CENTER FOR GLOBAL DEVELOPMENT

The Unfolding of Women's Economic Empowerment Outcomes

TIME PATH OF IMPACTS IN AN INDONESIA TRIAL

Mayra Buvinic, James C. Knowles, and Firman Witoelar

Abstract

This paper makes the case that some interventions designed to improve women's economic lives need to be tracked long enough for women to manifest new and beneficial behaviors. The study analyzes the time paths of the predicted impacts in a randomized trial providing financial incentives to bank agents (on the supply-side) and basic business training to women business owners (on the demand-side) to facilitate their access to and use of formal financial services. The trial took place over a two-year period in 401 villages in five regencies of East Java province, Indonesia. Although predicted impacts of the treatments are positive and increasing over time for seven of eight economic empowerment outcomes analyzed, their time paths vary, with business profits and household income increasing sharply in the second year after increasing very little during the first year while business capital, improved business practices, personal savings and women's agency increased linearly with time following treatment. A mediation analysis using a structural equations model finds significant direct and indirect effects between these economic empowerment-related outcomes: business practices, business capital and women's agency contribute directly to increased business profits while personal savings, business profits itself and household income are indirect drivers; further, both treatments had a strong direct impact on household income that was reinforced by the direct effects of increased business profits. The article's findings are consistent with other incipient evidence on the delayed effects of some interventions targeting women's economic empowerment. They underline the need to measure impacts over a sufficiently long period, which in this case is at least two years after exposure to the treatments.

KEYWORDS
women's economic empowerment, financial inclusion, entrepreneurship, and agency

JEL CODES
J16, B54, L26, and D91

This paper was originally published in June 2021. It was updated in June 2022. The original version can be viewed at https://www.cgdev.org/sites/default/files/ unfolding-womens-economic-empowerment-outcomes-time-path-impacts-indonesia-trial.pdf

The Unfolding of Women's Economic Empowerment Outcomes: Time Path of Impacts in an Indonesia Trial

Mayra Buvinic
Center for Global Development

James C. Knowles

Firman Witoelar
Crawford School of Public Policy, Australian National University

This research was supported by generous grants from the ExxonMobil Foundation to the Center for Global Development and from the Umbrella Facility for Gender Equality (UFGE) in partnership with the Australian Department of Foreign Affairs and Trade to the World Bank Group's East Asia and Pacific Gender Innovation Lab. Special thanks go to James M. Jones for his steadfast support of this work, to Mead Over for his many insightful comments, and to five anonymous readers for their careful reviews. We are also grateful to Erika Deserrano, Hillary C. Johnson, Gianmarco León-Ciliotta, Elizaveta Perova and our partners J-PAL SEA, Mercy Corps Indonesia Mitra Saniya and Survey Meter for their substantive contributions to the research trial on which this paper is based. We are especially grateful to Glory Sunarto, Andrea Adhi, Marifatual Amalia, Jenna Juwono, Lolita Moorena, Michael Tjahjadi, and Irwan Setyawan for their excellent field work.

Mayra Buvinic, James C. Knowles, and Firman Witoelar. 2022. "The Unfolding of Women's Economic Empowerment Outcomes: Time Path of Impacts in an Indonesia Trial." CGD Working Paper 586. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/unfolding-womens-economic-empowerment-outcomes-time-path-impacts-indonesia-trial

CENTER FOR GLOBAL DEVELOPMENT
2055 L Street, NW Fifth Floor
Washington, DC 20036
202.416.4000
www.cgdev.org

Contents

List of Figures

List of Tables

1. Introduction and literature review

1.1 Introduction

This paper is based on a randomized controlled trial in rural Indonesia designed to evaluate the impacts of alternative approaches to increasing women business owners' (WBO) access to and utilization of village-based mobile banking services, particularly savings accounts. The trial measured the impacts of both supply-side (agent incentives) and demand-side (basic financial literacy and business training) treatments on WBOs' business outcomes (profits, capital and labor inputs, business practices), on their personal savings, and on their agency and household income. On the supply-side, higher agent incentives were expected to motivate village-based bank agents to reach out to women business owners and assist them in opening and using individual savings accounts that would afford them increased privacy and agency as users of financial services. On the demand-side, the financial literacy and business training (hereafter referred to as "business training") were expected to improve WBOs' business practices and motivate them to save more. The results indicate that these expected effects occurred and were reinforced by substantial indirect effects that together led to positive impacts on WBOs' business practices, personal savings, business capital, women's agency, business profits and household income that increased significantly over time.

Due to delays in recruiting suitable bank agents, implementation of the treatments occurred over a 15-month period. As a result, there is considerable variation in the durations between first exposure to the treatments and the time when their effects were measured. This unanticipated feature of the data is exploited in this paper to investigate the time paths of the trial's impacts. This is important for several reasons. First, it provides information about whether the estimated impacts can be expected to continue, diminish or increase in the future, which affects the trial's economic evaluation. Second, knowledge of the time paths of impacts can be helpful in deciding when to measure impacts in similar trials (King and Behrman 2009). Third, the time paths of the impacts can help us to understand the process by which the impacts occurred. For example, to what extent were direct treatment effects accompanied by indirect effects that also contributed to the observed impacts.

The results indicate that the impacts on all outcomes, with the exception of the WBO's own labor inputs (for which there was no significant impact), increased with the duration of time since first exposure to the treatments. The time patterns of the increases are not uniform across the outcomes. In the case of business profits and household income, for example, the predicted impacts were small and insignificant during the first year following treatment but increased sharply in the second year.

1.2 Review of the literature

Along with other literature, we define empowerment as the expansion of agency and measure it by the WBO's reported say in household decision making, a commonly used measure of empowerment

that captures the exercise of agency or autonomy in financial decision making (Sen 1999; Donald and colleagues 2020; Laszlo and colleagues 2017). Increased agency mediates economic outcomes and economic empowerment measures should capture both the expression of agency and the resulting economic achievements (Kabeer 1999). Recent empirical work has shown that women's economic empowerment is boosted by both the financial privacy of individual bank accounts (Garz and colleagues 2020) and the role of savings in reducing pressures or obligations to share resources with others (Dupas and Robinson 2013; Hoff and Stiglitz 2016).

The original trial was designed to explore whether supply-side incentives that reward agents' good performance in terms of reaching women clients would help overcome potential gender biases in the business environment and increase women's uptake of financial services. To our knowledge there are no trials testing the effectiveness of supply-side incentives in increasing women's access to and use of financial services. Contrasting with the lack of evidence on supply-side interventions, on the demand-side a body of evidence suggests that only well-designed, high-quality financial literacy and business management training (with high attendance and low attrition rates) can help overcome information and skills deficits that can be particularly onerous for businesswomen, improve business practices and increase business profits (McKenzie and Woodruff 2015; McKenzie 2017; 2021).

The main question of this analysis, the trajectory of impacts over time, complements an incipient literature on the path of treatment effects on women's economic empowerment outcomes showing both short-lived and delayed effects. For example, Banerjee, Karlan and Zinman (2015) found that an increase in women's empowerment (measured through reported influence on household decision making) as a result of a capital transfer and complementary training as part of "graduation programs" for the very poor in six countries, picked up in most programs, dissipated within a year. Similarly, in Sri Lanka, a combination of business training and cash grants led to large and significant increases in WBO's profits during the first eight months post-treatment that largely disappeared in the second year (De Mel, McKenzie and Woodruff 2014).

In contrast, recent business training trials targeting micro-entrepreneurs have shown positive effects of business training on profits and/or sales, but with a significant delay. In Peru, WBOs who only received training caught up in terms of increased sales with those who also received technical assistance but only in the second year (Valdivia 2015). In South Africa, an intensive training course for business owners (45% of them female), increased profits significantly, but only after a year (Anderson-MacDonald and others 2018). In Kenya, an ILO training and mentoring program for WBOs showed significant positive effects on profits and sales only three years after the training (McKenzie and Puerto 2017). In Vietnam, another ILO training and mentoring project observed significant effects on women's profits and personal empowerment (internal control and intra-household decision making) after one year (Huis, Lensink, Vu and Hansen 2019). In Indonesia, business training employing local best practices for retailers, 70% of whom are women, combined with two

thirty minutes-long visits did not increase profits six months after the training but increased them significantly after 18 months (Dalton and colleagues 2021).

It makes intuitive sense that increases in the profits of WBO's firms, which on average are considerably smaller and less capital intensive and profitable than male-owned firms, would take time to materialize, especially after relatively 'light' interventions, such as the ones in this trial. One would also expect that interventions that boost women's empowerment would lead to more sustainable increases in profits and other business outcomes.

2. Background and description of the trial

2.1 Background

To promote full financial inclusion, the Government of Indonesia adopted a law in 2014 that authorized the establishment of banking services in the absence of branch offices, called "branchless banking."[1] The Indonesian model of branchless banking uses village-based agents and mobile telephone communications (including both SMS and internet access) to offer basic banking services, including interest-bearing savings accounts with no opening or maintenance fees. The agents are mostly existing shop-owners who are authorized to assist customers to open accounts, accept cash deposits and enable cash withdrawals from customers' savings accounts. Account holders can use their mobile phones to check their account balances, make payments and effect transfers. Only the account holders and their banks have access to account balances, which makes them less vulnerable to demands from spouses, relatives and other social claimants.

Although targeted more generally to its large, mainly rural unbanked population, Indonesia's branchless banking services provide a unique opportunity to improve rural women's access to and use of savings accounts and other formal financial services. Among underbanked rural women, WBOs are a potentially important target group for branchless banking services both because they are substantially under-banked relative to men business owners and possibly partly as a consequence, have lower savings, fewer business assets and lower business incomes (World Bank 2016; Buvinic, Knowles and Witoelar 2021).[2]

2.2 Description of the trial

The trial supported both supply-side and demand-side interventions designed to increase the utilization of saving and other branchless banking services by WBOs with established businesses

1 OJK Regulation No. 19/POJK.03/2014 dated 18 November 2014 ("*Laku Pandai* Regulation").

2 Sample WBOs earn about half as much as men business owners from their primary and secondary businesses and have only about a third of the business assets. WBOs are about half as likely as men business owners to have borrowed money during the past 12 months and their total savings during the last 12 months are about half as large in value, although 84% of WBOs report having saved during the last 12 months, compared to only 69% of men business owners.

in 401 predominantly rural villages of five regencies (*kabupaten*) of East Java province, Indonesia in which the study's partner bank planned to introduce branchless banking services.[3] The interventions were randomized within an experimental design that makes it possible to assess the effects of the supply-side and demand-side interventions, both separately and in combination. In each sample village, the partner bank (with assistance from the trial) recruited a branchless banking agent using the bank's standard selection criteria (the selected agents were in many cases clients with a good credit history).[4]

Sample villages were randomly assigned to high or low incentives, including one agent in each village (48.3% of whom were women). The random assignment of villages to treatments was stratified by village randomization group.[5] One hundred and ninety-six villages/agents (48.9%) were randomly assigned to the high-incentive treatment, while 205 villages/agents (51.1%) were randomly assigned to the low-incentive treatment. The standard agent fee for identifying a new savings account client that deposits at least Rp. 20,000 (about US$1.50) and who maintains an average savings balance of Rp. 20,000 over two weeks is Rp. 2,000 (about US$0.14). This is the "low-incentive treatment" in the trial. The randomly assigned "high-incentive treatment" is Rp. 10,000 (about US$0.71) for each new client enrolled.[6]

The sample WBOs were randomly selected from a village listing (census) of WBOs who were between 18 and 55 years of age, had a currently operating business in a sector other than agriculture, self-identified as the primary owner or manager of the business, resided in the sample village, and had an operational mobile phone. The random assignment of WBOs to the business training was also stratified by the same village randomization groups, resulting in 1,604 WBOs (56.2%) randomly assigned to the business training and the remaining 1,248 WBOs (43.8%) assigned to the control group. 1,401 (1,451) sample WBOs were in villages randomly assigned to receive high (low) agent incentives, 784 of whom received both treatments. Random assignment of the business training within villages raises the possibility of spillover effects from WBOs randomly assigned to receive the treatment to WBOs (primarily in the same village) assigned to the control group. However, the data do not point to the presence of significant spillover effects.[7]

3 The five sample regencies are: Tuban, Bojonegoro, Ngawi, Lamongan and Gresik, with a combined 2010 population of 5.56 million. 354 of the sample villages (88.3%) are rural, while 47 (11.7%) are semi-urban.

4 The standard criteria are: (1) the owner is a previous borrower from the bank, (2) the business is in a central location in the village, (3) the owner is mostly present at the business premises, (4) the owner has a good reputation among villagers (as reported by the village chief), (5) the owner is able to demonstrate sufficient financial liquidity, (6) the owner is not an agent for another bank, and (6) the owner is willing to participate as an agent.

5 The randomization groups were geographical clusters of villages that were recruited at approximately the same time.

6 The trial also included a village-level informational treatment (orthogonal to the main agent incentive treatment) that varied information about the level of incentives agents received in randomly selected villages.

7 For example, 39.3% and 43.2% of the WBOs randomly assigned to receive the business training reported that they "knew about" the partner bank's two e-savings products, compared to only 4.9% and 3.9% respectively of WBOs who were not assigned to receive the business training. By comparison, only 6.3% and 5.8% of the two groups of WBOs respectively "knew about" the most frequently known e-savings product of another bank. Only 9 and 6 respectively of the 58 and 46 WBOs not randomly assigned to receive the business training who knew about the partner bank's two e-savings products, reported that they learned about the products from "friends or relatives."

