Pre-analysis plan

Irene Furlani, Michael Grimm, Alia Bihrajihant Raya, Ishadiyanto Salim,
Nadila Puspa Arum Widjanarko, Subejo, Ratih Ineke Wati,
Anung Pranyoto, Rani Nur Rochim, Siti Fatonah

April 17, 2025

1 Trial title

Cultivating Connectivity: Measuring the Impact of a Digital Agricultural Extension Platform on Millennial Farmers and Extension Workers in Indonesia (LenteraDigiEx)

2 Country

Indonesia

3 Status

Ongoing

4 Keywords

Digital platform, agriculture, extension workers, millennial farmers, digital literacy, technology adoption, Indonesia, sustainable agricultural practices

5 Abstract

To attract Millennial Farmers (MFs) into an aging agricultural sector, Indonesian Extension Workers (EWs) must match the digital literacy level and offer digital solutions to their pool of farmers. Moreover, digital platforms offer a solution to reduce the cost of capacity building and activity support of EWs, as well as information and services for MFs. We assess the impact of different double side interventions (EW capacity building and MF information session) on the knowledge and adoption of the digital platform LenteraDESA with an experimental design. On the supply side, we examine the causal impact of LenteraDigiEx on digital literacy, extension capacity, performance, technology perception, and attitude of EWs. On the demand side, we consider the impact on digital literacy, income, loans and investments, employment, and sustainable agricultural and business practices. We hypothesize that the interventions will increase digital literacy, platform adoption, and usage. The use of the LenteraDESA platform may increase the EWs' performance and capabilities, and may also increase MFs' knowledge and use of sustainable agricultural and business practices, and ultimately improve MFs' overall welfare.

6 Trial Start Date

January 2025

7 Intervention Start Date

May 2025

8 Intervention End Date

May 2025

9 Trial End Date

March 2026

10 Outcomes

This pre-analysis plan was finalized prior to the availability of any outcome data related to our training intervention. The implementation of alternative training schemes as a randomized controlled trial (RCT) enables the identification of causal effects of these training options on digital literacy and platform uptake, training uptake behavior, and working life satisfaction for both respondent types, performance and capability of EWs, and individual welfare of MFs. Primary outcomes are reported in bold.

10.1 Shared outcomes

Digital Literacy and platform uptake

- **What is the digital literacy level of the respondent?** (composite index - primary outcome)
  - What is the level of digital information and data literacy of the respondent? (composite index)
  - What is the level of digital communication and collaboration literacy of the respondent? (composite index)
  - What is the level of digital content creation literacy of the respondent? (composite index)
  - What is the level of digital safety literacy of the respondent? (composite index)
  - What is the level of digital problem solving literacy of the respondent? (composite index)
  - What is the level of digital knowledge of the respondent? (composite index)
  - What is the level of digital attitude of the respondent? (composite index)
  - What is the level of digital skills of the respondent? (composite index)
- **Does the respondent know about LenteraDESA?** (primary outcome)
- **Does the respondent use LenteraDESA?** (primary outcome)

Training uptake behavior Within these outcomes, a special focus will be on the training topic of the intervention (digitization in agriculture).

- Has the respondent attended any training in the last six months?
- Has the respondent obtained any certificate for a training in the last six months?
- How many types of training did the respondent attend in the last six months?
- Has the respondent attended any training that covers the use of digital technology in supporting agricultural extension in the last six months?
- How many types of training on digital technology use did the respondent attend in the last six months?

Working life satisfaction

- What is the level of working life satisfaction of the respondent?

10.2 Outcomes only related to EWs

Performance and capability of extension workers

- Have the training sessions helped improve the respondents' ability to use digital technology for extension services?
- What is the respondent's level of acceptance of digital technology in extension activities?
- What is the level of personal digital innovation of the respondent?
- **Has the respondent integrated digital tools into activities in the last 12 months?** (primary outcome)
- Has the respondent integrated LD into their activities in the last 12 months?
- How many extension activities did the respondent perform during the last 12 months? (in total -primary outcome- and by type of activity)
- **Did the intervention improve the respondent's capability?** (primary outcome)
  - How many farmers does the respondent reach?
  - How many MFs does the respondent reach?
  - How many acres does the respondent supervise?

