

**Growing Markets through Business Training for Female Entrepreneurs:
A Market-Level Randomized Experiment in Kenya***

David McKenzie, *Development Research Group, World Bank*
Susana Puerto, *Employment Policy Department, International Labour Organization*

Abstract

A common concern with efforts to directly help some small businesses to grow is that their growth comes at the expense of their unassisted competitors. We test this possibility using a two-stage randomized experiment in Kenya which randomizes business training at the market level, and then within markets to selected businesses. Three years after training, the treated businesses are selling more, earn higher profits, and their owners have higher well-being. There is no evidence of negative spillovers on the competing businesses, and the markets as a whole have grown in terms of sales volume.

Keywords: Business Training, Spillovers, Microenterprise, Market Development

JEL codes: O12, O17, J16, L26.

* We gratefully acknowledge funding provided to this project from the International Initiative for Impact Evaluation (3ie), Private Enterprise Development in Low-Income Countries (PEDL), the International Labour Organization (ILO), the Strategic Research Program (SRP), and the World Bank Jobs Umbrella multi-donor trust fund. Human Subjects Approval was obtained from Innovations for Poverty Action (IPA, 13February-002) and the Maseno University Ethics Review Committee (MSU/DRPC/MUERC/000006/13). Authority to conduct research in Kenya was provided by the Kenyan Ministry of Science and Technology (NCST/RCD/14/013/553B). A pre-analysis plan and the associated trial were registered on the AEA Social Science Registry on February 21, 2014 (AEARCTR-0000287). We thank Valerie Breda, Faizan Diwan, Bilal Hakeem, Mary Kamore, Virginia Losada, Jane Maigua, Grace Makana, Frank Odhiambo, Silvia Paruzzolo, and Miriam Wekesa for their contributions to the study design and implementation, ICRW for their qualitative work, and participants at various seminars for useful comments.

1. Introduction

Governments and NGOs around the world promote and offer a variety of direct support to small businesses, with the provision of business training one of the most common services provided. Better business practices are strongly associated with better firm performance across a range of countries in both the cross-section and over time (McKenzie and Woodruff, 2015), and so the hope is that teaching these practices will help firms grow. However, a key concern around public or subsidized provision of business training is whether any growth of trained firms comes at the expense of their competitors (Rotemberg, 2014). This concern is particularly apparent when working with microenterprises in rural markets in developing countries, where it is easy to believe that if firms are all selling similar products in a small market, any extra sales made by trained firms must come from competing away these sales from neighbouring untrained firms. Conversely, if better management practices are best thought of as a technology (as in Bloom et al., 2015), then there might be positive spillover benefits as untrained firms copy the practices adopted by the trained and also become more productive. Furthermore, if training leads firms to expand the variety of products being offered and the customer service provided, the overall market size may increase, particularly in underdeveloped marketplaces.

We conduct a randomized experiment in 157 rural markets in Kenya to test how business training (the International Labour Organization (ILO) *Gender and Enterprise Together* program) affects the profitability, growth and survival of female-owned businesses, and to evaluate whether any gains in profitability come at the expense of other business owners. We work with a large sample of 3,537 firms, and use a two-stage randomization, first randomizing at the market-level, and then randomizing the offer of training to individuals within treated markets. A year and a half after the training has taken place, half of the sample assigned to training was then offered a subsequent mentoring intervention intended to test whether additional group-based and in-person support strengthens the impacts of training. Four rounds of follow-up surveys with low attrition are used to measure impacts at one and three years after training. This is complimented with data from a market census taken four years after training, that also included male-operated firms.

Firms assigned to training are 3 percentage points more likely to survive after three years, earn 18 percent higher sales, and make 15 percent higher profits. Their owners have better mental health

and a higher subjective standard of living. These gains are greater at three years than at one year after the training, persist into a fourth year, and are similar for firms assigned to training only as for firms also assigned to a mentor. These gains come with no significant spillover effects on untreated firms operating in the same markets, and total sales and the total number of customer transactions is higher in the treated markets than control markets. We also find no reduction in new entry into these markets after training, and the total sales of surviving firms, including those operated by males, is higher in treated markets four years after training. This market growth appears to stem from better customer services, better business practices, the introduction of new products, and multiplier effects from firms acting as customers to one another, with no significant impacts on access to finance or input management. We conclude that, in underdeveloped markets, microenterprise growth need not come at the expense of competitors, and business training can help the overall market grow.

There are three strands of this literature of particular relevance to this study. The *first* is evidence on the effectiveness of business training. McKenzie and Woodruff (2014) review this literature and note that overall evidence on effectiveness is mixed, in part because many studies have low statistical power and measure impacts over short durations. This existing literature also suggests that business training may be less effective for female business owners, either because they work in sectors with very low efficient scales or because they face many other constraints that limit the ability of their businesses to grow (e.g. de Mel et al., 2014; Berge et al. 2016). Get-Ahead, the training program studied here was designed especially for women with low education levels. Bulte et al. (2016) evaluate the impact of this same program on female microfinance clients in Vietnam and find some evidence of firm growth 12 months after training. Our results show stronger and longer-term evidence that this training can help female-owned businesses grow.

Second, several studies examine whether augmenting standard business training with mentoring can enhance its effectiveness. Valdivia (2015) find that individualized hand-holding or mentoring after training does not have sustained impacts relative to training alone. Brooks et al. (2016) find, also in Kenya, that assigning a mentor to a young firm

does increase profits in the short-run, but the effect fades over time. Our results are consistent with mentoring not delivering additional gains compared to training alone.

Third, almost all of the existing literature to date has been unable to measure impacts at the market level, so cannot tell if any gains come from untrained firms. The one exception is Ecrf gtóp"gv'crf (2013), who worked with 17 villages in rural Mexico, assigning 7 to treatment and 10 to control. They find no significant spillovers, although this may in part reflect low statistical power given the small number of villages and that they lose 18 percent of their sample to attrition and 41 percent to closure by their second follow-up. Our paper builds on this work and not only shows no significant spillovers, but significant growth at the market level.

The remainder of the paper is structured as follows: Section 2 discusses the selection of our sample and randomization procedure; Section 3 the training and mentoring interventions; Section 4 our data collection and estimation approach; Section 5 impacts on primary and secondary outcomes; Section 6 mechanisms for these results; Section 7 cost-effectiveness; and Section 8 concludes.

2. Selection of the Sample, Randomization Procedure, and Baseline Characteristics

2.1 Selecting a Sample

A participatory process involving the ILO, government ministries, and organizations serving female entrepreneurs¹ was used to select four counties for the study: Kakamega and Kisii in the Western region, and Embu and Kitui in the Eastern region.² These regions are largely rural, with an average population for each county of approximately 1 million, and the majority of the population below the poverty line. In each of the four counties, field staff from Innovations for Poverty Action, Kenya, mapped out all market centers deemed as *medium* or *large* outside of the main cities and conducted a market census of all female-owned businesses between June and November 2013 (see timeline in Appendix 1). After listing, an eligibility filter was applied

¹ Stakeholders consulted included the Department of Micro and Small Enterprise Development (DMSED) of the Ministry of Labour, Ministry of Youth Affairs, Ministry of Cooperative Development and Marketing, Ministry of Youth, Federation of women entrepreneurs associations (FEWA), Women Enterprise Fund (WEF), Youth Employment Development Fund (YEDF), Business Development Service providers, and Inoorero University.

² Kitui and Embu average 150 km distance from Nairobi, and Kakamega and Kisii 335km.

(Appendix 2) and a baseline survey administered to 3,537 individuals located in 157 separate markets.

These markets are typically small and remote, largely consisting of women operating a limited variety of businesses such as selling fruits, vegetables, grains, and dried fish products from tables; and offering services like hairdressing, dressmaking and small food kiosks. Our market census taken in mid-2017 (discussed below) finds female-owners to be 89% of the fruit and vegetable sellers, 87% of the grain sellers, 81% of the fish sellers and 59% of the restaurant owners in these markets.³ The market is a designated place in the village, and authorities typically build several stalls for the sellers to display their goods and services, and charge them a market fee to sell in the market.⁴ The average market in our study has 22 female-operated firms surveyed in it, with 75 percent in retail trade and 25 percent in services.⁵

2.2 Randomization Procedure

Individuals were then assigned to treatment and control for the business training intervention in a two-stage process, diagrammed in Figure 1. First, **Markets** were assigned to treatment (have some individuals in them invited to training) or control (no one in the market would be invited to training) status. Randomization was done within 35 strata defined by geographical region (within county) and the number of women surveyed in the market. The need to ensure sufficient numbers for training meant that more than half the markets were allocated to treatment, with 93 markets assigned to treatment and 64 to control.