All agents and randomly selected WBOs were trained by Mercy Corps Indonesia (MCI), an organization with extensive experience providing financial and business literacy training to Indonesian business owners and farmers. The agent and WBO training used separate training manuals based on those used in MCI's previous work with business owners in Indonesia. The training was done by 50 experienced MCI trainers. The agent training was provided at the agents' business sites, while the WBO training was conducted in the village (sub-village) hall or in one of the trainee's business sites/homes. Due to the delays in recruiting agents, both the agent and businesswomen's training were conducted in two batches, the first batch during the period December 2016 to July 2017 and the second batch during the period August 2017 to June 2018. The agent training was one-on-one, while the WBO training was provided to small groups of 4–7 WBOs, including the four randomly selected sample WBOs and (in some villages) up to 3 non-sample WBOs who were permitted to participate in the training.

The randomly assigned business training focused on financial and business literacy (tracking income and expenses, setting priorities, the importance of saving, financial planning, basic bookkeeping, cash flow planning, record-keeping) and on the partner bank's branchless banking products. Apart from marketing (not included) and the focus on saving and use of the branchless banking products (included), the topics covered are typical of those covered in business training trials (McKenzie and Woodruff 2015). The business training was provided in four small group sessions. The initial training session, held most often on the weekends so it would not conflict with businesswomen's regular work schedules, lasted about three hours (ranging from one to four hours) and focused on the content in the training manual, ending with the assignment of homework for review in the remaining three sessions, which focused on the WBOs' actual business practices while addressing any questions from the trainees. The second training session focused on financial planning, including both business and personal applications; the third, on bookkeeping and stocktaking; and the fourth, on challenges related to customer outreach and promotion. Program data indicate that the percentages of WBOs randomly assigned to the business training (N = 1,604) who participated in the initial training session and each of the three follow-up sessions are 94.3%, 84.4%, 83.5% and 82.7% respectively, indicating a relatively high level of compliance.

The agent training was provided to all agents in four personal (one-on-one) sessions that also averaged about three hours in length. The agent training included a module on marketing that emphasized the potential value of marketing to under-banked groups, particularly women.

An endline impact evaluation of the trial found that the interventions were successful in increasing WBO's profits, capital inputs, adherence to recommended business practices, agency in household decision-making, savings, household ownership of durable assets and reduced use of money lenders (Buvinic and colleagues 2020). However, it found little effect of the treatments on the take-up and utilization of the partner bank's branchless banking services.

3. Data

3.1 Description of the data

A Baseline Survey (BS) was conducted in 401 villages in two phases (November 2016–February 2017 in 107 villages; July–November 2017 in the remaining 294 villages). The BS collected extensive data on 2,852 business owners using a single questionnaire that required about 1.5 hours to administer. The first follow-up survey (FS1) of 1,399 WBOs interviewed in the BS was conducted in February 2018 in a non-random sample of 200 villages in the three districts in which the training of agents and WBOs had been completed prior to the survey. The second follow-up survey (FS2) of 2,841 WBOs interviewed in the BS was conducted in all 401 sample villages in November–December 2018.

Because actual implementation of the treatments extended over a 15-month period (December 2016–June 2018), there is considerable variation in the data between the time WBOs were first exposed to the treatments and when their effects were measured in FS1 and FS2. Table 1 shows that the combined data from both follow-up surveys provide 4,240 observations spread over four six-month intervals post-exposure, with the majority of observations in the first three 6-month intervals and relatively few in the last interval of 19–24 months. The analysis in this paper exploits the unique features of this panel data while addressing some of its limitations.

One limitation is endogeneity in the time since first exposure to the treatments in both follow-up survey rounds, which is addressed by obtaining village-level fixed-effects (FE) estimates that are limited to variation *within* village-level panels, effectively removing the effects of any unobserved village-level fixed effects that may have been present during the period of the data collection. Another limitation (possible sample selection bias in FS1, which was conducted in a non-random sample of 200 of the 401 sample villages) is addressed by assessing the robustness of the results by re-estimating the impact models without the FS1 data.

Annex 1 provides detailed information on the eight outcomes analyzed, while Table 4 reports selected statistics on the outcomes by survey round.

3.2 Sample attrition

Both follow-up surveys experienced relatively low rates of overall sample attrition (i.e., 3.9% in FS1 and 4.1% in FS2) when defined as the inability to collect follow-up data from WBOs who were interviewed in the baseline survey. Because the WBO sample was limited to women with established businesses (only 5.5% of the WBOs' primary businesses were less than one year old, while 54.1% were at least five years old), very few of the sample businesses were reported to have closed between the baseline and follow-up surveys. In most cases, sample attrition was due to refusal to respond rather than to inability to be contacted. However, attrition (non-reporting) rates are higher for business

profits (7.0%) and business capital (12.7%).[8] Possible attrition bias in WBO outcomes was assessed by estimating the following linear regression model with data from the two follow-up survey rounds (r = FS1, FS2):

$$A1_{i,j,r} = \beta_{1,0} + \beta_{1,1}High_j + \beta_{1,2}Train_{i,j} + \mathbf{x}_{i,j}\delta_1 + \varepsilon_{1,i,j,r} \qquad (r = FS1, FS2) \qquad (1)$$

where $A1_{i,j,r}$ indicates that an outcome for WBO i in village j is not reported in follow-up survey round r, $High_j$ is a dummy variable indicating that the agent in village j was randomly assigned to receive high incentives, $Train_{i,j}$ is a dummy variable indicating that WBO i in village j was randomly assigned to receive business training, $\mathbf{x}_{i,j}$ is a row vector of baseline covariates,[9] $\beta_{1,0}$, $\beta_{1,1}$, $\beta_{1,2}$ and the vector δ_1 are fixed coefficients, and $\varepsilon_{1i,j,r}$ is a random disturbance clustered at the village level. Equation (1) was estimated both by OLS and by fixed-effects estimation at the village level.[10] The results in Table 5 (columns 1 and 4) indicate that overall sample attrition is not significantly related to either one of the randomized treatments (or to both jointly). To check whether the observed and unobserved characteristics of attriters differ *across* the treatment arms (Dumville and others 2006, Bell and others 2013, Winston and others 2016, Millan and Macours 2017) the following equation was estimated for each treatment arm separately for the three attrition measures:

$$A_{i,j,r} = \beta_{2,0} + \mathbf{x}_{i,j}\boldsymbol{\delta}_2 + \varepsilon_{2,i,j,r} \qquad (r = FS1, FS2) \qquad (2)$$

where the variables are defined as in equation (1). The results in Tables A-1 to A-3 in the Online Appendix indicate that the joint hypothesis that the coefficients of the covariates (i.e., the predictors of attrition) are the same across the four treatment arms (i.e., control, high agent incentives only, business training only, and both treatments combined) is not rejected at conventional significance levels with either estimation method. However, several of the estimated coefficients of individual covariates vary significantly across treatment arms. Accordingly, for the two outcomes of business profits and business capital, with significantly higher attrition, the models are also estimated using either inverse probability weights or Heckman's sample selection model (Wooldridge 2010).[11] The estimates are presented in section D of the Online Appendix and their implications are discussed in the results section.

An additional sample selection problem arises from the non-random sampling of the villages included in the first follow-up survey (FS1). As a robustness check, the models are re-estimated with the FS1 data omitted. The results are presented in section E of the Online Appendix and are quite similar to the results reported in Tables 6–8 that are obtained with the full sample.

8 These overall non-reporting rates are based on both follow-up survey rounds, as reported in Table 5.

9 The covariates ($\mathbf{x}_{i,j}$) include all of the variables analyzed in Table B-1 in the Online Appendix.

10 Village-level fixed-effects estimates are obtained, as discussed in section 5.2, because the FS1 sample is a non-random sample of the villages.

11 The selection equation in Heckman's model is identified by the randomized treatments, which are significantly related to attrition in this case, as indicated by the results in Table 5.

4. Model specification

4.1 Time elapsed after treatment exposure

In this paper, we are mainly interested in how the effects of both the supply-side and demand-side treatments may vary with the duration of time since WBOs were first exposed to the randomly assigned treatments. In the case of the supply-side treatment, the time since WBOs were first exposed to the agent incentives treatment (DA) is assumed to begin in the month of the agents' initial deployment (MA) and to have continued up to the month in which the WBO was re-interviewed in a follow-up survey (M):

$$DA_{i,r} = M_{i,r} - MA_j \qquad (r = FS1, FS2) \qquad (3)$$

where i refers to the WBO, j refers to village, and r refers to the survey round. DA is assumed to be zero for all WBOs in both phases of the baseline survey (r = BSa, BSb), and DA is also assumed to be zero for WBOs in untreated (low incentive) villages in both follow-up survey rounds (r = FS1, FS2).[12]

In the case of the demand-side treatment, the time since the WBOs' first exposure to the business training treatment (DW) is assumed to have begun in the month in which the business training began (MW_i) and to have continued up to the month in which the WBO was re-interviewed in a follow-up survey (M):

$$DW_{i,r} = M_{i,r} - MW_i \qquad (r = FS1, FS2) \qquad (4)$$

where i, j and r are defined as above. DW is assumed to be zero for all WBOs in both phases of the baseline survey (r = BSa, BSb), and DW is also assumed to be zero for WBOs not randomly assigned to receive the business training in both follow-up survey rounds (r = FS1, FS2).

WBOs' initial training month (MW_i) varied in 24 of the 401 sample villages. Because this variation may not be random (King and Behrman 2009), the training month of all WBOs in a given village is assumed to be the earliest initial training month reported for a WBO in the same village ($MW_i = \min(MW_i)_j$, j = 1, ..., 401).

Some WBOs were exposed to both the supply and demand-side treatments. The time since first exposure to both treatments ($DAW_{i,r}$) is assumed to have begun in the month in which the WBO was first exposed to both treatments (MAW_i) and to have continued up to the month in which the WBO was re-interviewed in a follow-up survey (M):

$$DAW_{i,r} = M_{i,r} - MAW_i \qquad (r = FS1, FS2) \qquad (5)$$

12 $DA_{i,r}$ is also assumed to be zero in all survey rounds in villages in which the agent resigned or was discharged prior to receiving initial training and was not replaced (10 villages). However, if the resigned agent received initial training (11 villages), $DA_{i,r}$ is assumed to equal the difference between the month of resignation and the month of initial training. In cases where an initial agent resigned and was subsequently replaced (18 villages), MA_j is assumed to be the month in which the replacement agents received their initial training.

where MAW_i is equal to max(MA_j, MW_i).

4.2 Impact measurement models

Impact estimates are obtained using three alternative models. Model 1 estimates the impacts in successive six-month intervals since the time of first exposure to the treatments (i.e., at 1–6 months, 7–12 months, 13–18 months, and 19–24 months). The estimated impacts in Model 1 are unrestricted (free to vary) from one six-month interval to another. However, the precision of the estimates is sensitive to the number of observations in each six-month interval. In contrast, Model 2 restricts the estimates to a linear trend in the time since first exposure, although the estimated trend may be positive, negative or zero. The estimates obtained with Model 2 are more precise than those obtained with Model 1 because only one parameter is estimated for each treatment. Model 3 estimates the impacts as a quadratic function of the time since first exposure to the treatments, still restricted but with considerably more flexibility than with Model 2. However, the increased flexibility of Model 3 over Model 2 comes at the cost of decreased precision because two parameters, instead of only one, are estimated for each treatment.

Model 1 can be written as:

$$Y_{i,j,r} = \alpha_6 + \mathbf{ia_j\beta_6} + \mathbf{iw_i\gamma_6} + \mathbf{iaw_{i,j,}\delta_6} + \mathbf{s\rho_6} + \mathbf{z\lambda} + \varepsilon_{6,i,j,r} \quad (6)$$

where Y refers to an outcome, r refers to the survey round, j refers to the village, i refers to an individual WBO, **s** is a row vector of dummy variables indicating different survey rounds (r = 1b, 2, 3),[13] **z** is a row vector of covariates including the baseline value of the outcome but whose other components vary across outcomes, where α_6, $\boldsymbol{\rho_6}$ and $\boldsymbol{\lambda}$ are fixed parameters, $\varepsilon_{6,i,j,r}$ is a random disturbance term equal to $c_{6,j} + c_{6,i} + u_{6,i,j,r}$ (where $c_{6,j}$ is an unobserved village-level fixed effect that is assumed to be constant over time, $c_{6,i}$ is an unobserved WBO-level effect that may or may not be constant over time, and $u_{6,i,j,r}$ is an idiosyncratic error) and where

$$\mathbf{ia_j\beta_6} = \beta_{6,1}DA1_j + \beta_{6,2}DA2_j + \beta_{6,3}DA3_j + \beta_{6,4}DA4_j \quad (6a)$$

$$\mathbf{iw_i\gamma_6} = \gamma_{6,1}DW1_i + \gamma_{6,2}DW2_i + \gamma_{6,3}DW3_i + \gamma_{6,4}DW4_i \quad (6b)$$

$$\mathbf{iaw_{i,j,}\delta_6} = \delta_{6,1}DAW1_{i,j} + \delta_{6,2}DAW2_{i,j} + \delta_{6,3}DAW3_{i,j} + \delta_{6,4}DAW4_{i,} \quad (6c)$$

The DA1–DA4, DW1–DW4 and DAW1–DAW4 are dummy variables indicating different six-month post-exposure intervals respectively to the high agent incentives, business training and both treatments, and $\beta_{6,1}$–$\beta_{6,4}$, $\gamma_{6,1}$–$\gamma_{6,4}$ and $\delta_{6,1}$–$\delta_{6,4}$ are fixed parameters.