10.3 Outcomes only related to MFs

Welfare of MFs

- **How much income did the respondent get from their agricultural activities in the last 12 months?** (primary outcome)
- How many agricultural and business practices does the respondent know?
- How many agricultural and business practices does the respondent use?
- How many sources of agricultural and business practices does the respondent know?
- Does the business of the respondent have any level of formalization?
- How many certification types does the business of the respondent have?
- Is the respondent a member of a farmer group?
- Is any (other) household member part of a farmer group?
- Did the respondent have any loan in the last 12 months?
- How many loans did the respondent have in the last 12 months?
- What is the total loan amount of the(se) loans?
- Is there a loan from a formal institution?
- How many sources of loans does the respondent have?
- What is/are the purpose(s) of the(se) loan(s)?
- How many Rupiees has the respondent invested in their business in the last 12 months?
- Does the respondent currently employ any worker?
- How many people does the respondent currently employ?
- How many family members that are not paid for work does the respondent currently have?

Within the agricultural and business practices outcomes, a special focus will be given to the topics of the intervention (agricultural digitization, social media marketing, e-commerce, and digital farm recording).

11 Experimental Design

11.1 Sampling

Our sample of respondents is drawn from 51 subdistricts that are comparable to agricultural extension service offices[1], located in five regencies in the Yogyakarta special region of Indonesia. The 51 subdistricts cover almost the entire area under analysis. The covered area is incomplete as five offices have been excluded from the analysis since they reach at most two MFs. The reasons for the exclusion are implementation efficiency and cost-benefit.

The planned proportional random sample consists of 839 MFs and 171 EWs drawn from a sampling frame of 1,624 MFs and 303 EWs. The sampling frame comes from administrative data collected by the agriculture and food security office of the local government[2] and consists of governmental EWs and registered MFs. We selected the sample to have a similar number of respondents across treatment arms to increase statistical power. Therefore, we first considered 50% of the individuals to be distributed across the three treatment arms (two treatment groups and one control), such that

$$R_T = \frac{N * 0.5}{3}$$

where R_T is the ideal number of respondent per treatment group, and N is the total number of individuals in the sampling frame. In a second step, we select the number of respondent per subdistrict as explained in the following equation:

$$ceil\left(R_{TS}\right) = \frac{N_{TS}}{N_T} * R_T$$

In particular, the number of selected respondents R in a subdistrict S within a treatment arm T is rounded up integer of the proportion of individuals in a subdistrict of a given treatment arm N_{TS} over the total population of that treatment arm N_T, times the total number of respondents in a given treatment arm R_T. We rounded up all decimal points to select the number of respondents within each subdistrict, to ensure at least the ideal number of respondents per treatment arm to be reached.

During the baseline data collection (February 2025), we realized that the reliability of the MFs' data was low, as many the potential respondent only attended one or more events held by BPSDMP, therefore, we requested additional data from the local agricultural department (DINAS). After randomizing the potential replacements from this list, we obtained a final number of respondent of 784 MFs and 170 EWs. As we plan two additional waves, we aim at obtaining 2,862 observations from 954 respondents.

11.2 Data collection

Data collection is planned in three phases: Baseline, midline, and endline. These three phases will occur at the same time for both respondent types. Baseline data collection ended in March 2025. Midline data collection is planned for October 2025, and endline is planned for February and March 2026.

11.3 Interventions

The intervention will consist of two treatment groups. In each treatment group, we will implement two types of training interventions: One for MFs and one for EWs. The one for MFs will include information on digital literacy, the LenteraDESA platform, and its new feature for farm recording[3]. The training for EWs will include information on how to integrate digital platforms and digital content into their activities in addition to the other topics. In the first treatment group, the 'light' training, the

[1] The subdistrict level coincides most of the time with the same administrative categorization, except for the regency of Sleman where the administrative subdistricts are grouped into 8 offices.