Then within each market, **Individuals** were assigned to treatment (be invited to training) or control (not be invited to training) within treated markets by forming four strata, based on quartiles of weekly profits from the census (≤ 450 , 451-800, 801-1500, 1501-4000), and then assigning half

³ While men operate some businesses in these markets, they tend to specialize in other products such as butchery, carpentry, bicycle and motorcycle repair, artisans, and shoe repair. Women are 63% of the tailors and 43% of the hairdressers and barbers, but within these two sectors women tend to serve female clients and men male clients.

⁴ These structures are moveable, and business owners are typically free to move around the market as they like so long as they rent the market spot from the local authority. We include all women operating firms in the vicinity of the market, regardless of whether or not they are officially registered.

⁵ In addition, markets typically have a market day once a week, where outside vendors come to sell goods such as clothing, plastic housewares, and shoes. These vendors who are only in the market one day a week are not included in our study.

the individuals within each strata to training. This resulted in 1172 of the 2160 individuals in treated markets being assigned to treatment, and 988 to control.⁶

For the follow-up mentoring intervention, we dropped markets with 15 or fewer firms in the study, since these would have too few firms to make mentoring cost-effective. This left 73 of the 93 treated markets. We randomly assigned 37 of these (within the original randomization strata) to mentoring, and the remaining 36 to be training only. Then within the mentoring markets, all those individuals assigned to training were assigned to be offered mentoring services, although in practice only those businesses still surviving were actually offered these services. We then compare the mentoring and training-only markets to the 44 markets in the control group with 16 or more firms.

2.3 Characteristics of the Sample and Verification of Random Assignment

Table 1 provides some key characteristics of the women selected in our sample and their firms by treatment assignment (Appendix Table 1 provides a similar table by assignment to mentoring). The sample looks similar across the three groups and we cannot reject the joint orthogonality of baseline characteristics when comparing the treatment group to the spillover group, the treatment group to the pure control, or the spillover group to the pure control. The average woman in our sample is 36 years old, has 9 years of schooling, and has been running her firm for just over 6 years. Two-thirds of the women are currently married. The modal firm has no employees (only 20 percent have one or more employees). The mean firm earns 1100 KSH (US\$13) per week in profits on sales of 5500 KSH (US\$65), and has capital stock of 31,000 KSH (US\$370).⁷ One quarter have ever received financing from a bank or microfinance organization and 45 percent are registered at the local level. Only 35 percent of firms keep business records at baseline, and on average firms are using just over half of the 26 business practices in the McKenzie and Woodruff (2015) index. This suggests scope for improvement from business training.

⁶ When the number of firms within a strata was odd, the additional unit was also randomly allocated to training.

⁷ The exchange rate was approximately 1US\$=84 KSH in 2013 at the time of baseline.

3. The Training and Mentoring Interventions

3.1 The Get Ahead Program

Vj g'tclpłpi 'r tqxkf gf 'ku'j g'KQau'I gpf gt'cpf 'Gpvtgr tgpwtuj k' 'Vqi gj gt'o *Get-Ahead* for Women k' "Gpvtgr tkug'r tqi tco 0'Vj ku'r tqi tco "of k'htu"htqo "eqpxgpvkpcn'dwukpguu"tclpłpi materials by highlighting essential entrepreneurial skills from a gender perspective... It addresses the practical and strategic needs of low-income women in enterprise by strengthening their basic business and people management skills. It shows women how to develop their personal entrepreneurial traits cpf 'qdvclp'lw r qtv'j tqwi j 'i tqw u.'pgy qtm'cpf 'kpu'kwk'qpu'f gcnłpi 'y kj 'gpvtgr tkug'f gxgrqr o gpvö" (Bauer et al., 2004). The program began in Thailand in 2001, and has now been used in at least 21 countries, serving approximately 400,000 women.

The program has four key modules, with the following themes:

Module 1: Basics on Gender and Entrepreneurship. The module introduces *Get-Ahead* and delivers basic concepts on the promotion of gender equality between men and women and the life cycle of people and enterprises.

Module 2: The Business Woman and Her Environment. The module focuses on raising awareness among women about their strengths and weaknesses as well as current or future working environment and its effect on the business.

Module 3: The Business Project. The module offers trainees information on the development of business ideas, opportunities and challenges, and the basics around marketing, production, services and technology, and financing, including costing and book keeping.

Module 4: People, Organization and Management. It is a soft skills module around managing a business (or a family business), with information on management of self and others, business support and networking, and action planning.

Some of the topics covered throughout these modules are not often emphasized in general business training programs, for example the conversations about the difference between sex and gender, and the role of cultural constraints in shaping women in business; dividing household and business tasks and time management around this; and how to network with other women and the role of y qo gpa'cuuqek'v'qpu'0k' 'cf f k'kp. 'k'eqxgtu'c'pwo dgt'qh'qr leu'o qtg'v' r k'cn'qh'ucpf ctf 'r tqi tco u" such as recordkeeping and bookkeeping; separating business and household finances; marketing;

financial concepts; costing and pricing; generating and fine-tuning new business ideas; setting smart objectives; and traits needed for business success.

The program seeks to provide training to low-income women engaged in small-scale businesses. The training methodology is participatory, with practical exercises to teach concepts. Modules expose women to different situations and environments simulating real business life. For example, women learn about the different types of costs involved in production, and how to account for their own costs through making lemonade. They have role play exercises to practice different sales strategies for customers, and make necklaces to discuss a production process and the importance of different factors in product design.

The course is a five-day course, with Appendix 3 showing the daily schedule. All trainers had at least five years of experience in training small firms, and had tertiary qualifications. Training took place in two to three locations per county. The locations were chosen to be relatively central to clusters of marketplaces, and were typically held in local hotels or church buildings. Training was offered for free, and participants were provided transport subsidies of approximately US\$6 per day to cover the costs of travelling from their residences to these locations (an average of 14 kilometers). The cost of providing the training is estimated at between US\$222 and US\$333 per woman trained.⁸

3.2 Training Take-up

Training took place immediately after the baseline surveys in each county, between June and November 2013. Of the 1172 individuals assigned to training, 77.7 percent attended at least one day of training. Of the individuals who attended at least one day, 94.6% attended all five days. In Diwan et al. (2014) we report on a choice structure experiment intended to increase training attendance, and discuss the correlates of attendance. Age and marital status are strong and statistically significant predictors of attendance: all else equal, women aged above 35 are 35 percentage points more likely to attend training than those below 35, while married women are 24 percentage points less likely to attend than unmarried women. This potentially reflects the competing demands on their time from other household tasks. Women are also more likely to

⁸ The smaller number reflects workshops with 20 attendees, while the larger is for workshops with 30 attendees.

attend if they have previously participated in training (perhaps reflecting greater perceived benefits from attending), have a large household (potentially providing more people to undertake household and business tasks in their absence), and are located closer to the training venue (reducing travel time). Women who earn more profits are less likely to attend, perhaps reflecting a higher opportunity cost of time, or that they think there is less need to improve.

3.3 Mentoring Intervention

After seeing the one-year follow-up results and emerging evidence suggesting that many business training programs may be too short to show sizeable impacts (McKenzie and Woodruff, 2014, 2015), we decided to add a mentoring component for half the sample assigned to training. Mentors were female business owners of a similar average age to our study sample, with 75 percent having had post-secondary education, and 68 percent having studied business or accounting, and a median income of US\$1,000 per month. Mentors were recruited through the Women Enterprise Fund (WEF) and Kenya Industrial Estates (KIE).

Each mentor was assigned a group of three to six mentees. The mentor program then involved the mentor meeting in a group with the mentees for 10 sessions, comprising of one session every two weeks for five months followed a standardized structure (Appendix 5). In addition, the mentor would meet with the individual mentees once a month over this period to provide individualized guidance. The mentoring process aimed to reinforce the business training by having the business owner identify core goals for their business, consider where there are gaps, explore options to fill these gaps, and then take action towards meeting these goals.

There were 524 firms assigned to training in the mentoring markets. In practice, only the 446 women whose businesses were surviving at the time of the first follow-up were contacted about mentoring. Out of these 392 signed up for the program, and 388 attended at least one session. Conditional on attending, the median went to 77 percent of all sessions. Take-up is thus 388/524 (74 percent) of those assigned, and 388/446 (87 percent) of those in business at the time of the intervention. The cost of providing the mentoring is estimated at US\$553 per women trained⁹.

⁹ The estimate is based on the number of women that attended at least one session.

3.4 How might training and mentoring affect treated and spillover firms?

The aim of these programs is to help those participating to grow their businesses. Consider a production function $f(\cdot)$, chooses inputs of labor L and capital K to maximize profits:

$$\pi = pf(\theta, K, L) - wL - rK \quad \text{s.t. } K \leq \tau A \quad (1)$$

Where p is the market price of their product, w and r are the prevailing market input prices for labor and capital, respectively. τA is the firm's wealth level A .