13 Because the survey round dummies are included in the models to adjust for time (as well as variations in the definition of outcomes across survey rounds in some cases, as discussed in Annex 1). A dummy is also specified for the second phase of the baseline survey, which was separated from the first phase by several months (the first phase of the baseline survey is the omitted category).

The covariates in **z** are WBO-level baseline characteristics, including the baseline value of Y. In addition to restricting the covariates to baseline values, additional criteria for including a given covariate is that it is reported in the entire baseline sample and is statistically significant at the 0.05 level when included as an explanatory variable in a linear regression model estimated with baseline data and with a given outcome Y as the dependent variable. Consequently, the covariates included in **z** vary across outcomes.[14] The covariates are included to increase the precision of the estimates and to control at least partially for possible attrition bias.

The dummy variables in the row vector **s** are included to reflect possible intercept shifts in equation (6) over time (i.e., between survey rounds) as well as changes in the definitions of some outcome measures between survey rounds, as discussed in Annex 1.

$\varepsilon_{6,i,j,r}$ is assumed to be correlated with the dummy variables indicating alternative six-month post-treatment exposures to the treatment variables (i.e., A1–A4, W1–W4 and AW1–AW4), as discussed below. In addition, due to the presence of the fixed effect $c_{6,j}$, $\varepsilon_{6i,j,r}$ is assumed to be serially correlated in the same village j, but is assumed to be independent across villages with a variance that may be non-constant (heteroscedastic).

The following joint hypotheses are tested with Model 1: (1) agent incentives have no effect ($\beta_{6,1} = \beta_{6,2} = \beta_{6,3} = \beta_{6,4} = 0$); (2) the business training has no effect ($\gamma_{6,1} = \gamma_{6,2} = \gamma_{6,3} = \gamma_{6,4} = 0$); (3) there is no *interaction effect* between the two treatments ($\delta_{6,4} = \delta_{6,4} = \delta_{6,4} = \delta_{6,4} = 0$); and (4) there is no *combined effect* of exposure to both treatments ($\beta_{6,1} + \beta_{6,2} + \beta_{6,3} + \beta_{6,4} + \gamma_{6,1} + \gamma_{6,2} + \gamma_{6,3} + \gamma_{6,4} + \delta_{6,1} + \delta_{6,2} + \delta_{6,3} + \delta_{6,4} = 0$).

Model 2 (linear trends in the time since first exposure to the treatments) is written as:

$$Y_{i,j,r} = \alpha_7 + \beta_7 DA_j + \gamma_7 DW_i + \delta_7 DAW_{i,j} + \mathbf{s\rho_7} + \mathbf{z\lambda} + \varepsilon_{7,i,j,r} \quad (7)$$

where DA, DW DAW are the time (in months) since first exposure respectively to the high agent incentives, business training and both treatments, α_7, β_7, γ_7, and δ_7 are fixed parameters, and where the other variables and parameters are defined as in equation (6).

The following joint hypotheses are tested with Model 2: (1) agent incentives have no effect ($\beta_7 = 0$); (2) the business training has no effect ($\gamma_7 = 0$); (3) there is no *interaction effect* between the two treatments ($\delta_7 = 0$); and (4) there is no *combined effect* of the treatment arm for WBOs exposed to both treatments ($\beta_7 + \gamma_7 + \delta_7 = 0$).

Model 3 (quadratic trends in the time since first exposure to the treatments) is written as:

$$Y_{i,j,r} = \alpha_8 + \beta_{8,1} DA_j + \beta_{8,2} DA^2_j + \gamma_{8,1} DW_i + \gamma_{8,2} DW^2_{i,j} + \delta_{8,1} DAW_{i,j} + \delta_{8,2} DAW^2_{i,j} + \mathbf{s\rho_8} + \mathbf{z\lambda} + \varepsilon_{8,i,j,r} \quad (8)$$

14 Section B (Table B-1) in the Online Appendix reports the regressions used to select the covariates to include in the models for each outcome.

The following joint hypotheses are tested with Model 3: (1) agent incentives have no effect ($\beta_{8,1} = \beta_{8,2} = 0$); (2) the business training has no effect ($\gamma_{8,1} = \gamma_{8,2} = 0$); (3) there is no *interaction effect* between the two treatments ($\delta_{8,1} = \delta_{8,2} = 0$); and (4) there is no *combined effect* of the treatment arm for WBOs exposed to both treatments ($\beta_{8,1} + \beta_{8,2} + \gamma_{8,1} + \gamma_{8,2} + \delta_{8,1} + \delta_{8,2} = 0$).

Joint tests can also be used to identify which model is most appropriate for estimating impact for a given outcome. For example, in Model 1, joint tests for equality in the impacts of each treatment across six-month intervals ($\beta_{6,1} = \beta_{6,2} = \beta_{6,3} = \beta_{6,4}$), ($\gamma_{6,1} = \gamma_{6,2} = \gamma_{6,3} = \gamma_{6,4}$) and ($\delta_{6,1} = \delta_{6,2} = \delta_{6,3} = \delta_{6,4}$) provide a basis for comparing Model 1 to the typical impact estimation model in which impact is assumed not to vary with the time elapsed since treatment exposure. In Model 2, which assumes a linear trend in the time pattern of effects, a joint test of the hypothesis that impacts do not vary with the time elapsed since exposure can provide a basis for choosing between Model 1 and Model 2. Lastly, a joint test of the significance of the quadratic terms in Model 3 provides a basis for choosing between Model 2 and Model 3.

5. Estimation

5.1 Transformations of outcomes prior to estimation

Average monthly profits in the WBOs' primary business and the total value of their business capital are both highly skewed (skewness = 17.48 and 8.97 respectively) with distributions that are approximately lognormal with only 3 and 13 zero values respectively. Both outcomes are transformed to inverse hyperbolic sine (IHS) values after being re-scaled (multiplied by 1,000) so that their distributions more closely approximate lognormal distributions, making the estimates easier to interpret (Bellemarre and Wichman 2019). In this case, the estimated average treatment effect on both dependent variables can be converted to a percentage change relative to the baseline sample mean in Model 1 by using the Kennedy (1981) approximation (equation A2-1 in Annex 2) or to the projected percentage changes in Models 2 and 3 using the formula in equation (A2-2) in Annex 2.

Labor inputs (hours worked in a typical month by WBOs in their primary business) is a discrete count variable bounded between zero and 720 with an approximately normal distribution, apart from the truncated left tail (skewness = 1.32). Business practices (the sample proportion of 13 recommended businesses practices followed) is a scaled discrete count variable between 0 and 1, while WBO agency (the sample proportion of five household decisions in which the WBO is reported to participate) is also a scaled discrete count variable between 0 and 1.[15] These outcomes are analyzed as linear regression models.[16]

15 The fact that these two outcomes are scaled count variables (i.e., the counts are divided by 13 and 5 respectively) affects only the scaling of the estimated coefficients, not their significance levels.

16 Section C of the Online Appendix compares the linear regression model estimates to those obtained from a Poisson regression model.

The two savings measures are based on two separate questions: (1) did you save at all during the last 12 months (or last 3 months, in the first follow-up survey)? and, if yes, (2) how much did you save in various savings instruments? The responses are analyzed as two separate outcomes, the first as a zero-one (binary) dependent variable in a linear probability model ("any savings")[17] and the second as the non-zero amount saved ("non-zero savings"), which is highly skewed (skewness = 44.71) and approximately lognormally distributed and is modeled as a natural log transformed dependent variable in a linear regression model.[18] The estimated average treatment effect on the logged non-zero value of the amount saved can be converted to a percentage change relative to the baseline sample mean in Model 1 by using Kennedy's approximation (equation A2-1 in Annex 2) or to the projected percentage changes in Models 2 and 3 using the formula in equation (A2-2) in Annex 2.

Household income is a composite variable (asset index) equal to the first principal component of a set of standardized indicators referring to durable asset ownership.[19] Since the component indicators are standardized, the distribution of the index is centered on zero, with values ranging from −5.323 to 10.571 and a distribution that is similar to a normal distribution (skewness = 0.64), apart from an extended positive tail. Accordingly, the household income variable is not transformed.

5.2 Model estimation

The times since WBOs first exposure to the treatments (DA, DW and DAW), unlike the treatments themselves, were not randomly assigned. In particular, the times since first exposure are shorter in villages in which the agent or business training was conducted later (as proxied for by the baseline survey month) due to delays in recruiting agents (Table 2).

In addition, the first follow-up survey (FS1) sample is limited to 200 of the 401 sample villages in which the training of agents and WBOs had been completed. This is clearly a non-random sample of the 401 sample villages, as evidenced by differences in the mean times since first exposure, which are substantially higher in the second follow-up survey in villages that were included in the FS1 (Table 3).

Fortunately, the time since first exposure for the WBOs is endogenous and/or selective only at the village level because the random assignment of WBOs to the training intervention was done *within* villages, with the time since first exposure based on the date on which the first-trained WBO in each village began her training.

The random assignment of the business training within each village panel removes any correlation between the unobserved WBO-level fixed effect (c_i) and the random disturbance term (ε).

17 Estimating "any savings" as a probit (or logit model) instead of a linear regression model does not significantly affect the estimated average treatment effects, as shown in Section C of the Online Appendix.

18 Responses of "yes" to the first question with responses of zero to the second question (15 responses in all survey rounds) are assumed to indicate that there was zero saving during the reference period.

19 The properties of the asset index as a measure of household income are discussed in Section F of the Online Appendix.

Accordingly, fixed-effects (FE) estimation based solely on the variation in the times since first exposure *within* village panels yields consistent estimates of the duration of post-treatment exposure effects and was used to estimate the models in equations (6)—(8). Consistent estimates of the standard errors are obtained by adjusting them for possible heteroscedasticity, serial correlation and clustering within each village panel using the Huber-White sandwich estimator (Wooldridge 2010).

Estimates of the post-exposure treatment effects of high versus low agent incentives are in this case based on the observed variation *within* each village panel in the outcomes between the baseline values (when there were no high agent incentives) and the values of the same outcomes in the follow-up surveys in which some villages had high agent incentives while others had low agent incentives. Differences in outcomes *between* villages are ignored. Similarly, estimates of the post-exposure effects of the business training are based on the observed variation within each village panel in the outcomes between the baseline values (when the training had not yet occurred) and the values of the same outcomes in the follow-up surveys in which some WBOs in each village had received the business training while other WBOs in the same village had not. Variation in outcomes between WBOs in different village panels are ignored. Since there is no variation in the village-level unobserved fixed effect (c_j) within each village panel, the treatment measures are uncorrelated with the random disturbance term (ε) in all three models, and the estimates are consistent.

6. Results

Village-level fixed-effects estimates of Models 1–3 (equations 6–8) are reported in Tables 6–8. The tables report the estimates of the average treatment effects, estimated standard errors and significance levels (using the algebraic symbols used in equations 6–8 to represent the treatments) as well as the p-values for the joint hypotheses tested for each outcome, the baseline sample means, standard deviations, root mean square errors (RMSE) of the outcomes in their post-transformation metrics (as described in section 5.1), and the R-squared and estimation sample size (N).[20] The estimated coefficients of the DA and DW treatments in Tables 6–8 are the estimated effects of each treatment provided alone, whereas the estimated coefficient of DAW refers to the marginal estimated effects of receiving both treatments (i.e., the interaction term). Estimates of the effect of receiving both treatments combined (one of the four treatment arms) can be calculated as the sum of the estimated coefficients of the treatments in Tables 6–8 *(e.g.,* $\beta_7 + \gamma_7 + \delta_7$ in Model 2).[21]

Figures 1–8 show graphically the statistically significant predicted impacts of the treatments on each outcome in the common metric of standard deviations from the baseline means together with

20 The estimated coefficients of the dummy variables referring to different survey rounds and the included covariates for each outcome are not reported in Tables 6–8 but are available upon request.

21 The other three treatment arms are control (represented by the intercept), high agent incentives alone, and the business training alone.

95% confidence intervals for four successive six-month intervals of time since first exposure to the treatments.[22] Figures 1–8 provide a clearer indication of the impacts of the treatments over the four six-month post-treatment intervals than the estimated coefficients in Tables 6–8. Whereas Tables 7 and 8 show only the estimated coefficients of the treatments, Figures 1–8 show the mean *predicted* impacts of the treatments in each six-month post-treatment time interval. Estimates with 95% confidence intervals (CIs) that do not include a value of zero are referred to below as "significant" (at the 0.05 level), while estimates that are significant at only the 0.10 level are referred to as "marginally significant."[23]

Most of the predicted impacts (and all of the significant predicted impacts) in Figures 1–8 are consistent with Model 2 in adhering to a significant positive linear trend in the time since first exposure to the business training (or to both treatments combined, in the case of business practices). However, the time patterns of the predicted impacts vary with the other models. The variation is greatest with Model 1, which places no constraints on the time patterns of the predicted impacts. However, the increased flexibility of Model 1 comes at a cost of lower precision (wider CIs) compared to Model 2.[24] Model 3 restricts the predicted impacts to a more flexible quadratic trend in the time since treatment, but with less precision in most cases than Model 2. Additional analysis using the Benjamin-Hochberg approach (Section G of the Online Appendix) underscores the limited reliability of the Model 1 estimates which use the same sample to estimate many more parameters than those estimated in Models 2 and 3.

