	10187	10187	10187	10187	10187
Mean of control group	46.949	0.781	6.600	34.842	1.865
p-value of joint F-test	0.446	0.669	0.472	0.402	0.135

Note: The sample includes all farmer promoters that responded to the baseline survey. All variables are measured in the baseline. All regressions include randomization strata fixed effects. Robust standard errors in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

Table 5: Balance checks on farmer promoters' past performance and future goal

	2021A Performance			2022A Goal		
	(1) Inputs	(2) Sessions	(3) Farmers	(4) Inputs	(5) Sessions	(6) Farmers
Random group	4.793***	4.756**	-0.005	-0.027	-0.053	-1.366
	(1.659)	(2.121)	(0.045)	(0.055)	(0.948)	(1.251)
Match group	3.180*	3.810	-0.048	-0.021	-0.540	-1.388
	(1.804)	(2.392)	(0.042)	(0.054)	(0.957)	(1.348)
Mismatch group	4.699**	3.974*	0.027	0.027	0.582	-0.898
	(1.846)	(2.177)	(0.044)	(0.056)	(0.921)	(1.243)
N	10181	10183	9967	9886	9918	10153
Mean of control group	96.169	131.917	3.493	4.181	58.210	82.405
p-value of joint F-test	0.648	0.903	0.248	0.593	0.448	0.918

Note: The sample includes all farmer promoters that responded to the baseline survey. All variables are measured in the baseline. All regressions include randomization strata fixed effects. Robust standard errors in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

Table 6: Balance checks on farmer promoters' personality traits

	(1) Reciprocity	(2) Scarcity	(3) Commitment	(4) Consensus
Random group	-0.005	0.008	0.006	-0.015
	(0.019)	(0.021)	(0.012)	(0.016)
Match group	-0.018	0.011	-0.005	-0.028
	(0.016)	(0.019)	(0.013)	(0.019)
Mismatch group	0.003	0.016	0.006	-0.023
	(0.017)	(0.020)	(0.012)	(0.018)
N	10187	10187	10187	10187
Mean of control group	5.846	4.913	6.127	5.648
p-value of joint F-test	0.362	0.904	0.510	0.719

Note: The sample includes all farmer promoters that responded to the baseline survey. All variables are measured in the baseline. All regressions include randomization strata fixed effects. Robust standard errors in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

Appendix 2. Timeline