Business training and mentoring can then potentially increase profit levels through several channels. Increasing entrepreneurial ability may allow the business owner to produce more with the same inputs, giving the business more to sell. Better record-keeping and financial accounts may make it easier for the owner to obtain financing, alleviating credit constraints and again allowing the business to produce more. If markets are perfectly competitive, then firm owners will be able to sell all they produce at the market price p , and there will be no spillover impact on other firms in the market.

However, if firms are competing with one another for customers, then the price p need not remain constant as firms produce more. Better marketing practices, better customer service, and more efficient cost control may enable trained firms to compete away customers from other firms in the market. As a consequence, any gains to treated firms could come at the expense of lower sales and lower profits for non-treated firms in these markets.

Conversely, one of the hopes of the training providers is the possibility of positive spillovers to other firms in the market. A first potential channel is that the trained business owners might teach others in the market the skills learned, or that others in the market may be able to observe and adopt the new practices irrespective of whether or not they are directly taught. Second, the Get Ahead training emphasizes the role of networks and women working together to seek financing or purchase products together to obtain bulk discounts, which could result in lower costs and higher profits for both treated and untreated. Thirdly, there may be multiplier effects in which trained women spend their additional income on items sold by other businesses in the market.

Implicit in this discussion is the assumption that the production function $f(\cdot)$ is itself unchanged. But business training may encourage firm owners to consider producing new products or offering new services. This offers another means for treated firms to grow without negatively impacting on untreated firms ó they can expand the market through increasing variety.

4. Follow-up Surveys and Estimation Approach

4.1 Follow-up Surveys

Four rounds of follow-up surveys were conducted, in order to measure outcomes approximately one year and three years after training occurred (see timeline Appendix 1). Two types of surveys were used. A comprehensive long-form survey collecting data on a wide range of business outcomes was used in rounds 2 and 4. These were supplemented by much shorter surveys in rounds 3 and 5. These short surveys were conducted two or three months after the long surveys, and were intended to provide a second observation on volatile business outcomes like sales and profits, as well as an additional opportunity to gather data from individuals who could not be found at the time of the long survey rounds. Appendix 6 describes how key outcomes are measured.

Appendix Table 2 details response rates. Overall we were able to interview 95.0 percent of the sample in at least one of round 2 or 3, and 92.3 percent in at least one of round 4 or 5. In addition, in cases where we were unable to interview someone due to refusal, travel, death, or other reasons, we collected information from other household members or close contacts on whether the individual in our sample was currently operating a business. This enables us to have data on survival status for 99.3 percent of the sample at one year, and 97.2 percent at three years. There is no significant difference in data availability with treatment status at the three year horizon, although those assigned to treatment are 1 to 2 percentage points more likely to have data available at the one year horizon.

These four rounds of follow-up surveys enable us to measure the direct impacts of training, as well as spillover impacts on women in the experimental sample operating in the same markets. In order to test for broader spillovers, we collected three additional types of surveys. The first was market censuses of all women operating in these markets at the time of the round 2 and round 4 follow-

up survey. This enables us to capture new entrants to the markets, as well as women who ran existing firms but who did not pass our eligibility filters. Secondly, between June and August 2017, approximately four years post-intervention, we returned to these markets and conducted a market census of all women from our original sample that were still operating, as well as all male business owners operating in the same markets. This allows measurement of any spillover impacts on male business owners. Finally, at the same time as conducting this 2017 market census, we conducted a customer survey of 3,073 customers, randomly selecting 20 customers from each market to ask a very short survey about their frequency of shopping, share of food expenditure made at that market, and whether they shop at other markets. We use this to help assess potential concerns about spillovers on firms in other markets.

In addition to the survey data we have two other sources of information that aid in assessing impact. The first are photos of the inventories of the businesses, which were taken at the time of the baseline survey and in rounds 2 and 4. We had two independent field staff (blinded to treatment status) value these inventories based on the market prices of the different items, and average these values to get a photo-based measure of the size of the firm. They also visually compared the firms in photos to determine whether the business looked bigger in relative terms to what it was at baseline. Secondly, intensive qualitative work for impacts over the first year of the study was carried out by ICRW (2015).

4.2 Estimation Approach

A pre-analysis plan and the associated trial were registered on the AEA Social Science Registry on February 21, 2014. The assigned registry number is AEARCTR-0000287.¹⁰ This plan pre-specified the primary and secondary outcomes of the study, the estimation approach, and the causal chain to be investigated prior to the collection of any follow-up data. Subsequent funding enabled us to conduct the short follow-up surveys to collect additional data on profits and sales immediately following the long-form surveys, and to add the mentoring intervention. The market censuses conducted in 2017 were not included as part of this original plan, but conducted in response to feedback on the initial draft of this paper.

¹⁰ <http://www.socialscienceregistry.org/trials/287>

Following McKenzie (2012) we pool the short and long-run follow-up data to get average effects at one and three years, and use an ANCOVA specification where the baseline data are available in order to maximize power. We then examine outcomes at the individual firm level by estimating for firm i in market j at time $t=2,\dots,5$:

$$Y_{i,j,t} = \beta_0 + \beta_1 T_{i,j} * 1Year_t + \beta_2 T_{i,j} * 3Years_t + \beta_3 S_{i,j} * 1Year_t + \beta_4 S_{i,j} * 3Years_t + \pi Y_{i,j,t=1} + \gamma BM_{i,j,t=1} + X'_{k,i,j} \theta + \sum_{s=2}^5 \delta_s 1(s = t) + \varepsilon_{i,j,t} \quad (2)$$

Where $Y_{i,j,t}$ is the given outcome variable measured in round t , $Y_{i,j,t=1}$ is its baseline value and $BM_{i,j,t=1}$ a dummy variable indicating whether or not this baseline value is missing, $T_{i,j}$ is an indicator for being in a treatment market and being assigned to treatment, $S_{i,j}$ is the spillover term, measuring whether firm i is a control firm in a market assigned to treatment; $1Year_t$ is a dummy taking the value one in follow-up rounds 2 and 3; $3Years_t$ is a dummy taking the value one in follow-up rounds 4 and 5; we follow Bruhn and McKenzie (2009) in using X_k as a vector of randomization strata dummy variables (geographic region*market size*profit range), δ_s are survey round dummies, and $\varepsilon_{i,j,t}$ is the error term, which we cluster at the market level to account for the market level random assignment. β_1 and β_2 provides the intent-to-treat effects at one- and three-year horizons, which is the effect of being assigned to treatment relative to being a firm in the control markets. Note that the three-year impact represents the impact of having been assigned to receive business training three years earlier, followed by having a fifty percent chance of being assigned to receive mentoring approximately 9 months earlier. We test $\beta_1 = \beta_2$ to determine whether the impacts differ over time, and $\beta_1 = \beta_2 = 0$ to determine whether we can reject that there is no treatment effect in any period.

One of the main questions of interest in this study is also whether there are spillovers from the treated firms to other firms within the same villages. These spillovers could be positive (e.g. treated women share knowledge with control women in the same markets) or negative (e.g. treated women compete away the sales of control women from the same markets). β_3 and β_4 measure these spillover effects at the one and three year horizons by comparing control firms in treated markets to control firms in control markets. Section 5.3 then discusses estimation of spillover effects on other firms outside of our experimental sample.

We then test whether being assigned to the mentoring treatment had a differential impact from training alone via the following regression, estimated using only rounds 4 and 5:

$$Y_{i,j,t} = \gamma_0 + \gamma_1 \text{Mentor}_{i,j} + \gamma_2 \text{Trainonly}_{i,j} + \gamma_3 \text{Spillovermentor}_{i,j} + \gamma_4 \text{SpilloverTrainonly}_{i,j} + \pi Y_{i,j,t=1} + \gamma \text{BM}_{i,j,t=1} + X'_{k,i,j} \theta + \delta_4 1(t = 4) + \varepsilon_{i,j,t} \quad (3)$$

Where *Mentor* indicates being in a market assigned to mentorship and being assigned a mentor, *Trainonly* indicates being in a market assigned to training only and being assigned training, and *Spillovermentor* and *SpilloverTrainonly* are the spillover groups in the mentor and training only markets. We estimate this equation only for markets eligible for the mentoring intervention (those with more than 15 firms in them at baseline). The standard errors are again clustered at the market level. We then test $\gamma_1 = \gamma_2$ in order to determine whether mentoring has a different effect from training alone.

We focus on the intention-to-treat effects in our main analysis. In Appendix 8 we also report the local average treatment effects (LATE) of receiving training on our primary outcomes, by instrumenting training attendance with training assignment in equation (2). None of the control group attended training, so the LATE is the same as the average treatment effect on the treated. The LATE impacts are approximately 24 percent higher than the ITTs.