The estimates in Tables 6–8 are reduced form estimates of the impacts of the treatments on a given outcome, without showing the paths through which they occur (they tell us what happened as the result of the treatments but not how it happened). Tables 9 and 10 present standardized estimates of the parameters of a structural equations model (SEM) used in mediation analysis to reveal how the estimated impacts interacted causally to produce the results in Figures 1–8.[25] The outcomes in the SEM are the same as those used in Model 2, but with the two personal savings variables combined into a single savings variable.[26] The estimation sample is limited to the baseline (BS) and endline (FS2) data (the FS1 data was not used since it does not include data on business practices).[27] The *direct effects*

22 The predicted effects are obtained from the estimated regressions reported in Tables 6–8 using Stata's "adjust" command. Treatment variables and covariates included in the estimation command but not included in the "adjust" variable list (which identifies the treatment variables for which predicted values are calculated) are left at their current values, observation by observation.

23 Because the 95% confidence intervals shown in Figures 1–8 are symmetric, an estimated effect with a confidence interval that does not include zero has a probability less than 0.025 of being equal to zero.

24 The lower precision of the Model 1 estimates is due to the estimation sample being used to estimate four parameters, compared to only one parameter in Model 2 and two in Model 3.

25 The estimated coefficients in Tables 9 and 10 are maximum likelihood estimates obtained using the "sem" command in Stata 17.0's SEM software module. They are standardized so that they are easy to compare across outcomes.

26 This is possible because the variables in the SEM are measured as deviations about their means, which provides equivalent fixed-effects (FE) estimates (since Stata 17.0's SEM software does not provide an FE estimation option). These transformed variables do not have a high concentration of zero values.

27 Section E of Online Appendix provides equivalent estimates to those in Table 7 and Figures 1–8 obtained from a sample that excludes the FS1 data.

shown in Table 9 are the estimated coefficients of the treatments, as in Tables 6–8, as well as the estimated coefficients of any outcomes that are also included in each structural equation as direct determinants of the outcome. The *total effects* (impacts) shown in Table 10 are the sum of the direct effects in Table 9 and the indirect effects of outcomes included in each structural equation that are also included as endogenous variables in other structural equations. For example, the total effects of business practices on business profits in column 1 of Table 10 reflect not only their direct effect on business profits (shown in column 1 of Table 9), but also their indirect effect from the inclusion of business practices as a direct determinant of business capital in column 2.

More detailed descriptions of the results for each outcome follow.

Profits. Figure 1 shows the predicted impacts of the business training alone (DW) and of both treatments combined (DAW) in all three models (the predicted impacts of high agent incentives alone are not shown because they are statistically insignificant in all three models). Although the predicted impacts are positive in year one, they are relatively small and not statistically significant. Beginning in the third six-month interval (13–18 months after first exposure to the treatments), the predicted impacts of DW increase sharply and become significant in Model 1, increasing more rapidly than both a linear trend (Model 2) or quadratic trend (Model 3) during the second year. The mediation analysis in Table 9 (column 1) shows that, in addition to the treatments, important direct drivers of profits include business capital, business practices and agency.

An important question is whether the absence of significant impacts on business profits during the first year reflects an actual delay in the effects or merely the time lag in the way profits were measured ("average monthly profits *during the past 12 months*"). It makes intuitive sense that increases in the profits of WBO's firms, which on average are considerably smaller and less capital intensive and profitable than male-owned firms, would take time to materialize, especially after relatively 'light' interventions, such as the ones in this trial. If the effects had begun without delay, one would expect to see a pattern in which the predicted impacts from Model 1 would increase noticeably between the first and second six-month intervals. However, such a staggered pattern is observed most clearly only in year two. The conclusion, therefore, is that most of the impact of the treatments on business profits occurred in year two.

Business capital. Figure 2 indicates that the predicted impacts of the combined treatments (DAW) on business capital based on the estimates in Tables 6–8 are not even marginally significant. However, because sample attrition was highest for capital inputs (i.e., 12.7% in both follow-up surveys combined in Table 5), additional analysis was done to see whether adjustments for sample attrition would show stronger effects.[28] Figure 2 shows that the predicted impacts based on a Heckman sample selection model are positive and marginally significant in both Models 2 (linear trend)

28 The results from using either inverse probability weights or a Heckman sample selection model to remove possible selectivity bias are reported in section D of the Online Appendix.

and 3 (quadratic trend). The mediation analysis in Tables 9 and 10 indicate that improved business practices and increases in business profits, personal savings and household income also contributed to the increases in business capital, while increased business capital contributed to business profits.

Labor inputs. Figure 3 indicates that the predicted impacts of the combined treatment (DAW) on WBO's own labor inputs (hours worked in a typical month in the primary business) are not even marginally significant in any of the models. The results of the mediation analysis in row 3 of Table 10 indicate that the WBO's own labor inputs do not contribute even indirectly to the increases in any other outcome.[29]

Business practices. The business training (possibly in combination with agent incentives) was expected to have a direct positive effect on adherence to recommended business practices.[30] Figure 4 shows that the predicted impacts of the combined treatments (DAW) on business practices follow a significant positive linear trend in Model 2, reaching +0.17 standard deviations in the fourth six-month interval.[31] The mediation analysis in column 4 of Table 10 indicates that the positive effects of the combined treatments on business practices were not significantly increased by any direct or indirect effects of other outcomes.[32] However, the results in row 4 of Table 10 indicate that improved business practices contributed to increases in business profits and business capital.

Savings. The business training was expected to have a direct positive effect on WBOs' personal savings, reflecting the emphasis given to the importance of saving in the training, while the agent incentives were also expected to have a direct positive effect on savings by making individual savings accounts more accessible to WBOs. Figure 5 indicates that the predicted impact of the business training (DW) alone on the probability of non-zero savings follows a significant positive linear trend, while Figure 6 indicates that the predicted impact of DW on the non-zero amount saved follows a marginally significant positive linear trend, with both predicted impacts reaching +0.1 standard deviations in the fourth six-month interval. The predicted impacts on savings of high agent incentives alone (DA) and of both treatments combined (DAW) are not shown in Figures 5 and 6 because they are uniformly insignificant in all three models (Tables 6–8). The mediation analysis in column 5 of Table 9 and 10 indicates that increases in both business profits and household income contributed to increased savings, while the estimates in row 5 indicates that increased savings contributed to increases in business capital.

29 Although the results in column 3 of Tables 9 and 10 of the mediation analysis indicate that the combined treatment (DAW) has a significant positive effect (consistent with the results in Tables 6 and 7) that is reinforced by the direct effect of increased business profits, the significant negative direct and indirect effects of increased household income in column 3 are apparently strong enough to offset these positive effects.

30 For example, if WBOs are interested in applying for a bank loan, they would be expected to ensure that they have the necessary records (one of the 13 recommended business practices).

31 The estimates in the literature on the positive experimental effects of business training on business practices range from 0.1 to 0.2 standard deviations (McKenzie and Woodruff 2015). The hypothesis that there is no combined effect in Model 2 is rejected at the 0.05 level in Table 7.

32 The estimate of the combined treatment (DAW) in Tables 9 and 10 is the sum of the three estimated coefficients of the treatments (i.e., DA + DW + DAW).

Agency. The agent incentives were expected to have a positive direct effect on WBOs' agency by improving their access to secure and private individual savings accounts. Although the results in Figure 7 do not show significant predicted impacts of the agent incentives (DA) on WBOs' agency, either alone or in combination with the business training, they do show positive and significant predicted impacts in Model 2 from the business training (DW), which emphasized the importance of saving.[33] Although the unconstrained predicted impacts of DW in Model 1 are stronger and highly significant during the first six months (for which only FS1 data are available), they decrease in the second six months and become insignificant in year two. The predicted impacts from Model 2, with narrower confidence intervals, show that the effects of the business training on WBOs' agency are not short-lived, but rather follow a significant positive linear trend reaching +0.15 standard-deviations in the fourth six-month interval. The results of the mediation analysis in column 6 of Tables 9 and 10 (which does not include the FS1 data) indicate that the direct effects of the business training on agency were not significantly increased by the direct or indirect effects of other outcomes. Although the mediation analysis (row 6) indicates that the treatment-induced effects on agency contributed to the impact on business profits, the absence of other significant direct or indirect effects in row 6 suggest that women's agency did not play a major role in driving the impacts reported in Figures 1–8 beyond its modest contribution to business profits.

Household income. Figure 8 indicates that the predicted impacts of the business training alone (DW) on household income follow a significant positive linear trend (Model 2) in the time since treatment.[34] In addition, Figure 8 indicates that the predicted impacts of DW increase more rapidly than a linear trend in year two, adhering more closely to a quadratic trend (Model 3). That the predicted impacts of the DW and DAW treatments in all three models are lower in the first year of exposure than in the second year is consistent with the conclusion above that increased profits occurred mainly in the second year, particularly since the mediation analysis in column 7 of Tables 9 and 10 indicates that business profits is the main driver of household income (apart from the treatments).

7. Conclusions

The results of this trial support the notion that comparatively 'light' interventions designed to improve women's economic lives—in this case a small group financial literacy and business training plus more motivated bank agents—need to be tracked long enough for businesswomen to manifest new and beneficial behaviors. The unfolding of economic empowerment outcomes takes time, and this must be especially the case for women operating very small firms in traditional rural business environments.

33 The predicted effects from Model 3 are not shown because they are uniformly insignificant for all three treatments.

34 The predicted effects of high agent incentives alone (DA) are not displayed in Figure 8 because, although they are also uniformly positive in the third and fourth six-month intervals, they are either insignificant or only marginally significant.

In this trial in predominantly rural villages in East Java, Indonesia, the expected direct effects of these "light treatments" on business practices, personal savings and agency occurred and were reinforced by important direct and indirect effects on additional economic empowerment-related outcomes, including business profits, business capital and household income.

The impacts of these 'light' treatments on most of the WBO outcomes analyzed were not fixed in time, but rather increased with the duration of time since first exposure to the treatments. Although the time path of most impacts adheres to a significant positive linear trend, there are exceptions. In the case of business profits and household income, for example, the predicted impacts are small and insignificant during the first year but increase more rapidly than a linear trend in the second year.

Returning to the three questions posed in the introduction regarding the durability of impacts over time, the best time to measure them, and possible causal links between the impacts, we begin with the question of when to measure impacts. The results for both business profits and household income indicate that if their impacts had been measured only at the end of the first year, the conclusion would have been that they were ineffective. In both cases, waiting at least two years was necessary to obtain an accurate estimate of impact.

In terms of possible causal links, the mediation analysis found evidence of strong causal links between several of the outcomes that contributed importantly to the relatively large impacts on business profits and household income. In the case of business profits, business capital, personal savings and household income, causal links between impacts existed in both directions. In the case of improved business practices, however, which had a positive impact on both business profits and business capital, the causal links *to* improved business practices are limited to the direct effects of the treatments. Although increased agency had a significant (but modest) positive impact on business profits, the causal links *to* increased agency which, unlike past research is not short-lived, are also limited to the direct effects of the treatments.

The question concerning the durability of the impacts cannot be answered by traditional constant impact models. The results reported here suggest that most of the impacts were not only durable but actually increased in magnitude with the passage of time. Whether these positive upward trends in the impacts continued beyond two years is an unanswered question that can only be resolved if a third follow-up survey is conducted.

References

Anderson-MacDonald, S., R. Chandy and B. Zia. 2018. "Pathways to Profits: The Impact of Marketing vs. Finance Skills on Business Performance." *Management Science* 64(12): 5559–5583.

Banerjee, A., D. Karlan and J. Zinman. 2015. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." *American Economic Journal: Applied Economics* 7(1): 1–21.

Bell, M., M. Kenward, D. Fairclough, and N. Horton. 2013. "Differential dropout and bias in randomized controlled trials: When it matters and when it may not." *British Medical Journal* 346: e8668.

Bellemarre, M. and C. Wichman. 2019. "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics* July: 1–12.

Buvinic, M., H. C. Johnson, E. Perova, and F. Witolear. 2020 "Can boosting savings and skills support female business owners in Indonesia? Evidence from a Randomized Controlled Trail," East Asia and Pacific Gender Innovation Lab (EAPGIL), World Bank Group (March).

Buvinic, M., J. Knowles and F. Witoelar. 2021. "Persistent Gender Gaps in Business Profits in Indonesia: Implications for Covid-19 Recovery Policies," Working Paper 603, Center for Global Development, Washington, DC (December).

Calderon, G., J. M. Cunha and G. De Giorgi. 2013. "Business Literacy and development: Evidence from a randomized controlled trial in rural Mexico." Working Paper No. 19740, National Bureau of Research.

Dalton, P.S., J. Rüschenpöhler, B. Uras and B. Zia. 2021. "Curating Local Knowledge: Experimental Evidence from Small Retailers in Indonesia." *Journal of the European Economic Association*, 19(5): 2622–2657.

De Mel, S., D. McKenzie and C. Woodruff. 2014. "Business training and female enterprise startup, growth, and dynamics: Experimental evidence from Sri Lanka." *Journal of Development Economics* 106(C): 199–2010.

Deserrano, E., G. Leon-Ciliotta and F. Witoelar. 2021. "When Transparency Fails: Financial Incentives for Local Banking Agents in Indonesia," Unpublished manuscript (January).