[2] Precisely, it comes from the *Unit Pelaksana Teknis Daerah (UPTD) Balai Pengembangan Sumber Daya Manusia Pertanian (BPSDMP) Dinas Pertanian dan Ketahanan Pangan DIY* that is the Regional Technical Implementation Unit (UPTD) of the Agricultural Human Resources Development Center (BPSDMP) of the Agriculture and Food Security Office of the special region of Yogyakarta (DIY).

[3] The characteristics of the farm record (in Bahasa, *Catatan Tani*) have already been defined, but the technical implementation is still under development.

recipients will be offered an online session, while the second treatment group, the 'intensive' training, will replicate the same session in-person(offline). Both types of treatment will be followed by a self-paced online training on LenteraDESA, covering various topics related to digitization in agriculture and digital solutions for agriculture. The training will be concluded with an online presentation by the participants who will obtain a certificate of completion at the end of the course. MFs will also be able to use the digital farm record for loan requests.

11.4 Randomization procedure

We applied a stratified random assignment with misfit corrections based on the distribution of EWs across subdistricts. First, we considered the distribution of EWs in our sampling frame and created strata based on whether the number of EWs in a subdistrict is above the median number of EWs per subdistrict. Then, we randomly assigned one-third of the subdistricts (17) to each treatment arm (two treatment groups and one control), making sure that misfit corrections were globally applied based on the distribution of EWs in the subdistricts.

11.5 Spillovers

We do not expect major spillovers in terms of training participation, as the intervention is based on subdistrict level and is only offered to the selected respondents (due to budget constraints), preventing EWs from sharing their freshly acquired knowledge with MFs outside of their working subdistrict area. Moreover, access to self-paced training will be first offered only to selected respondents[4]. Still, given that digital platforms are nonexclusive[5], the respondents in the control group could access the platform themselves. Therefore, the control group can become aware of the platform and access its general content, learning about those agricultural and business practices that might affect their behavior. As we plan to analyze intent-to-treat effects, the impact of the training effect on these outcomes will only be underestimated.

11.6 Model specification

The primary objective of this analysis is to estimate the impact of the intervention on the outcome variables measured during the follow-up surveys using an ANCOVA estimator. Given the random nature of the treatment assignment, we can use an ANCOVA estimator to increase power by controlling for baseline levels of the outcome variables. The analysis will be based on different specifications based on the outcome types (shared outcomes, EWs' outcomes, and MFs' outcomes). **Shared outcomes**

$$Y_{ist} = \beta_0 + \beta_1 Light_s + \beta_2 Intensive_s + \beta_3 Farmer_{is} + \beta_4 Y_{is0} + X'_{is0}\gamma + \epsilon_{st} \quad t = 1, 2 \tag{1}$$

where Y_{ist} is the post-treatment outcome for respondent i in subdistrict s at time t. Depending on the outcome type (short-term vs mid-term outcome, t will be equal to 1 for the first follow up survey (i.e., midline), or equal to 2 for the endline data collection (second follow up). $Light_s$ is the treatment indicator equal to 1 if a subdistrict s received the 'light' training treatment. Similarly, $Intensive_s$ is the treatment indicator equal to 1 if a subdistrict s received the 'intensive' training treatment. β_1 and β_2 are the effects of the respective treatment type relative to the control group. $Farmer_{is}$ is an indicator equal to 1 if the respondent is a MF. β_3 is the average difference in the outcome variable when the respondent is a MF, compared to an EW. Y_{is0} are the values of baseline outcomes (i.e., at time zero) like initial digital literacy and platform uptake, training uptake behavior and working life satisfaction. X_{is0} is a vector of control variables (age, gender, education, marital status, household size) and ϵ_{st} is the error term. As explained before, the treatment status is assigned at the subdistrict level, therefore the standard errors are clustered at the subdistrict level. **Individual level outcomes (EWs and MFs)**

$$Y_{ist} = \beta_0 + \beta_1 Light_s + \beta_2 Intensive_s + \beta_3 Y_{is0} + X'_{is0}\gamma + \epsilon_{st} \quad t = 1, 2 \tag{2}$$

[4]Please note that the training options will only be accessible to the selected respondents in treated subdistricts, while the platform features will be accessible to everyone.