5. Impacts on Primary and Secondary Outcomes

We begin by examining impacts on our primary and secondary outcomes at the individual level, then examine impacts on the markets as a whole, including discussion of whether there were spillover impacts on women outside of our experimental sample, or on male-owned businesses.

5.1 Impacts on Primary Outcomes at the Individual Level

Our primary outcomes are whether the business training succeeds in helping the firm survive, and increases firm sales, profits, and size (as measured by the photo value of inventories). We consider several measures of these variables, as specified in the pre-analysis plan, along with an aggregate index which is the average of standardized z-scores of these outcomes and provides an overall measure of whether the intervention has succeeded in increasing firm performance as well as a way for controlling for multiple hypothesis testing.

Table 2 provides the results. Panel A shows the one and three year impacts of assignment to training, while panel B separates the three year results by whether the firm is in a market also assigned to mentoring or to training only. Over the one-year horizon the direct impacts are all positive, but only the impact on daily sales is statistically significant at the 5 percent level, and there is a 0.05 standard deviation increase in the aggregate index, significant at the 10 percent level. These estimates are all larger in magnitude by year three, with statistically significant impacts on survival (3 percentage points), daily sales (171 KSH, or 14.6% of the control mean), weekly sales (1038 KSH, or 18.0% of the control mean), weekly sales of the main product the business sells (454 KSH, or 13.4% of the control mean), and weekly profits (221 KSH, or 15.4% of the control mean). The overall aggregate index shows a 0.088 standard deviation increase, which is significant at the 1 percent level. The year three results are statistically different from the year one results for the key outcomes of survival, weekly sales, and weekly profits, and we can reject the null hypothesis of no treatment effect in either period.

The impact on our objective and independent measure of business size: the value of inventories as photographed in the firm is positive (818 KSH, or 9.5% of the control mean), but not statistically significant. Appendix 9 considers a second measure from the photographs, which comes from comparing the round 4 business to the baseline business and classifying it as bigger, smaller, or the same size. We do find treated businesses are 4.2 percentage points more likely to be classified as bigger, which is significant at the 5 percent level. Appendix 9 also considers two additional sets of robustness checks. Our main analysis codes profits and sales as zero for firms which have closed down (since they are not earning profits or sales). Appendix 9 shows that firms assigned to training have significantly higher weekly profits, weekly sales, and a higher aggregate index even when we condition on survival. Appendix 9 also shows the results remain significant after using an alternative definition of firm survival, and when using alternative transforms of profits and sales such as the inverse hyperbolic sine and logarithmic transformations. Finally, it also addresses the potential concern that business training affects how profits and sales are reporting by showing no treatment effect on the number of reporting errors, nor on the difference between the reported inventories and the size based on valuing photographs of inventories.

Figures 2 and 3 show the quantile treatment effects on profits and sales for the three-year time horizon. The quantile effects increase with quantile, showing that the impacts were larger at the top of the distribution than the bottom, but are statistically significant from the 20th percentile onwards until at least the 80th percentile.

Consider next the evidence for spillover effects. Over a one-year horizon we see negative spillovers on weekly profits and weekly sales, with this effect significant at the 10 percent level for sales. However, we also see positive and insignificant spillover estimates for several of our other profits and sales measures and, as a result, the impact on the aggregate index is small, with the point estimate being a 0.01 standard deviation reduction, which is not statistically significant. Over the three-year horizon, none of the outcomes show a significant spillover impact, and the aggregate index has a 0.002 standard deviation increase, which is small and not statistically significant. Likewise, the quantile treatment effects in Figures 2 and 3 show the spillover impacts to be much smaller than the direct effects, and not statistically significant. As a result, we cannot reject that there are no spillover effects (either positive or negative) on untreated firms operating in the same markets as the treated.

This picture of either small or no spillovers in sales and profits is consistent with the qualitative work, with some respondents saying they saw no change in the market competition as a result of some women being trained, while others noting a mix of positive (passing on knowledge to other firms) and negative (treated women providing better customer service to attract customers) spillovers.

Panel B examines whether the impacts differ for those assigned to mentoring versus training alone. The impacts are similar in magnitude for most outcomes, with the one exception being survival, where training alone appears to have a larger impact than training followed by mentoring. When we consider our overall index, which accounts for multiple testing, we find a 0.098 standard deviation increase for mentoring and 0.104 standard deviation increase for training alone. These impacts are very similar in magnitude, and we cannot reject equality ($p=0.901$).

5.2 Impacts on Secondary Outcomes at the Individual Level

To examine the extent to which these improvements in profits and sales translate into broader individual wellbeing, our pre-analysis plan specified four domains of secondary outcomes to examine. The impacts on these outcomes are shown in Table 3.

The first set of secondary outcomes concerns employment, whether in self-employment or wage work. Column 1 of Table 3 shows that 88.5 percent of the pure control group are engaged in some form of employment for pay at the time of the three-year follow-up surveys, and the increase in firm survival also translates into an increase in any employment. Column 2 looks at all income from work, which combines profits from self-employment with any earnings from wage labor. We see the higher profits are not being offset by lower wage earnings, as total income from work increases by 389 KSH per week, or 18.1% of the control mean.

The training emphasized a number of topics and approaches that had the goal of empowering women in terms of decision-making around finances and business. We measure 10 different outcomes in this domain (e.g. are they compelled to spend money on their husband or family, do sole control over, etc.). The average individual in the control group is able to do 7 out of 10 of these decisions, and Column 3 shows that training is not found to have any sizeable or significant impact on this measure of empowerment at either time horizon. This is consistent with the qualitative assessment, which noted that training did not appear to change individual or household decision-making dynamics (ICRW, 2015).

The third domain we examine is subjective well-being and mental health. We measure subjective well-being today and anticipated subjective standard of living in 5 years on a Cantril ladder, and mental health using the MHI-5 index of Veit and Ware (scored so that higher scores indicate better mental health). Respondents show a great deal of optimism about the future, seeing themselves as being on step 5.0 out of 10 on the life ladder currently, but expecting to be on step 8.0 in 5 years. We find that training increases subjective well-being by 0.2 to 0.3 steps. The impact on mental health is positive and also statistically significant in both time periods. Finally, we examine the impact on ownership of household durable assets. There is a

positive, but marginally significant impact in year one, and positive but insignificant impact in year three. The only evidence for a significant difference between the mentoring and training alone groups comes for the any employment outcome, which is consistent with the differential survival effect. Taken together, these results show the higher profits and sales do appear to have translated into higher overall well-being for the women given training.

5.3 Market Level Impacts and Potential Spillovers on Other Types of Firms

Our results show that firms assigned to treatment have grown, with no significant negative spillover for the untrained women in our experimental sample in the same markets. This suggests that the overall markets have grown. To formally test this hypothesis, we aggregate outcomes to the market-level, and estimate the following equation at the level of the 157 markets:

$$TotalY_{j,t} = \beta_0 + \beta_1 MarketTrained_j * 1Year_t + \beta_2 MarketTrained * 3Years_t + \pi TotalY_{j,t=1} + X_j' \theta + \sum_{s=2}^5 \delta_s 1(s = t) + \varepsilon_{j,t} \quad (4)$$

Where *MarketTrained* indicates market *j* was assigned to training, *TotalY* is the total of outcome *Y* (profits, sales, or customers) in market *j* at time *t*, and the standard errors are clustered over time at the market level.

Table 4 shows the results of estimating equation (4). By the third year, overall profits have grown 2554 KSH per week (9.2% of the control mean, not statistically significant), overall sales in the market by 15353 KSH per week (13.7% of the control mean, significant at the 5 percent level), and the total number of customer transactions¹¹ in the market per week by 318 (16.7% of the control mean). These results are consistent with the overall market growing, at least over the longer-term.

Our study was designed to measure spillovers on similar women to the training firms operating in the same markets. Since these are women operating firms of similar size and in similar industries to the treated firms, our prior is that any spillovers would be largest for this group. The analysis in the first three columns of Table 4 therefore considers the ðmarketö to be the set of women in our experimental sample at baseline. However, because we do not find significant spillovers on these

¹¹ Each firm is asked how many customers they had in the past week. The total number of customer transactions aggregates these responses across all firms. Thus, if a customer shops at two firms, it is counted as two customer transactions here.

women, it is natural to wonder if spillovers are instead taking place for other groups. Our market census data and customer survey data allow us to examine this.

The rounds 2 and 4 market censuses enable us to examine spillovers on two other potential groups of women who might be affected. The first is female new entrants to these markets. The second is existing firms operated by women in these markets that did not pass our eligibility filters. Columns 3 through 6 of Table 4 show positive and insignificant impacts on the numbers of new entrants in these markets, the total number of female-operated firms not in our survey, and the weekly profits of these firms. The success of those trained therefore also does not come from crowding out new entrants or other women in these same markets.