Donald, A., G. Koolwal, J. Annan, K. Falb and M. Goldstein. 2020. "Measuring women's agency." *Feminist Economics* 26(3): 200–226.

Dumville, J. C., D. J. Torgerson and C. E. Hewitt. 2006. "Reporting attrition in randomized controlled trials," *British Medical Journal* 332(7547): 969–71.

Dupas, P. and Robinson, J. 2013. "Why Don't the Poor Save More?: Evidence from Health Saving Experiments," *American Economic Review* 103(4): 1138–1171.

Garz, S., R. Heath, E. Kipchumba and M. Sulaiman. 2020. "Evidence of Digital Financial Services Impacting Women's Economic Empowerment: What Explains the Impacts and what is Left to Learn," WEE DiFine, Brac Institute of Governance and Development, Brac University, Bangladesh.

Goldberger, A. S. 1968. "The Interpretation and Estimation of Cobb-Douglas Functions," *Econometrica* 36(3): 464–72.

Hoff, K. and J. E. Stiglitz. 2016. "Striving for Balance in Economics: Towards a theory of the social Determination of Behavior." *Journal of Economic Behavior and Organization* 126: 25–57.

Huis, M., R. Lensink, N. Vu and N. Hansen. 2019. "Impacts of the Gender and Entrepreneurship Together Ahead (GET Ahead) training on empowerment of female microfinance borrowers in Northern Vietnam," *World Development* 120: 46–61.

Kabeer, N. 1999. "Resources, Agency, Achievements: Reflections on the Measurement of Women's Empowerment," *Development and Change* 30: 435–464.

Kennedy, P. E. 1981. "Estimation with correctly interpreted dummy variables in semilogarithmic equations," *American Economic Review* 71(4): 801.

King, E. M. and J. R. Behrman. 2009. "Timing and Duration of Exposure in Evaluations of Social Programs." *World Bank Research Observer* 24(1): 55–82.

Laszlo, S., K. Grantham, E. Oskay and T. Zhang. 2017. "Grappling with the challenges of measuring women's economic empowerment," ISID (Institute for the Study of International Development), McGill University. http://grow.research.mcgill.ca/pulbications/workingpapers/gwp-2017-12.pdf

Lin, W., D. P. Green, and A. Coppock. 2016. "Standard operating procedures for Don Green's lab at Columbia." Version 1.05 (June 7, 2016). https://alexandercoppock.com/Green-Lab-SOP/Green_Lab_SOP.html

McKenzie, D. 2017. "How Effective are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence." *World Bank Research Observer* 32(2): 127–154.

McKenzie, D. 2021. "Small Business Training to Improve Management Practices in Developing Countries: Re-assessing the Evidence for 'Training Doesn't Work." *Oxford Review of Economic Policy*, Volume 37, Issue 2, Summer 2021, Pages 276–301. https://doi.org/10.1093/oxrep/grab002

McKenzie, D. and S. Puerto. 2017. "Growing Markets through Business Training for Female Entrepreneurs: A Market-Level Randomized Experiment in Kenya," Policy Research Working Paper 7993, The World Bank, Washington, D.C.

McKenzie, D. and C. Woodruff. 2015. "Business Practices in Small firms in Developing Countries." NBER Working Paper No. 21505. National Bureau of Economic Research, Cambridge, MA.

Millan, T. and K. Macours. 2017. "Attrition in Randomized Control Trials: Using Tracking Information to Correct Bias." IZA Discussion Paper No. 10711, Institute of Labor Economics, Bonn (April).

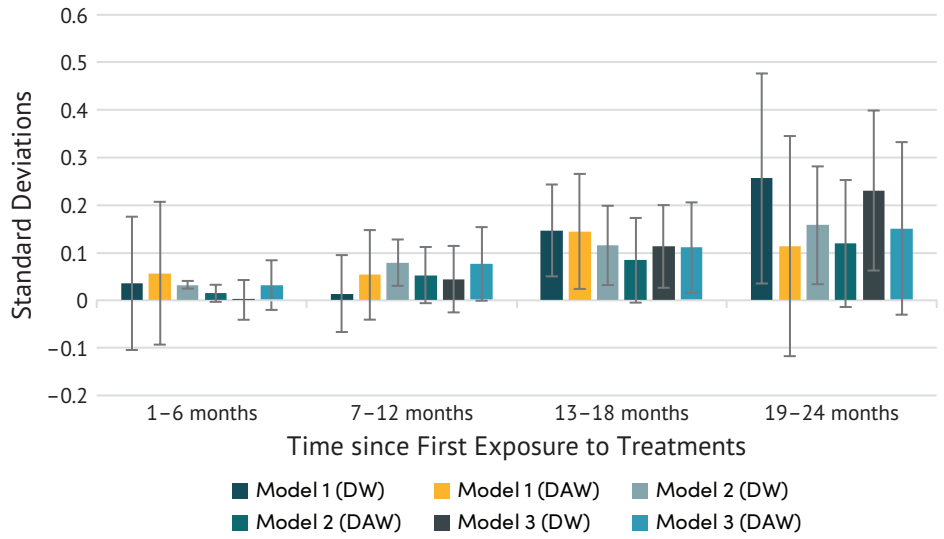
Sen, A. K. 1999. *Development as Freedom*. New York: Oxford University Press.

Valdivia, M. 2015. "Business training plus for female entrepreneurship? Short and medium-term experimental evidence from Peru." *Journal of Development Economics* 113(C): 33–51.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis for Cross-section and Panel Data*. Second edition. Cambridge, MA: MIT Press.

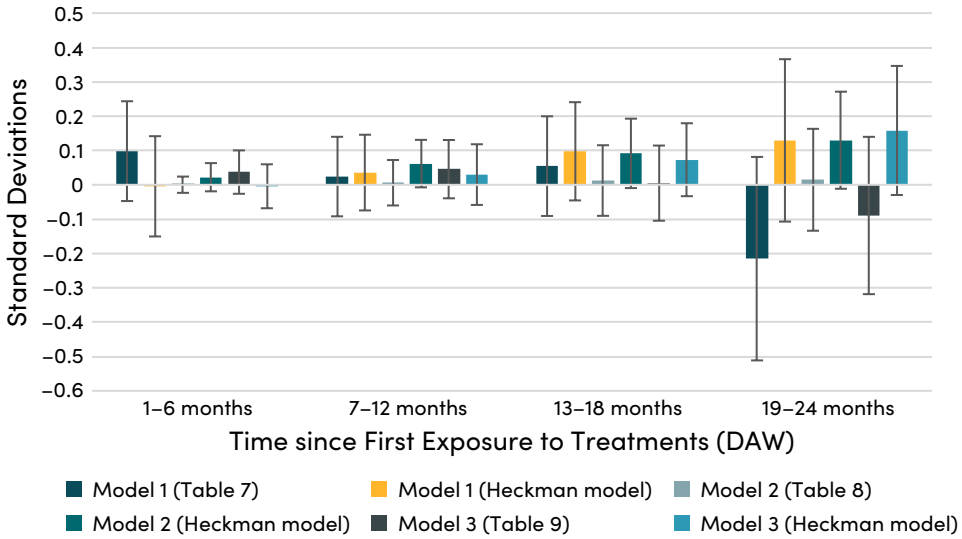
World Bank. 2016. Women Entrepreneurs in Indonesia: A Pathway to Increasing Shared Prosperity. Finance and Markets Global Practice, East Asia Pacific Region, Jakarta, Indonesia.

FIGURE 1. Predicted effects of the treatments on business profits over four six-month intervals of post-treatment exposure (with 95% confidence intervals adjusted for clustered sampling)



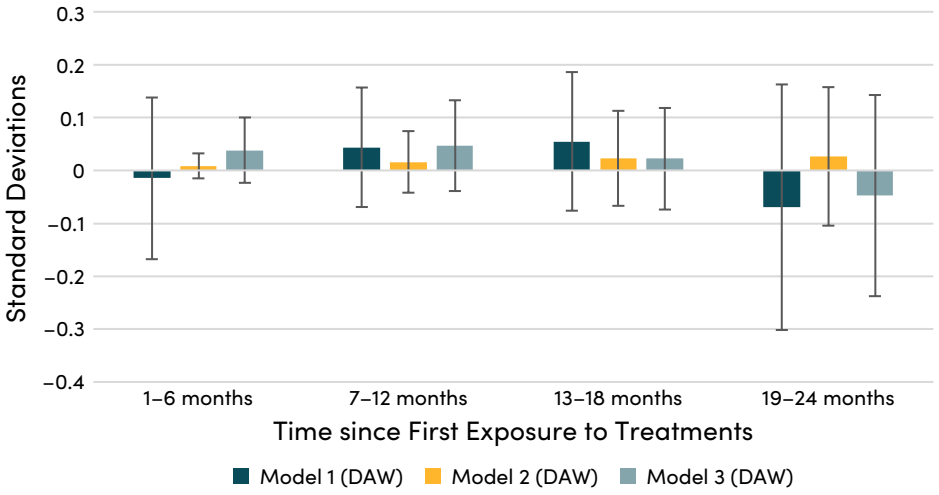
Source: Tables 7–9 (see text for prediction methods)

FIGURE 2. Predicted effects of the treatments on capital inputs over four six-month intervals of post-treatment exposure (with 95% confidence intervals adjusted for clustered sampling), showing alternative Heckman model estimates to adjust for sample attrition



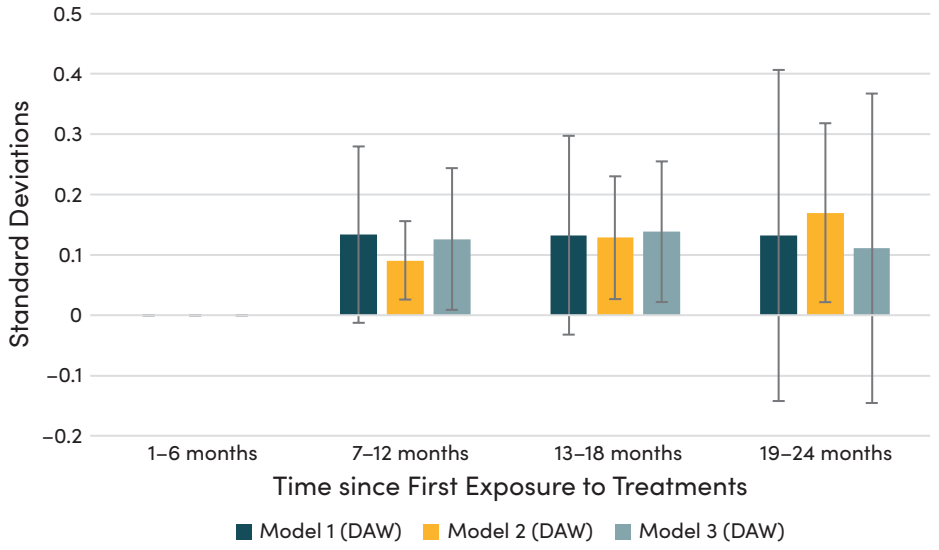
Source: Table 6–8 and Online Appendix Table D-2 (see text for prediction methods)

FIGURE 3. Predicted effects of the treatments on labor inputs over four six-month intervals of post-treatment exposure (with 95% confidence intervals adjusted for clustered sampling)



Source: Tables 6–8 (see text for prediction methods)

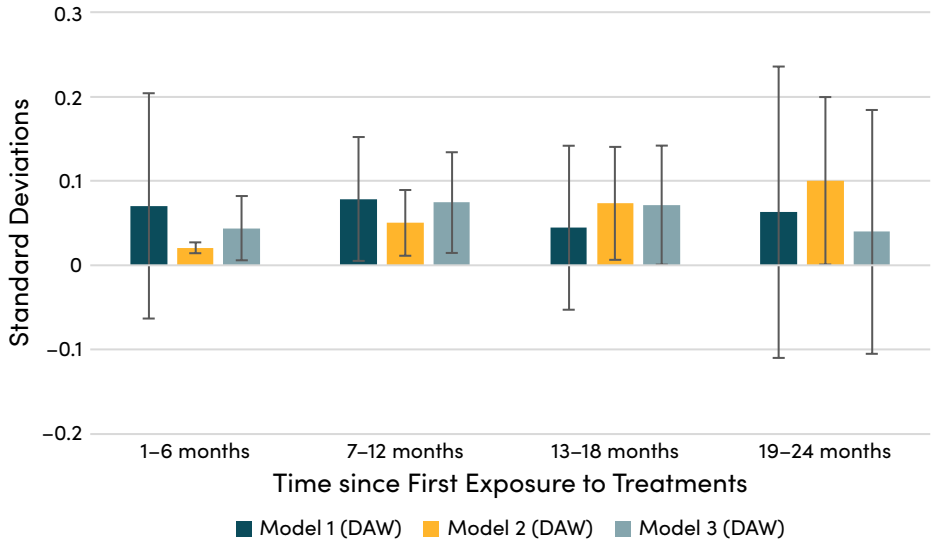
FIGURE 4. Predicted effects of the treatments on adherence to recommended business practices over three* six-month intervals of post-treatment exposure (with 95% confidence intervals adjusted for clustered sampling)



Source: Tables 6–8 (see text for prediction methods)

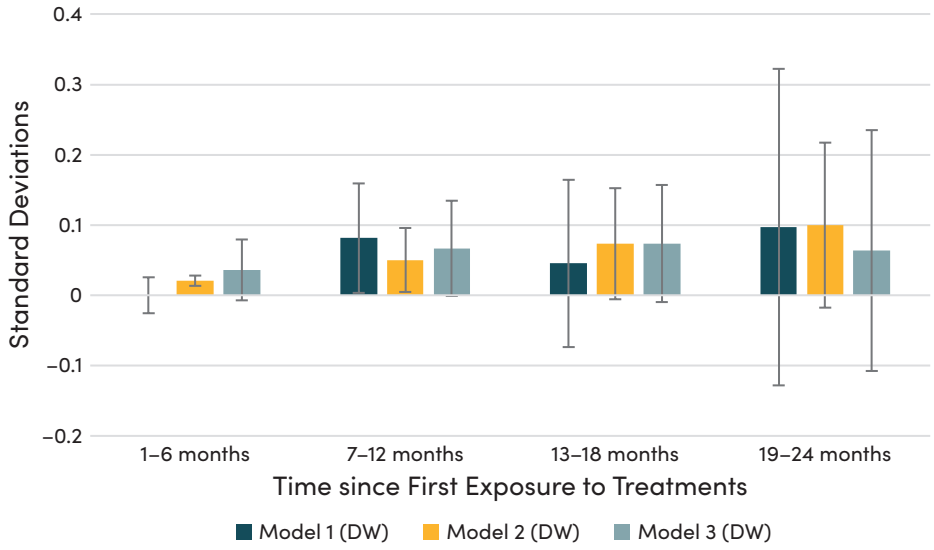
*No estimates are available for the first six-month interval because no data on business practices were collected in the first follow-up survey.