[5]Digital platforms are nonexclusive, except when: (i) an area lacks digital infrastructure (ii) an individual lacks digital tools, has no (or low) digital literacy, or has financial constraints when the platform is not for free.

For individual type-specific outcomes, the specification is similar. There is no MF indicator, but there are additional covariates that are part of the vector X_{is0}. For EWs, it will also include working experience, while for MFs, it will also include a measure of assets as a proxy for wealth, additional household income, entrepreneurship ability. As before, Y_{is0} will include the baseline outcomes: For example, for MFs analysis, it can include initial welfare measures (e.g.,income level, or employment level), while for EWs analysis it can include baseline values of performance and capability.

Alternative specifications will also include: Regency fixed effects and subsector fixed effects (crop cultivation, livestock, fishery, forestry, agro-processing). We will consider heterogeneous effects by gender, agricultural subsector, age, education, and initial dicital literacy level.

In particular, for heterogeneous effects based on gender, the specification will be the following:

$$\begin{aligned} Y_{ist} =& \beta_0 + \beta_1 Light_s + \beta_2 Intensive_s + \beta_3 Light_s * Female_{is} + \beta_4 Intensive_s * Female_{is} + \\ &+ \beta_5 Y_{is0} + X'_{is0}\gamma + \epsilon_{st} \quad t = 1, 2 \end{aligned}$$

where the coefficients β_3 and β_4 will capture the additional effect of being in a treated subdistrict when the respondent is female. Please note that the gender indicator $Female$ will be included in the vector X'_{is0}. A similar specification will apply for the heterogeneous effect by the agricultural subsector.

For those outcomes for which we do not have baseline information, we estimate simple models comparing treatment to control at t=1, t=2 and pooled.

11.7 Multiple hypothesis testing

In our study, there is a high probability of falsely rejecting the null hypothesis due to the large number of outcomes. Therefore, to ensure that our findings are robust to p-value adjustments for multiple hypothesis testing, we will control for the False Discovery Rates as in Benjamini and Hochberg (1995). Specifically, we will calculate sharpened q-values within each family of outcomes.

11.8 Cost-benefit analysis

We plan to perform a cost-benefit analysis to inform the local government about the value of a potential scale-up of the project.

12 Was the Treatment Clustered?

Yes

13 Planned Number of Clusters

51

14 Planned Number of Observations

The planned total of 2,862 observations (2,352 for MFs and 510 for EWs) from three survey waves from 170 EWs and 784 MFs. If attrition rates are greater than 10% and we find evidence of differential attrition by treatment status, we will estimate the pairwise Lee bounds for our treatment effects.

14.1 Power analysis

We updated our power calculations using our baseline data collection for our composite digital literacy indicator (scale 0-4) based on a 95% confidence interval and a power of 80%. The baseline data shows a mean of digital literacy of 1.74 scale points, with a standard deviation of 0.49 and an intra-cluster correlation (ICC) of 0.069. Given our average cluster size of 18 respondents per cluster and 17 clusters per treatment arm, we are powered to detect a minimum effect of 0.165. For a separate analysis, we can detect an effect of 0.189 for MFs and 0.205 for EWs.

15 IRB approvals

We obtained ethical clearance from multiple offices:

- IRB approval from the Ethics Committee on Social Studies and Humanities National Research and Innovation Agency (BRIN) on January 8th, 2025 with a letter on January 10th, 2025. It followed the application number 12122024000015. The approval number is 017/ KE.01/SK/01/2025.
- Approval from the ethics commission for research of the University of Passau on January 28th, 2025 with procedure number III/GRIMM.I-07.5095/250128.
- Internal review of the Digital Platform Ecosystem (DPE) research training group management on December 19th, 2024.

References

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: series B (Methodological)*, 57(1):289–300, 1995.