Although most of the firms operated by women in our sample are in sectors that are female-dominated (section 2.1), there remains the possibility that trained women took customers from male-operated firms in these sectors, or even that customers substituted spending on male-dominated sectors to buy more from the trained women. Our 2017 market census was designed to investigate this possibility by conducting a short census of both our original firms and of male-owned firms operating in our study markets. Column 1 of Table 5 shows that 83.6 percent of the pure control group were able to be re-interviewed in this exercise, with this not differing by treatment status. In columns 2 and 5 we then estimate the impacts on log sales and log profits for operating firms, measuring the direct impact for women assigned to training, the spillover impact on women in the experimental sample not invited to training, and the spillover impact on men operating in the same markets. We see that the treatment group still has 14 percent higher sales and 15 percent higher sales (conditional on survival and being re-interviewed) four years after treatment, which is statistically significant at the 5 percent level.¹² In contrast, the spillover impacts for untrained women and for untrained men operating in the same markets are positive, smaller in magnitude, and not statistically significant.

Columns 3 and 6 of Table 5 split the male sample into males working in male-dominated sectors (butchery, artisans, carpentry, electronics, hardware, shoe repair, cyber cafes and computers) from

¹² It continues to be the case that we cannot reject equality of treatment effects for the mentoring and training only treatments ($p=0.316$ for sales, $p=0.884$ for profits).

males working in the same sectors as most of the women in our study. The spillover effect is not statistically significant for either, and we cannot reject that the spillover impacts are the same for both groups ($p=0.809$ for sales, $p=0.166$ for profits). The result is that in columns 4 and 7, when we look at average sales and profits of all these firms in the market, they have significantly increased. This again shows the overall market has grown.

A final type of spillover that might then occur could be if customers switch from shopping at other markets to buying more from their local market. The average distance within county from one study market to another is 76km, and the average distance to the closest market is 5.9km. Our customer survey taken in 2017 asked a random sample of 20 customers per market how many days a week they shop at the study market. The mean is 4.9 days, and median 6, with no significant difference by market treatment status ($p=0.394$). They spend a mean of 71 percent and median of 80 percent of their total food expenditure in the study market. Appendix 10 shows our main results are robust to excluding markets where customers do not spend at least 70 percent of their food expenditure in the market, and to excluding markets which are within 5km of another market. This suggests our results are not being driven by spillover effects occurring in other nearby markets.

6. Why do these impacts arise? Causal Chain and Mechanisms

Our surveys and pre-analysis plan enable us to trace out the causal chain from providing training through to changes in business outcomes, and to examine the different mechanisms through which training may or may not have an effect. Since we examine a number of different mechanisms, in Appendix 11 we provide sharpened q-values that hold constant the false discovery rate when reporting results for specific outcomes. All outcomes at the three-year horizon that have p-values below 0.05 also have sharpened q-values below 0.05, whereas only the one year impacts that are significant at the 1 percent level have sharpened q-values below 0.10.

6.1. Changes in Business Knowledge, Practices, and Use of a Mentor

The first step in the causal chain is for training to lead to changes in the business knowledge and business practices of the women taking training. Business knowledge was assessed through giving respondents a description of a business and then asking them seven questions that involve calculating the revenue, value of stock on hand, variable costs, total expenses, profits, fixed costs,

and break-even point. This proved very difficult for most participants, with the median respondent only getting 2 out of 7 questions right, and only 0.5 percent getting all the answers correct. This question was only asked in the first long follow-up survey and was asked of both those with surviving businesses as well as those whose business had closed down. Column 1 of Table 6 shows that there is no significant treatment effect or spillover effect on business knowledge. This is consistent with the financial literacy results of Carpena et al. (2017) who find that financial literacy training does not improve performance on questions involving numerical calculations.

We measure business practices through a set of 26 questions that measure the marketing, record-keeping, buying and stock control, and financial planning of the firm. These questions are only measured in the long follow-up survey rounds and only for firms that survive. These questions have been shown to correlate strongly with business performance in a range of countries by McKenzie and Woodruff (2015), and to predict future survival and growth of the firm. The mean firm in the pure control group is employing 53 percent of these practices. Column 2 shows that the impact of being invited to training is a statistically significant increase in the use of business practices at both the 1 and 3 year horizons: treated firms have a 0.05 to 0.07 increase in the proportion of practices used. This is approximately a 10-13 percent increase on the control mean. There is a marginally significant positive spillover in business practices to untreated firms in the same marketplaces over 1 year, however the magnitude is very small (0.01), and the three-year impact is smaller still and not significant. The increase in practices is three times as large for those assigned to mentoring (0.096), compared to those assigned to training alone (0.033), suggesting the mentoring did help reinforce business practices.

Finally, column 3 confirms that those assigned to the mentoring treatment are indeed much more likely to say they have used a mentor. However, we also see a significant impact on using a mentor from those firms assigned only to training, suggesting that training did lead to a doubling in the use of mentor, even without the separate mentorship intervention.

6.2 Impact on Dealing with Customers

Increased marketing and better presentation of the business may enable the firm to increase sales through gaining more customers and being better able to retain existing customers. We examine

this in the first three columns of Table 7. The women in our sample serve a median of 60 and mean of 100 customers per week in the control markets (after top-coding at the 99th percentile to reduce the influence of outliers). Column 1 shows that those assigned to training are serving an additional 14 customers a week after three years, significant at the 1 percent level. Columns 2 and 3 show treated firms are more likely to have gained a new customer during the last 3 months, and less likely to have lost a regular customer during the same time period. There is no significant spillover impact, which is consistent with the previous evidence that the total number of customers in the market increased.

Total sales of the business increased, as did the number of customers. Column 4 considers the sales per customer. The point estimates are positive, suggesting more revenue per customer, but not statistically significant. If we instead consider outcomes in log form, so that we have the identity:

$$\ln(\text{sales}) = \ln(\text{sales}/\text{customers}) + \ln(\text{customers}) \quad (5)$$

Then the treatment effects are 17.6 percent on sales ($p=0.006$), 9.0 percent on sales per customer ($p=0.107$) and 8.5 percent on the number of customers, suggesting that the increase in sales volume is coming approximately evenly from expanding the number of customers as well as from obtaining more revenue per customer.

Why might business owners be able to get more customers? The qualitative work in part points to a better customer experience, with shops being cleaner, more likely to open on time, and owners being nicer to the customers. For example, the qualitative work noted one woman saying *“Vj gtg” ctg”vj qug”y j q”f kf pø”mpqy ”j qy ”v”cwtcev”ewuqo gtu.”dw”pqy ”Kecp”ugg”vj g{”j cxg”dggp”cdng”vq” attrcev”ewuqo gtu...Vj g{”ctg”cmkpi ”vq”vj go ”plegn{”wprkng”dghqtg”y j gtg”vj g{”y qwf ”cm’twf gn{”o” and another noting that now ðI try to talk [with] customers in a nice and professional way. I try to make delicious meals and make my place of work look attractive to vj go ...Ko cng”uwtg”}o {” eqqmkpi _”k”u”w ”vq”vj g”ewuqo gtuø”ucpf ctf.”pqv”y cvgt{”hqqf ø”(ICRW, 2015). In qualitative kpgtxlgy u”y kj ”vj qug”y j q”j cf”i qpg”vj tqwi j ”o gpvtkpi.”qpg”r ctvek cpv”uckf ”ðKwugf ”vq”pqv”ectg” much about my business, I could open late, and sqo gvko gu”vj g”lj qr ”y qwf ”tgo clkp”enqugf ”kh”Kf qpø” hggnrkng”qr gpkpi OP qy ”Kmpqy ”vj cvk”k”u”ko r qtcvp”vq”r w’ghqtv”kpvq”o {”dwukpguuø0*

We incorporated this potential channel into the three-year follow-up survey, asking whether the business always opened at a set time each day. Column 5 shows only 34 percent of the control group do this, and training increases this by a significant 6 percentage points. Mentoring appears to have a greater impact than training alone on having regular opening hours. Appendix 12 shows that the route to more customers does not appear to be through price competition, with no significant impact of treatment on the prices charged (or input costs of) the most commonly sold goods.

A second reason firms were able to attract more customers was by diversifying the range of products they sold. Banerjee and Duflo (2008) note the preponderance of small businesses in developing countries all selling similar things to one another, with no reason for customers to seek them out in terms of either product line or shopping environment. Column 6 shows that firms assigned to training were 9 to 11 percentage points more likely to have introduced a new product to their business, significant at the 1 percent levels.¹³ There was tremendous variety in the new products introduced, with almost every business that said it had introduced a new product giving a different answer for what product it had introduced. Examples included avocados, oranges, carrots, garlic, fertilizer, body oil, cowpeas, belts, hair oil, tea, tobacco and Weetabix. Since most firms were selling only a handful of products, this diversification could have attracted customers that would not have otherwise purchased from them.