FIGURE 5. Predicted effects of the treatments on any non-zero reported savings over four six-month intervals of post-treatment exposure (with 95% confidence intervals adjusted for clustered sampling)



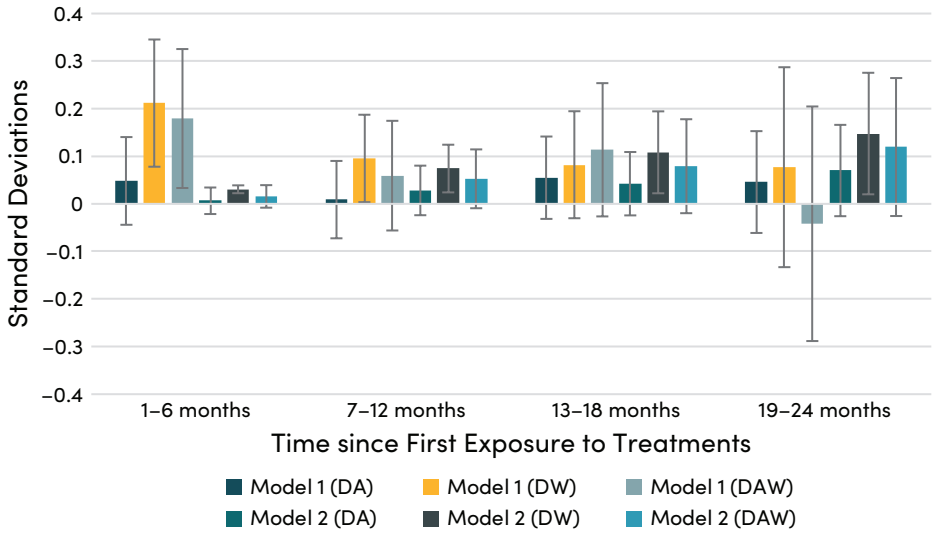
Source: Tables 6–8 (see text for prediction methods)

FIGURE 6. Predicted effects of the treatments on the amount of non-zero savings over four six-month intervals of post-treatment exposure (with 95% confidence intervals adjusted for clustered sampling)



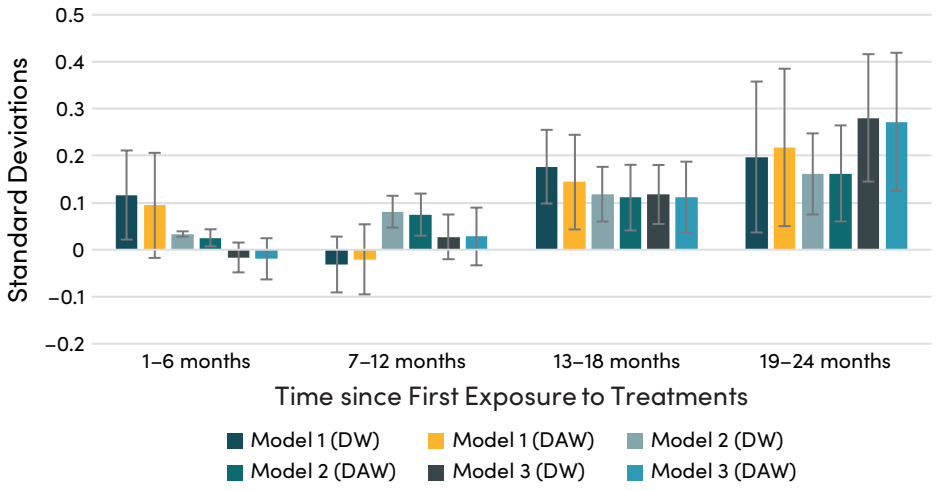
Source: Tables 6–8 (see text for prediction methods)

FIGURE 7. Predicted effects of the treatments on WBO agency over four six-month intervals of post-treatment exposure (with 95% confidence intervals adjusted for clustered sampling)



Source: Tables 6–8 (see text for prediction methods)

FIGURE 8. Predicted effects of the treatments on household welfare over four six-month intervals of post-treatment exposure (with 95% confidence intervals adjusted for clustered sampling)



Source: Tables 6–8 (see text for prediction methods)

TABLE 1. Observations by time elapsed since assigned treatment and follow-up survey rounds (see notes)

Treatment→	Agent Incentives			WBO Training & Mentoring			Combined Treatments		
Follow-up Survey Round→	First	Second	Total	First	Second	Total	First	Second	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Assigned to treatment	651	1,363	2,014	800	1,600	2,400	372	764	1,136
1–6 months	385	51	436	508	0	508	248	14	262
7–12 months	140	466	606	292	772	1,064	124	371	495
13–18 months	126	580	706	0	626	626	0	286	286
19–24 months	0	266	266	0	202	202	0	93	93
Not assigned to treatment	748	1,478	2,226	599	1,241	1,840	1,027	2,077	3,104
Total sample	1,399	2,841	4,240	1,399	2,841	4,240	1,399	2,841	4,240

Notes: The numbers reported in columns 1–6 refer to the numbers of sample WBOs who received the indicated treatment. To calculate the numbers of sample WBOs who received *only* the treatment, it is necessary to subtract the corresponding numbers who received both treatments combined in columns 7–9. 11 WBOs included in the baseline survey were not included in the second follow-up survey sample, including 4 who were randomly assigned to the WBO training and mentoring treatment.

TABLE 2. Conditional mean of time elapsed (in months) by treatment and by month in which the baseline survey interview was conducted (see note)

Baseline Survey Month	Agent Incentives		WBO Training & Mentoring		Combined Treatments	
	Mean	N	Mean	N	Mean	N
	(1)	(2)	(3)	(4)	(5)	(6)
Nov-16	18.44	210	15.88	238	15.93	120
Dec-16	14.71	304	13.48	448	12.94	176
Jan-17	15.13	32	12.73	22	12.56	16
Feb-17	13.00	49	11.33	60	9.86	28
Jul-17	10.89	367	9.07	395	9.33	209
Aug-17	10.25	615	9.26	713	8.86	351
Sep-17	11.40	190	9.56	188	9.64	108
Oct-17	10.19	170	9.29	204	8.52	88
Nov-17	10.00	114	9.62	136	9.03	64
Total	12.09	2,051	10.80	2,404	10.43	1,160

Notes: Mean time elapsed since exposure in this table are conditional on sample WBOs receiving each treatment. Mean time elapsed depend on both the month in which the relevant training began (i.e., agent training in column (1), WBO training in column (3), both in column (5) and on the month in which the WBO was interviewed in the two follow-up surveys.

TABLE 3. Mean time elapsed (months) since exposure to treatments in the second follow-up survey according to whether the village was included in the first follow-up survey sample

	Agent Incentives	WBO Training & Mentoring	Combined Treatments	N
Included in first follow-up survey sample:				
Yes	8.03	8.78	4.08	1,398
No	5.69	6.09	2.88	1,443
Overall mean (second follow-up survey)	6.84	7.42	3.47	2,841

TABLE 4. Characteristics of the eight outcomes analyzed by survey round

	Mean	Standard Deviation	Minimum Value	Maximum Value	Median Value	No. of zero values	No. of Reported Values	No. of Non-Responses	Total Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Baseline survey (both phases)									
Profits	1.651	3.065	0.0000	101.1	0.963	1	2,835	17	2,852
Capital inputs	21.016	71.502	0.0268	1622.4	4.574	0	2,685	167	2,852
Labor inputs	244.855	104.188	1.0000	720.0	240.000	0	2,849	3	2,852
Business practices	0.228	0.186	0.0000	1.0	0.154	124	2,846	6	2,852
Any savings	0.837	0.369	0.0000	1.0	1.000	464	2,850	2	2,852
Non-zero level of savings	7.096	13.759	0.0019	246.1	2.929	10	2,376	0	2,386
HH decision-making	0.794	0.293	0.0000	1.0	1.000	147	2,846	6	2,852
HH welfare	0.040	1.818	−5.322	7.042	0.032	0	2,847	5	2,852
First follow-up survey									
Profits	1.802	2.717	0.0000	36.6	0.962	1	1,312	87	1,399
Capital inputs	24.793	84.092	0.0000	2035.8	5.957	10	1,324	75	1,399
Labor inputs	225.850	105.244	0.0000	630.0	230.000	1	1,325	74	1,399
Business practices	NA	NA	NA	NA	NA	NA	0	0	0
Any savings	0.845	0.362	0.0000	1.0	1.000	208	1,345	54	1,399
Non-zero level of savings	5.011	19.712	0.0001	370.2	1.227	1	1,136	54	1,191
HH decision-making	0.841	0.261	0.0000	1.0	1.000	46	1,345	54	1,399
HH welfare	−0.064	1.680	−3.382	10.571	−0.384	0	1,345	54	1,399

	Mean	Standard Deviation	Minimum Value	Maximum Value	Median Value	No. of zero values	No. of Reported Values	No. of Non-Responses	Total Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Second follow-up survey									
Profits	1.755	9.347	0.0000	465.7	0.931	1	2,655	186	2,841
Capital inputs	27.920	90.800	0.0000	2355.1	7.591	3	2,600	241	2,841
Labor inputs	198.633	98.004	0.0000	570.0	182.000	1	2,679	162	2,841
Business practices	0.283	0.227	0.0000	1.0	0.231	101	2,679	162	2,841
Any savings	0.914	0.280	0.0000	1.0	1.000	234	2,724	117	2,841
Non-zero level of savings	7.565	14.155	0.0092	254.1	3.165	4	2,457	146	2,607
HH decision-making	0.754	0.318	0.0000	1.0	1.000	181	2,724	117	2,841
HH welfare	−0.044	1.492	−4.175	9.731	−0.246	0	2,724	117	2,841
Pooled data									
Profits	1.721	6.280	0.0000	465.7	0.946	3	6,802	290	7,092
Capital inputs	24.489	82.126	0.0000	2355.1	5.972	13	6,609	483	7,092
Labor inputs	223.111	104.107	0.0000	720.0	210.000	2	6,853	239	7,092
Business practices	0.255	0.209	0.0000	1.0	0.154	225	5,525	168	5,693
Any savings	0.869	0.337	0.0000	1.0	1.000	906	6,919	173	7,092
Non-zero level of savings	6.892	15.250	0.0001	370.2	2.602	15	5,969	200	6184
HH decision-making	0.787	0.299	0.0000	1.0	1.000	374	6,915	177	7,092
HH welfare	−0.013	1.670	−5.322	10.571	−0.190	0	6,916	176	7,092

NA = no data collected. HH = household.

Notes: The data on outcomes in this table are in their original (pre-transformed) metrics, with any subsequent transformations described in section 5.1. Detailed information on the individual outcomes is provided in Annex 1. Non-responses in column 8 include 54 WBOs in the first follow-up survey and 117 WBOs in the second follow-up survey who were interviewed in the baseline survey but who could not be interviewed in the follow-up survey. Units for profits, capital inputs and non-zero levels of savings are Rupiah (Rp.) millions in November 2016 Surabaya Municipality prices, units for labor inputs are number of days, units for business practices, any savings, and household decision-making are proportions. The values reported for household welfare are the predicted scores of the first principal components of indicators referring to household ownership of 20 durable goods.