A final reason firms may have more customers is through general equilibrium multiplier effects, whereby the extra earnings of business owners are spent on products offered by other firms in the market. The 2017 market census asked firms what share of the customers were also business owners in the same market, with five possible ranges as the answer. The median firm says between 26 and 50 percent of its customers are also business owners in the same market. Table 8 estimates the treatment impact on this response. Column 1 uses the range response, where higher numbers indicate a larger percentage, while column 2 discretizes the response as the midpoint of each range. We see that the treatment impact is positive and significant for firms assigned to training.

¹³ Our survey asked whether firms had introduced a new product or service, and what the main product or service introduced was. We are therefore unable to measure what share of sales come from this new product, or whether there are other firms in the market also selling this product.

Moreover, the spillover effects are also all positive, and are statistically significant for males operating firms in male-dominated sectors. This suggests that part of the way the market has grown has been through successful firms shopping with one another, and with other firms in the market.

6.3 Impact on Financial Control, Access to Finance, and Capital Levels

We next examine the extent to which training enables firm owners to produce more or earn higher profits by managing resources more effectively, and overcoming financial constraints. Several studies have emphasized the possibility that business training may have its strongest impact on sales during a bad month by helping participants identify strategies to reduce downward fluctuations in sales through diversifying the products they offer, as well as being more proactive about alternative activities during slow months. McKenzie and Woodruff (2014) note however that the evidence for this has been mixed in existing studies. We examine this channel in columns 1 and 2 of Table 9. We see that although businesses are more likely to regularly use business records to know if sales of a particular product are increasing or decreasing, this does not translate into any impact on reducing fluctuations in profits in bad months relative to usual months.

Columns 3, 4, and 5 of Table 9 examine the inventory management and purchasing outcomes of the firms. We see no significant impacts of treatment on the fraction of stock lost to spoilage, the

The latter is measured with considerable noise, so the insignificant point estimate admits a wide confidence interval.

Better business practices may enable firms to obtain more financing. This may work through either the demand or supply side for formal finance. On the demand side, individuals who have gone through training may have more plans for expansion or feel more confident approaching banks. On the supply side, banks may be more willing to lend to firms which keep better records, or which have higher profitability and sales. However, column 6 of Table 9 shows that access to finance is limited in our sample, with only 17 percent of the pure control group receiving a loan from a bank or microfinance organization in the past year, and that treatment does not have a significant impact on this. This is consistent with the qualitative work, which found a number of challenges on both the demand and supply side for formal finance: women were often afraid of what might happen if

they fell behind on payments, found the loan application process cumbersome and time-consuming, and also faced challenges in terms of ability to provide documentation and/or collateral in some cases (ICRW, 2015).

We do find firms to be 5.8 percentage points more likely to be using business bank accounts (column 7). Columns 8 and 9 then look at whether firms have grown in size according to their inventory levels and capital stock. By year three both the inventory levels and capital stocks are higher in the treatment group, which is consistent with them having more to sell, and with building up the size of the firm. The inventory impact is larger after mentoring than after training alone. Since firms did not receive more finance, this build-up in inventories and capital is likely to reflect re-investment of profits. This may also explain why impacts are stronger after three years than one year, since it takes time for firm owners to slowly reinvest profits into more merchandise to sell.

6.4 Impacts on Owner Hours, Attitudes, and Social Capital

The final set of intermediate outcomes and mechanisms are examined in Table 10. We start by examining whether women change the amount of time they are devoting to their business. The qualitative work suggested this might be the case, as evidenced by this quote from one participant

“Then Kwugf "q"qr gp"]o { "dwulpguu_cp { "ko g"Ky kuj gf ...Ky qwf "qr gp"o wej "rcvgt...dw'vj gug'f c { u"kwu'dgwgt'ulpeg'Kcpucpvn("qr gp"cv'plpg"cpf "emug"cv'plj j v'cv'ctqwpf "gki j v'Vj qug'f c { u"Klwu'wugf " to do a little work, I could not stay for long. Whenever I got some money to pay for my merry-go-round, then I would just close my business for the day. Also, then if I got someone who was buying five bags of maize, then I could just close business and leave for home, but these days I stay until Kco "eqpxlpegf "j cv'kwu" time to leave.ö"KTY ."4237-0Vj ku's wrcv'kg'hp'kpi does not show up in the one year horizon, but is there over three years for both mentoring and training alone, with women working 4 hours more per week.¹⁴

¹⁴ Part of this reflects the extensive margin of firms more likely to be operating, and hence their owners not just working zero hours. But even conditional on survival, firms assigned to training work a statistically significant 2.8 more hours per week after three years. We did not pre-specify looking at other employment. The control group averages 0.14 paid workers per firm. Assignment to treatment increases this by 0.041 workers (p=0.031) when we do not condition on survival, and 0.042 workers (p=0.056) conditional on survival. Thus while there is a significant employment impact, it is very small in magnitude.

The training was intended to also increase the confidence of women in their ability to perform business tasks. We measure entrepreneurial self-efficacy through 10 questions that measure the ability to develop new products, sell a product to a customer they are meeting for the first time, and persuade a bank to lend them money for their business. The mean control group individual is very confident in their ability to do 4.5 out of these 10 tasks, and Column 2 shows that we find no significant effect of training on self-efficacy. We also ask eleven questions intended to measure the types of attitudes that Get Ahead tries to encourage. These include willingness to take risks to get ahead in business, planning for the future, feeling confident that one can find solutions to problems that arise, and keeping your eyes open for ways to improve the business. These are scored on a 5 point scale, where 5 indicates the most agreement. The mean score across these questions is 3.8 for the pure control group. Column 3 shows no significant effect of treatment on these.

The training also emphasized and encouraging cooperating with other women in the marketplace. The remaining columns of Table 10 examine aspects of this. Columns 4, 5, and 6 consider working together with other women to obtain bulk discounts or to purchase goods together. We find no significant treatment effects on any of these measures over the one-year horizon, but do find treated individuals to be more likely to belong to associations and discuss business with others over the three year horizon. Despite the hope of the trainers that treated women would band together with untreated women to purchase inputs at discounts together, there is no direct effect nor spillover effect on this outcome.

7. Cost-Benefit

Our analysis shows that, over a three year horizon, the training has grown these underdeveloped markets, increasing the total volume of sales and number of customers. As a result, gains to the treated firms do not come at the expense of lost customers for the untreated firms in the same markets. The cost of training was approximately US\$200 per person assigned to training, and the gain in weekly profits was 221 KSH (\$2.60). The gain in profits would therefore need to last for at least 76 weeks, or approximately 1.5 years, for the benefits of the program to exceed the costs. Although we do not have a long time series to examine these benefits, we do see them being greater

at three years than one year, and continuing to be higher after four years, suggesting this appears plausible. The additional mentoring treatment cost approximately \$553 USD per individual assigned to mentoring. Since we are unable to reject that the impacts are the same for mentoring as for training alone for our primary outcomes, the evidence here suggests that adding mentoring does not pass a cost-benefit test.

8. Conclusions

We find the Get Ahead business training program does improve the survival, profitability, and growth of businesses which receive training. These impacts are much clearer over a three-year horizon than one year after training, persist into a fourth year, and suggest the scope for small businesses run by women to expand. Moreover, we find that this growth does not appear to come from taking sales away from competitor businesses, but instead arises from growing the overall market. Many markets in developing countries are underdeveloped, with poorly managed firms selling a narrow variety of products. The results of this paper show that training can be one way to develop these markets.

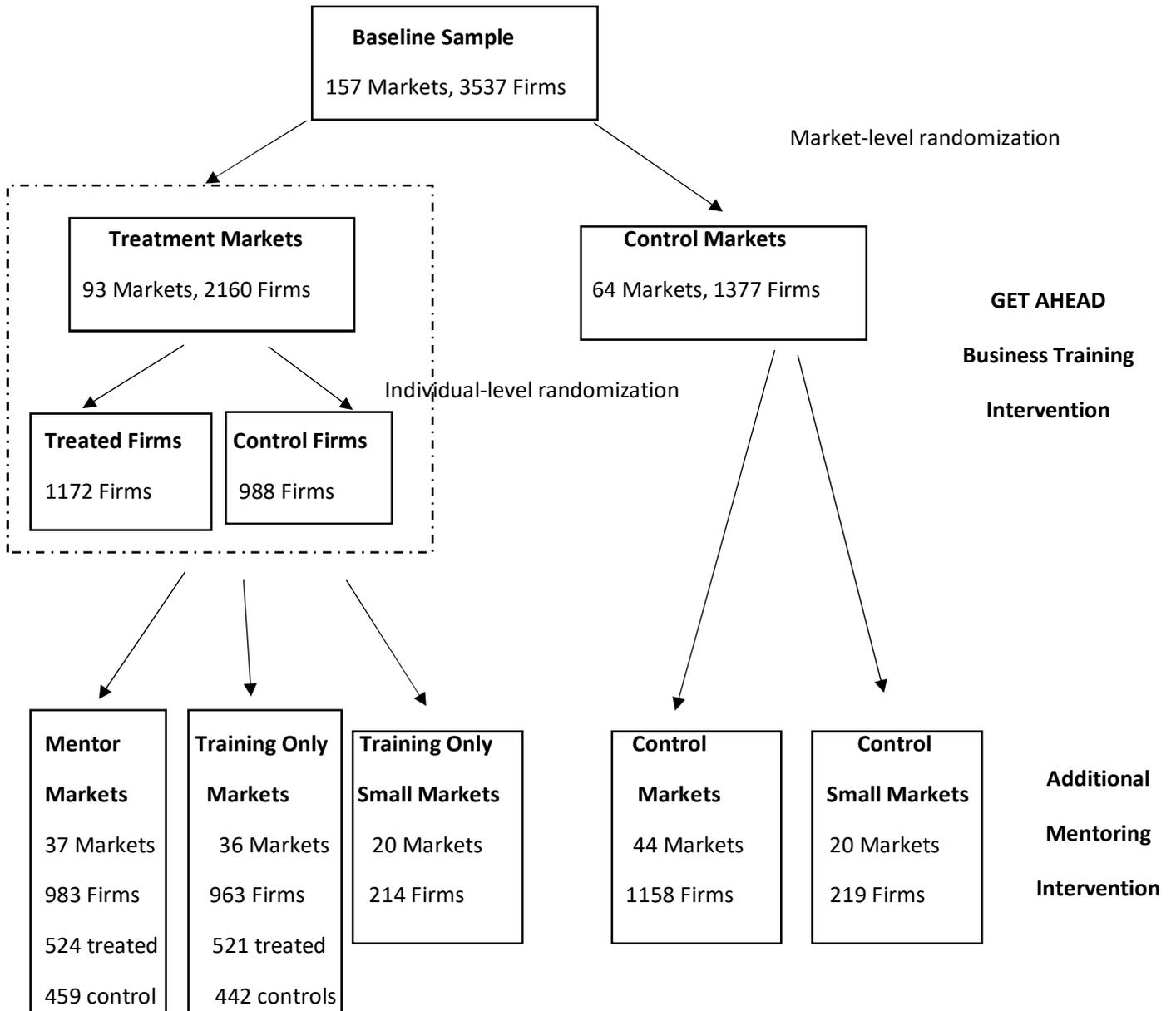
References

- Dcpgtlgg.'Cdj kkk'cpf 'Guj gt'F whm'422: +dY j cvku'O kf f rg'Ercuu'cdqw'vj g'O kf f rg'Ercuugu'Ctqw'pf " vj g'Y qtrf Aö." *Journal of Economic Perspectives* 22(2): 3-28.
- Bauer, Susanne, Gerry Finnegan and Nelien Haspels (2004) *Gender and Entrepreneurship Together: GET Ahead for Women in Enterprise: Training Package and Resource Kit*, Bangkok, Berlin and Geneva, International Labour Office.
- Berge, Lars, Kjetil Bjorvatn and Bertil Tungodden (2015) dJ wo cp"end Financial Capital for O letqgpvgr tkug'F gxgrqr o gpv'Gxkf gpeg'ltqo 'c'Hgrf 'cpf 'Ncd'Gzr gtko gpvö." *Management Science* 61(4): 707-22.
- Druqo . 'P leqruc.'Tchcgm'Ucf wp'cpf 'Lqj p'Xcp'Tggpgp'4237+'dO cpci go gpv'cu'c'Vej pqmji {ö." Mimeo. Stanford.
- Brooks, Wyatt, Kevin' F qpqxc' cpf " Vgtgpeg" Lqj puqp " 4238+' dMentors or Teachers? Microenterprise Training in Kenyaö.'O ko gq0
- Dtwj p.'O ktko 'cpf 'F cxf 'O eMgp| lg'422; +dK'Rwtuwk'qh'Dcrrpeg'ZTcpf qo k cvkq'kp'Rtceveg'kp" Dgxgrqr o gpv'Hgrf 'Gzr gtko gpvö." *American Economic Journal: Applied Economics*, 1(4): 200-232.
- Dwng.'Gty kp.'Tqdgvt'Ngpukm'Tqukg'xcp'Xgr gp'cpf 'Pj wpi 'Vj kJ qpi 'Xw'4238+'dF q'I gpf gt" and Business Trainings Affect Business Outcomes? Experimental Evidence from Xlqpcö." *Management Science*, forthcoming.
- Calderon, Gabriela, Jesse Cunha, and Giacomo de Giorgi (2013) dDwulpguu'Nkgtce{" end Development: Evidence from a Tcpf qo k gf "Vtkr'kp"Twcr'O gzleqö, *NBER Working Paper*
- Ectr gpc.'Hggm.'Uj cy p'Eqr.'Igtgo {"Uj cr ktq'cpf "Dkr'\ k"4233+'dThe ABCs of Financial Education: Experimental Evidence on Attitudes, Behavior, and Cognitive Biasesö." *Management Science*, forthcoming.
- Chandy, Rajesh, Stephen Anderson-O eF qpcr' cpf " Dkr'\ k"4238+' dRcy y c{u'v' Rtqhk' Kgp'kp' Ugr ctcv'Ej cpgnu'qh'Hko 'I tqy vj 'Vj tqvi j 'Dwulpguu'Vtclpki ö. *Management Science*, forthcoming.
- F g'O gn'Uwtgij . 'F cxf "O eMgp| lg"cpf "Ej tkqr j gt"Y qqf twh'4236+'dDwulpguu'Vtclpki "cpf" Female Enterprise Start-up, Growth, and Dynamics: Experimental evidence from Sri Ncpmö." *Journal of Development Economics*, 106: 199-210.
- F ky cp.'Hk cp.'I tceg'O cmepc.'F cxf 'O eMgp| lg.'cpf 'Ukr'k'Rctw| | qm'4236+'dKpxkcvkp'Ej qlg' Utwewtg'J cu'P q'K r cev'q'CWgpf cpeg'kp'c'Hgo cng'Dwulpguu'Vtclpki 'Rtqi tco 'kp'Mgp{cö" *PLOS ONE*, 9(10): e109873
- I kpé.'Zcxkt'cpf "I j c| cr' "O cpwtk'4238+'dMoney or Management? A Field Experiment on Constraints to Entrepreneurship in Rurcn'Rcnk'cpö.'O ko gq0Y qtrf 'Dcpn0
- ET Y "4237+'döKco 'f gvgto kpgf . 'pq'b cvgt'j qy 'j ctf 'k'dgeqo gu.'Ky kmpqv' kxg'wr ö'c's wrcv'kxg' cuuguo gpv'qh'vj g'KNQai'I GV'Cj gcf 'Dwulpguu'Vtclpki 'Rtqi tco ö.'ET Y "42370
- McKenzie, David (2012) "Beyond Baseline and Follow-up: The Case for More T in Experiments", *Journal of Development Economics* 99(2): 210-221.
- O eMgp| lg.'F cxf " cpf "Ej tkqr j gt"Y qqf twh'4237+' dDwulpguu'Rtceveg'kp" Uo cm'Hko u'kp" Developing Cqwp'kguö." *Management Science*, forthcoming.
- McKenzie, David and Christopher Woodruff (2014) dY j cv'ctg'y g'rgctpki 'ltqo 'dwulpguu'vclpki " gxcn'cvkqu'ctqw'pf 'vj g'f gxgrqr kpi 'y qtrf Aö." *World Bank Research Observer*, 29(1): 48-82.
- Tqvgö dgti . 'O ctvp'4236+'dGs wkr'dtkwo 'Ghgw'qh'Hko 'Uwdul'kguö.'O ko gq0

Xcrf klc."O ctvkp"4237+"öDwulpguu"vtcklpi "r nu"lqt"lgo crg"gpvtgr tpgwtuj k AUj qtv'cpf "o gf kwo " vgtö "gzi gtlö gpvri'gxlk gpeg"ltqo "Rgtwö."*Journal of Development Economics* 113:33-51.