TABLE 5. Multivariate analysis of attrition in the two follow-up surveys

Dependent variables →	Overall sample attrition	Attrition of business profits[a]	Attrition of business capital[a]	Overall sample attrition	Attrition of business profits[a]	Attrition of business capital[a]
Estimation method →	OLS	OLS	OLS	VL_FE[b]	VL_FE[b]	VL_FE[b]
	(1)	(2)	(3)	(4)	(5)	(6)
TREATMENTS						
High agent incentives	0.000	0.008	0.010	NA[c]	NA[c]	NA[c]
	(0.007)	(0.009)	(0.015)			
WBO training & mentoring	−0.008	−0.018*	−0.025**	−0.008	−0.019*	−0.024**
	(0.008)	(0.010)	(0.012)	(0.008)	(0.010)	(0.012)
COVARIATES (Baseline values)						
Location (urban = 1, rural = 0)	−0.007	0.001	−0.030	NA[c]	NA[c]	NA[c]
	(0.012)	(0.014)	(0.027)			
WBO's age	0.005	0.004	0.004	0.002	0.004	0.004
	(0.004)	(0.005)	(0.007)	(0.004)	(0.005)	(0.006)
WBO's age squared	−0.000	−0.000	0.000	−0.000	−0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
WBO is currently married	0.001	−0.015	−0.024	−0.003	−0.019	−0.030
	(0.011)	(0.016)	(0.023)	(0.012)	(0.018)	(0.023)
WBO has children	0.004	0.012	0.007	0.008	0.014	0.019
	(0.010)	(0.013)	(0.015)	(0.011)	(0.014)	(0.014)
Number of WBO's children	−0.011**	−0.016**	−0.009	−0.008	−0.011*	−0.001
	(0.005)	(0.006)	(0.009)	(0.005)	(0.007)	(0.009)
WBO's household size	−0.002	−0.001	0.000	−0.002	0.000	0.001
	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.004)
WBO's highest completed level of schooling	0.001	0.012**	0.022***	0.001	0.014**	0.020**
	(0.004)	(0.005)	(0.008)	(0.006)	(0.007)	(0.009)
WBO's cognitive score (1–4)	−0.007	−0.011*	−0.027***	−0.008	−0.009	−0.016*
	(0.004)	(0.006)	(0.008)	(0.005)	(0.006)	(0.009)
WBO owns a smart phone	0.010	0.005	0.013	−0.001	−0.008	0.005
	(0.007)	(0.010)	(0.012)	(0.008)	(0.010)	(0.013)
WBO knows partner bank agent's name	0.001	−0.008	−0.025*	0.004	−0.006	−0.032**
	(0.008)	(0.010)	(0.015)	(0.009)	(0.012)	(0.016)
Number of other sample WBOs who are known by WBO	−0.005*	−0.007*	−0.005	−0.005	−0.009**	−0.006
	(0.003)	(0.004)	(0.005)	(0.003)	(0.004)	(0.005)
Number of other sample WBOs who are close friends, relatives or family members of WBO	−0.004	−0.001	−0.005	−0.005	−0.005	−0.000
	(0.004)	(0.005)	(0.007)	(0.004)	(0.005)	(0.007)
Number of other sample WBOs with whom WBO talks weekly	−0.001	−0.001	−0.002	0.001	0.002	0.001
	(0.003)	(0.004)	(0.006)	(0.003)	(0.004)	(0.006)

Dependent variables →	Overall sample attrition	Attrition of business profits[a]	Attrition of business capital[a]	Overall sample attrition	Attrition of business profits[a]	Attrition of business capital[a]
Estimation method →	OLS	OLS	OLS	VL_FE[b]	VL_FE[b]	VL_FE[b]
	(1)	(2)	(3)	(4)	(5)	(6)
Number of other sample WBOs with whom WBO talks weekly about business matters	−0.001	−0.005	−0.010*	0.002	0.002	−0.005
	(0.003)	(0.003)	(0.005)	(0.003)	(0.004)	(0.006)
Primary business is registered	−0.001	−0.008	−0.028	0.004	−0.011	−0.016
	(0.011)	(0.013)	(0.019)	(0.012)	(0.015)	(0.019)
OTHER STATISTICS						
Joint test: randomized treatments = 0 (p)	0.548	0.107	0.083*	0.289	0.056	0.034**
Joint test: covariates = 0 (p)	0.029**	0.000***	0.000***	0.280	0.034	0.001***
Sample mean	0.040	0.070	0.127	0.040	0.070	0.127
R-squared	0.014	0.019	0.026	0.010	0.013	0.018
Sample size	4240	4240	4240	4240	4240	4240

*** p<0.01 ** p<0.05 * p<0.10.

Notes: Estimation sample includes data from both follow-up surveys. Estimated standard errors are adjusted for heteroscedasticity and for serial correlation and clustering at the village level.

[a] Includes attrition in baseline values. [b] Village-level fixed-effects. [c] Not available (village-level variable).

TABLE 6. Village-level fixed-effects estimates of Model 1 (treatment effects permitted to vary freely between six-month post-treatment exposure intervals)

	Profits (IHS)	Business Capital (IHS)	WBO Labor	Business Practices	Any Savings (0–1)	Non-Zero Savings (log)	Agency	Household Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TREATMENT EFFECTS								
DA1: 1–6 months	−0.116	−0.040	−5.761	−0.019	−0.006	−0.168	0.047*	0.129
	(0.098)	(0.136)	(8.988)	(0.026)	(0.036)	(0.157)	(0.026)	(0.130)
DA2: 7–12 months	−0.041	0.035	6.733	0.026	0.009	0.055	−0.004	−0.033
	(0.069)	(0.115)	(7.623)	(0.018)	(0.024)	(0.105)	(0.025)	(0.101)
DA:3 13–18 months	−0.042	−0.086	−9.502	−0.012	0.031	−0.142	0.050**	0.345***
	(0.071)	(0.101)	(7.191)	(0.016)	(0.026)	(0.103)	(0.023)	(0.106)
DA4: 19–24 months	0.011	−0.049	9.891	−0.015	0.020	−0.203	0.040	0.140
	(0.115)	(0.167)	(10.403)	(0.026)	(0.036)	(0.173)	(0.031)	(0.144)
DW1: 1–6 months	0.040	0.096	−8.621	NA	0.026	0.088	0.062***	0.212**
	(0.091)	(0.111)	(8.573)		(0.028)	(0.136)	(0.023)	(0.099)
DW2: 7–12 months	0.016	0.089	1.861	0.019	0.029	0.132*	0.028	−0.057
	(0.061)	(0.100)	(6.265)	(0.014)	(0.018)	(0.077)	(0.018)	(0.070)
DW3: 13–18 months	0.165**	−0.061	−3.541	0.014	0.017	0.072	0.024	0.321***
	(0.068)	(0.104)	(6.370)	(0.015)	(0.021)	(0.103)	(0.020)	(0.084)

	Profits (IHS)	Business Capital (IHS)	WBO Labor	Business Practices	Any Savings (0–1)	Non-Zero Savings (log)	Agency	Household Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DW4: 19–24 months	0.288**	0.210	–12.666	0.034	0.023	0.131	0.023	0.359**
	(0.135)	(0.200)	(11.178)	(0.022)	(0.034)	(0.179)	(0.033)	(0.157)
DAW1: 1–6 months	0.137	0.112	18.381	–0.048	–0.031	0.095	–0.049	–0.134
	(0.131)	(0.175)	(11.886)	(0.038)	(0.042)	(0.187)	(0.031)	(0.139)
DAW2: 7–12 months	0.117	–0.026	5.013	0.001	–0.017	–0.031	–0.028	–0.013
	(0.087)	(0.148)	(8.699)	(0.021)	(0.027)	(0.112)	(0.027)	(0.104)
DAW3: 13–18 months	0.008	0.233	23.935**	0.039*	–0.010	–0.034	–0.013	–0.349***
	(0.107)	(0.160)	(9.408)	(0.022)	(0.033)	(0.141)	(0.031)	(0.132)
DAW4: 19–24 months	–0.385*	–0.509	2.528	–0.005	0.040	0.183	–0.104*	–0.010
	(0.200)	(0.324)	(16.132)	(0.041)	(0.061)	(0.248)	(0.059)	(0.232)
JOINT TESTS								
No agent incentives effect (DA1–DA4 = 0)	0.818	0.809	0.168	0.236	0.763	0.204	0.085*	0.001***
No WBO training effect (DW1–DW4 = 0)	0.022**	0.444	0.456	0.265	0.535	0.548	0.090*	0.000***
No interaction effect (DAW1–DAW4 = 0)	0.041	0.235	0.122	0.188	0.829	0.877	0.197	0.108
No combined effect (ΣDA+ΣDW+ ΣDAW = 0)	0.431	0.992	0.262	0.567	0.163	0.613	0.364	0.012**
Equal effects across intervals (DA1 = DA2 = DA3 = DA4)[a]	0.835	0.705	0.093*	0.138	0.691	0.155	0.123	0.001***
Equal effects across intervals (DW1 = DW2 = DW3 = DW4)[a]	0.020**	0.330	0.421	0.703	0.952	0.944	0.411	0.000***
Equal effects across intervals (DAW1 = DAW2 = DAW3 = DAW4)[a]	0.039**	0.157	0.362	0.284	0.787	0.752	0.356	0.105
OTHER STATISTICS								
Baseline sample mean	7.497	9.191	244.855	0.228	0.834	1.012	0.794	0.040
Baseline standard deviation	1.125	1.661	104.188	0.186	0.372	1.478	0.293	1.818
Baseline RMSE	0.917	1.406	73.359	0.142	0.274	1.233	0.229	0.966
R-squared	0.178	0.190	0.435	0.481	0.289	0.230	0.336	0.593
Sample size	6780	6387	6853	5525	6919	5969	6915	6916

*** p<0.01 ** p<0.05 * p<0.10.

Notes: IHS = inverse hyperbolic sine transformation applied. NA = not available (no data on business practices were collected in the first follow-up survey). The baseline sample means are in their pre-transformation metrics, with profits, capital inputs and the non-zero level of savings in Rp. millions in November 2016 Surabaya Municipality prices.

[a] test of Model 1 vs constant impact model.

TABLE 7. Village-level fixed-effects estimates of Model 2 (Linear trends in treatment effects)

	Profits (IHS)	Business Capital (IHS)	WBO Labor	Business Practices	Any Savings (0–1)	Non-zero Savings (log)	Agency	Household Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TREATMENT EFFECTS								
DA	−0.0016	−0.0035	−0.1818	−0.0007	0.0018	−0.0083	0.0025*	0.0149**
	(0.0044)	(0.0063)	(0.4065)	(0.0008)	(0.0015)	(0.0061)	(0.0014)	(0.0060)
DW	0.0089**	0.0035	−0.3106	0.0011	0.0019	0.0074	0.0022*	0.0147***
	(0.0043)	(0.0067)	(0.3515)	(0.0008)	(0.0012)	(0.0055)	(0.0012)	(0.0049)
DAW	0.0005	0.0018	1.1544**	0.0016	−0.0007	0.0019	−0.0031*	−0.0129*
	(0.0065)	(0.0100)	(0.4968)	(0.0012)	(0.0019)	(0.0073)	(0.0018)	(0.0072)
JOINT TESTS								
Agent incentives have zero effect (DA = 0)	0.716	0.574	0.655	0.438	0.219	0.173	0.063*	0.013**
WBO training has no effect (DW = 0)	0.041**	0.605	0.377	0.139	0.106	0.178	0.064*	0.003***
No interaction effect (DAW = 0)	0.933	0.855	0.021**	0.171	0.716	0.793	0.078*	0.075*
No combined effect (DA + DW + DAW = 0)	0.071*	0.802	0.138	0.018**	0.053*	0.873	0.252	0.006***
DA effect is constant across six-month intervals (test of Model 2 vs Model 1)	0.696	0.630	0.079*	0.155	0.693	0.118	0.193	0.001***
DW effect is constant across six-month intervals (test of Model 2 vs Model 1)	0.104	0.858	0.595	0.665	0.706	0.974	0.093*	0.001***
DAW effect does not vary across six-month intervals (test of Model 2 vs Model 1)	0.015**	0.150	0.444	0.709	0.993	0.762	0.108	0.092*

	Profits (IHS)	Business Capital (IHS)	WBO Labor	Business Practices	Any Savings (0–1)	Non-zero Savings (log)	Agency	Household Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OTHER STATISTICS								
Baseline sample mean	7.497	9.191	244.855	0.228	0.834	1.012	0.794	0.040
Baseline standard deviation	1.125	1.661	104.188	0.186	0.372	1.478	0.293	1.818
Baseline RMSE	0.917	1.407	73.409	0.142	0.274	1.233	0.229	0.969
R-squared	0.177	0.189	0.433	0.480	0.289	0.228	0.334	0.590
Sample size	6780	6387	6853	5525	6919	5969	6915	6916

*** p<0.01 ** p<0.05 * p<0.10.

Notes: IHS = inverse hyperbolic sine transformation applied. RMSE = root mean square error. NA = not available (no data on business practices were collected in the first follow-up survey). The baseline sample means are in their pre-transformation metrics, with profits, capital inputs and the non-zero level of savings in Rp. millions in November 2016 Surabaya Municipality prices.

TABLE 8. Village-level fixed-effects estimates of Model 3 (Quadratic trends in treatment effects)

	Profits (IHS)	Business Capital (IHS)	WBO Labor	Business Practices	Any Savings (0–1)	Non-Zero Savings (log)	Agency	Household Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TREATMENT EFFECTS								
DA	−0.0155	−0.0182	0.2764	0.0037	−0.0013	−0.0017	0.0006	0.0019
	(0.0107)	(0.0168)	(1.1868)	(0.0030)	(0.0039)	(0.0166)	(0.0038)	(0.0150)
DA-squared (DA^2)	0.0008	0.0009	−0.0251	−0.0002	0.0002	−0.0004	0.0001	0.0007
	(0.0006)	(0.0009)	(0.0620)	(0.0002)	(0.0002)	(0.0009)	(0.0002)	(0.0007)
DW	−0.0034	−0.0028	0.8304	0.0023	0.0049	0.0152	0.0064**	−0.0166
	(0.0109)	(0.0164)	(1.1257)	(0.0029)	(0.0032)	(0.0146)	(0.0032)	(0.0126)
DW-squared (DW^2)	0.0008	0.0004	−0.0771	−0.0001	−0.0002	−0.0005	−0.0003	0.0021***
	(0.0006)	(0.0010)	(0.0679)	(0.0002)	(0.0002)	(0.0009)	(0.0002)	(0.0008)
DAW	0.0410**	0.0460*	0.5498	−0.0012	−0.0021	−0.0074	−0.0014	0.0069
	(0.0169)	(0.0265)	(1.6873)	(0.0046)	(0.0057)	(0.0225)	(0.0055)	(0.0208)
DAW-squared (DAW^2)	−0.0028***	−0.0031*	0.0392	0.0002	0.0001	0.0006	−0.0001	−0.0013
	(0.0010)	(0.0016)	(0.1011)	(0.0003)	(0.0004)	(0.0014)	(0.0004)	(0.0013)
JOINT TESTS								
Agent incentives have zero effect ($DA = DA^2 = 0$)	0.326	0.550	0.835	0.258	0.260	0.373	0.117	0.020**
WBO training has no effect ($DW = DW^2 = 0$)	0.047**	0.777	0.278	0.281	0.196	0.353	0.084*	0.001***
No interaction effect ($DAW = DAW^2 = 0$)	0.025**	0.142	0.041**	0.374	0.895	0.876	0.158	0.150

	Profits (IHS)	Business Capital (IHS)	WBO Labor	Business Practices	Any Savings (0–1)	Non-Zero Savings (log)	Agency	Household Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No combined effect ($DA + DA^2 + DW + DW^2 + DAW + DAW^2 = 0$)	0.061*	0.189	0.174	0.138	0.698	0.713	0.128	0.672
$DA^2 = DW^2 = DAW^2 = 0$ (test of Model 3 vs Model 2)	0.058*	0.219	0.627	0.511	0.419	0.934	0.210	0.046**
OTHER STATISTICS								
Baseline sample mean	7.497	9.191	244.855	0.228	0.834	1.012	0.794	0.040
Baseline standard deviation	1.125	1.661	104.188	0.186	0.372	1.478	0.293	1.818
Baseline RMSE	0.917	1.406	73.409	0.142	0.274	1.233	0.229	0.968
R-squared	0.178	0.190	0.433	0.480	0.289	0.228	0.335	0.591
Sample size	6780	6387	6853	5525	6919	5969	6915	6916

*** p<0.01 ** p<0.05 * p<0.10.

Notes: IHS = inverse hyperbolic sine transformation applied. RMSE = root mean square error. NA = not available (no data on business practices were collected in the first follow-up survey). The baseline sample means are in their pre-transformation metrics, with profits, capital inputs and the non-zero level of savings in Rp. millions in November 2016 Surabaya Municipality prices.

TABLE 9. Structural equation model (SEM) estimates: direct effects (N = 5,098)

	Business Profits (IHS)	Capital (IHS)	WBO Labor	Business Practices	Personal SSavings	Agency	Household Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Business profits (IHS)		0.1274***	0.0439***	0.0183	0.0975***	0.0136	0.0423***
Business capital (IHS)	0.1202***			−0.0254			
WBO labor	−0.0083						−0.0120
Business practices	0.0755***	0.1551***	0.0173*				
Personal savings		0.0812***					
Agency	0.0318**	0.0048	0.0183*		0.0127		0.0114
Household income		0.1738***	−0.0301***		0.0578***	0.0015	
DA	0.0103	−0.0118	−0.0063	−0.0287*	−0.0222	0.0465**	0.0472***
DW	0.0759***	0.0007	−0.0176	0.0330*	0.0269	0.0310	0.0473***
DAW	−0.0217	−0.0058	0.0402**	0.0494***	0.0012	−0.0415**	−0.0307**
Sample size	5,098	5,098	5,098	5,098	5,098	5,098	5,098

*** p<0.01 ** p<0.05 * p<0.10.

Notes: IHS = inverse hyperbolic sine transformation applied. Endogenous variable names are in italics. Estimation sample does not include data from FS1.

TABLE 10. Structural equation model (SEM) estimates: Total (i.e., direct + indirect) effects (N = 5,098)

Variable	*Business profits (IHS)*	*Business capital (IHS)*	*WBO labor*	*Business practices*	*Personal savings*	*Agency*	*Household income*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Business profits	0.0190***	0.1479***	0.0439***	0.0149	0.1020***	0.0140	0.0427***
Business capital	0.1200***	0.0135***	0.0047**	−0.0235	0.0120***	0.0016	0.0050***
WBO labor	−0.0088	−0.0034	0.0000	−0.0001	−0.0016	−0.0001	−0.0124
Business practices	0.0954***	0.1683***	0.0213**	−0.0025	0.0095***	0.0013	0.0038***
Personal savings	0.0097***	0.0823***	0.0004*	−0.0019	0.0010***	0.0001	0.0004***
Agency	0.0331**	0.0126	0.0194*	0.0003	0.0167	0.0005	0.0126
Household income	0.0217***	0.1811***	−0.0293**	−0.0042	0.0600***	0.0018	0.0013**
DA	0.0088	−0.0080	−0.0070	−0.0284**	−0.0179	0.0467**	0.0482***
DW	0.0831***	0.0287	−0.0143	0.0338*	0.0384**	0.0322*	0.0513***
DAW	−0.0205	−0.0069	0.0404**	0.0492**	−0.0033	−0.0418**	−0.0325**
Sample size	5,098	5,098	5,098	5,098	5,098	5,098	5,098

*** p<0.01 ** p<0.05 * p<0.10.

Notes: IHS = inverse hyperbolic sine transformation applied. Endogenous variable names are in italics. Estimation sample does not include data from FS1.

Annex 1. Description of the outcomes analyzed

Profits in all survey rounds is average monthly profits during the past 12 months in the primary and secondary business combined in millions of Indonesia Rupiah (Rp. millions) in November 2016 Surabaya Municipality prices.[35] It is based on a single question and with 3 zero responses (1 in the baseline survey, 1 in the first follow-up survey and 1 in the second follow-up survey) and no negative responses. Several WBOs failed to respond to the question on business profits: 17 in the baseline survey, 87 in the first follow-up survey, and 186 in the second follow-up survey.[36]

Capital inputs is the value of all business capital in all businesses owned by the WBO in millions of Indonesia Rupiah (Rp. millions) in November 2016 Surabaya Municipality prices. It is based on questions of whether the WBO has any of six different types of business capital (i.e., own shop premises, advances paid for rented shop, furniture and fixtures, equipment, inventories, and other), and if so, its current value (i.e., "how much would you sell this for?"). There are only 13 zero values in the follow-up surveys (10 in the first follow-up survey and 3 in the second follow-up survey), but there are significant numbers of non-responses (mostly responses of "don't know" with relatively few "refuse to answer" responses): 167 in the baseline survey, 75 in the first follow-up survey, and 241 in the second follow-up survey.

Labor inputs is the number of hours worked in a typical month by the WBO in her primary business.[37] This variable is the product of responses to the following two questions: (1) "How many days do you work in your primary business during a typical month?" and (2) "How many hours per day do you work in your primary business?" Many WBOs responded that they work 30 days per month (5,132), while 57 WBOs reported that they worked 16 or more hours per day. There are only two zero responses (1 each in the first and second follow-up surveys). Although only 3 WBOs did not report labor inputs in the baseline survey, 74 did not respond in the first follow-up survey and 162 did not respond in the second follow-up survey.

Business practices is the mean proportion of WBOs indicating that they adhere to the following 13 recommended business practices: (1) ask a supplier which products are selling well in your industry, (2) used a special offer to attract customers in the last three months, (3) done any form of advertising in the last six months, (4) done anything to measure the effectiveness of advertising (coded zero if (3) is no), (5) attempted to negotiate with a supplier for a lower price in the last three months, (6) have a record-keeping system that allows you to know your current inventory, (7) keep written business records, (8) record every purchase or sale (coded zero if (7) is no), (9) able to know cash on hand at any point in hand (coded zero if (7) is no) (10) use records to know whether sales of a particular product are

35 Profits in the secondary business were assumed to have been zero if the WBO reported not having any secondary business.

36 The number of non-responses in the two follow-up surveys include 54 WBOs who could not be interviewed in the first follow-up survey and 177 WBOs who could not be interviewed in the second follow-up survey.

37 WBOs total labor inputs in both primary and secondary businesses is not meaningful because of the tendency of WBOs to report the same labor inputs in both businesses.

increasing or decreasing from one month to another (coded zero if (7) is no), (11) worked out the cost of each product sold, (12) have a written monthly budget, and (13) have records needed to apply for a bank loan. Only six WBOs did not respond to these questions in the baseline survey, while 162 did not respond in the second follow-up survey (no data on adherence to recommended business practices were collected in the first follow-up survey). There are 124 zero responses in the baseline survey and 101 in the second follow-up survey.

Savings refers to WBOs' total reported savings during the past 12 months (or past 3 months in the first follow-up survey) in all savings instruments. It is based on responses to a question whether any amount was saved during the reference period (any savings) and if so, how much was saved in each of several alternative savings instruments (non-zero level of savings in millions of Indonesia Rupiah in November 2016 Surabaya Municipality prices). The list of alternative savings instruments varies to some extent between survey rounds. In the baseline survey, they include: (1) formal bank account, (2) electronic savings account, (3) hiding place at home, (4) with friends or family, (5) cooperative, (6) informal saving network, (7) community savings fund (BMT), (8) rotating saving/credit association (ROSCA), and (9) other. The following savings instruments were added in the first follow-up survey: (10) other household members' saving, (11) physical assets (e.g., jewelry), (12) *laku pandai* (LP)[38] savings account of the partner bank, (13) electronic savings account of another bank,[39] and (14) LP savings account of another bank. In the second follow-up survey, the savings instruments included those in the baseline survey plus (10) LP savings account of the partner bank and (11) cash wallet account of the partner bank, while saving instrument (2) was changed to "electronic savings account in other banks." The change in the reference period in the first follow-up survey and the changes in the lists of savings instruments need to be taken into account in the analysis. Although all but two WBOs responded to both savings questions in the baseline survey, 54 did not respond to either question on savings in the first follow-up survey, while 117 did not respond to the question on any savings and 146 did not respond to the question on the non-zero level of savings in the second follow-up survey.

Agency is the average proportion of WBOs who indicate that that they are either the sole decision-maker or participate with their spouse/partner or with another household member in the following five decisions: (1) whether to purchase an appliance for the home, (2) how household members may work outside the home, (3) whether to support other family members, (4) whether to save for the future, and (5) whether to sign up for a new banking product. There are 147 zero responses in the baseline survey, 46 in the first follow-up survey and 374 in the second follow-up survey. Although all but six WBOs responded to these questions in the baseline survey, 54 did not respond to all of these questions in the first follow-up survey, and 117 did not respond to them in the second follow-up survey (i.e., the WBOs who were not interviewed).

38 *laku pandai* is an Indonesian acronym for "branchless banking."

39 Instrument (2) in the baseline survey was also changed to "electronic savings account of the partner bank" in the first follow-up survey.

Household income is measured by an asset index, based on the household's reported ownership (yes-no) of 20 consumer durables. The asset index is calculated as the first principal component of the available indicators, as is the usual practice (Filmer and Scott 2012). Although all but five WBOs responded to the questions on durable goods ownership in the baseline survey, while 54 did not respond to the same questions in the first follow-up survey, and 117 did not respond in the second follow-up survey (i.e., the WBOs who were not interviewed).

Annex 2. Re-transformation of estimates to their original metrics

The estimated average treatment effects, when the treatments are represented by dummy variables (as in Model 1), can be converted to approximate percentage changes from the baseline sample mean value using the following transformation (Kennedy 1981):

$$P \approx 100^*(\exp(\hat{\beta} - 0.5\,\hat{\mathrm{Var}}(\hat{\beta})) - 1) \qquad \text{(A2-1)}$$

where P is the percentage change estimate, $\hat{\beta}$ is the estimated treatment effect in the IHS (or natural log) metric, and $\hat{\mathrm{Var}}(\hat{\beta})$ is its estimated standard error squared.

When the treatment effects are represented by linear or quadratic time trends (as in Models 2 and 3) and the dependent variable is transformed to an IHS value (profits, capital inputs) or to a natural log value (non-zero level of savings), the projected percentage changes can be obtained using the following formula:

$$P_t = 100^*[(\exp(\mathbf{x}_t\hat{\boldsymbol{\theta}}t + 0.5^*(\mathrm{RMSE})^2) - 1]/\bar{Y}_{t=0} - P_{t=0} \qquad (t = 0, 1, \ldots, 24) \qquad \text{(A2-2)}$$

where t refers to month, $\mathbf{x}$ refers to the row vector of treatment measures in Models 2 and 3 (e.g., DA in Model 2 or both DA and DA^2 in Model 3), $\hat{\boldsymbol{\theta}}$ refers to the vector of estimated treatment effects ($\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$), RMSE refers to the root mean square error, and $\bar{Y}_{t=0}$ refers to the baseline sample mean value of the outcome.[40]

40 The formula in equation (A2-2) assumes that the projections are based on the conditional mean, rather than the conditional median. If Y has a lognormal distribution with mean μ and constant variance σ^2, $\exp(\mu)$ is the conditional median, while the conditional mean is equal to $(\exp(\mu + 0.5^*\sigma^2)$ (Goldberger 1968).