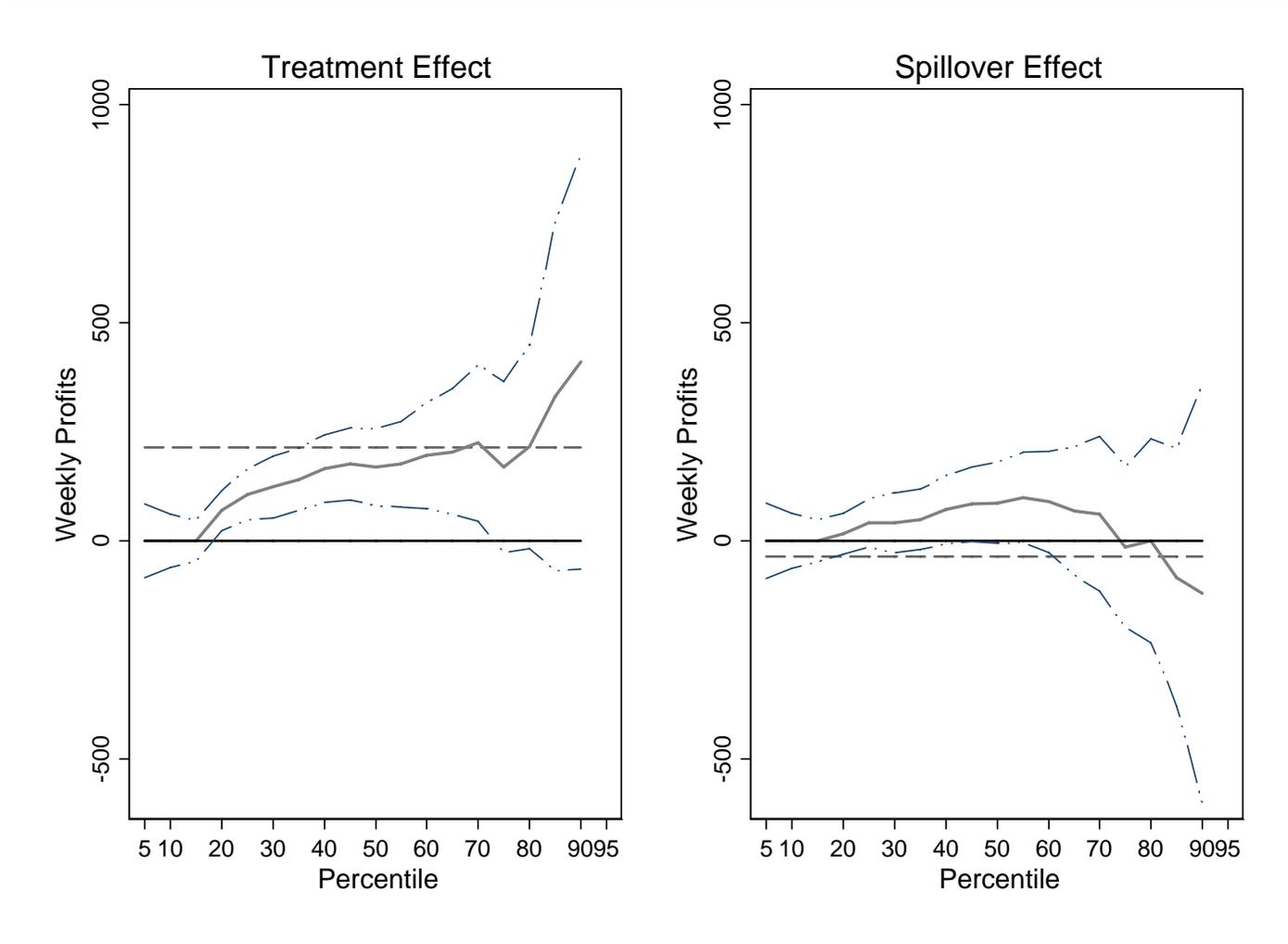
Veit, C.T. and Ware, Jnr, J.E. (1983). öVj g"utwewtg"qh'r u{ej qmji ical distress and well-being in i gpgtcnr qr wrcvkuö."*Journal of Consulting and Clinical Psychology*, 51, 730-742

Figure 1: Randomization Design



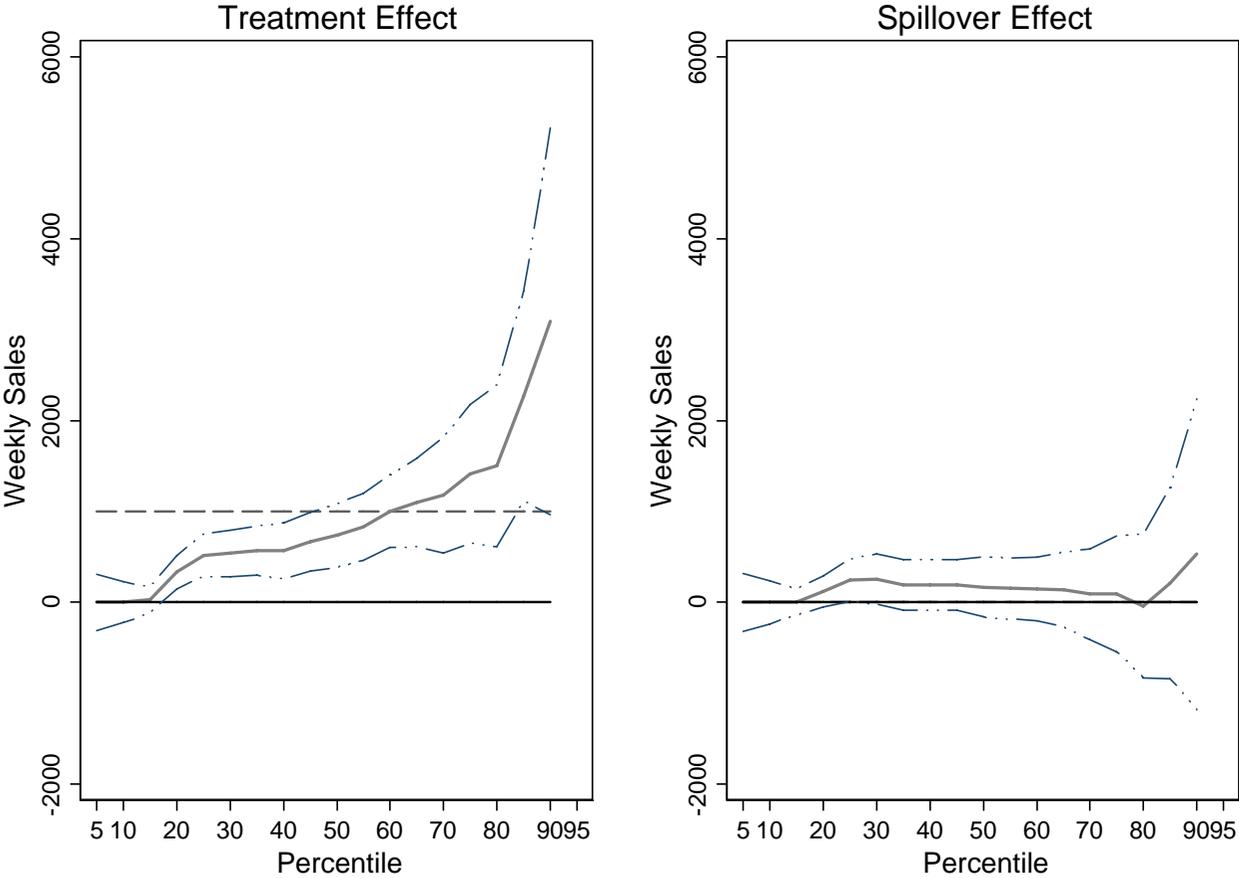
Notes: Graphic shows assignment to treatment. The first intervention of business training assigned firms to the GET AHEAD business training program using a two-stage randomization: first markets were assigned to be treatment or control, and then, within markets, firms were randomly selected to be assigned to training. Then in a second stage, markets with fewer than 16 firms at baseline (“small markets”) were dropped, and half the training treatment markets were assigned to also have those assigned to training within the market receive mentoring. Mentoring was only offered conditional on the firm surviving.

Figure 2: Quantile Treatment Effects on Profits After Three Years



Notes: Intent-to-treat effects from quantile regressions controlling for randomization strata and baseline profits, combining round 4 and round 5 survey data. Straight dashed line shows OLS estimate. 95 percent confidence intervals displayed allowing for clustering at market level.

Figure 3: Quantile Treatment Effects on Sales After Three Years



Notes: Intent-to-treat effects from quantile regressions controlling for randomization strata and baseline sales, combining round 4 and round 5 survey data. Straight dashed line shows OLS estimate. 95 percent confidence intervals displayed allowing for clustering at market level.

Table 1: Individual Characteristics and Verification of Randomization

	Means by Treatment Assignment			Pure Control S.D.	P-value (1) vs (2)	P-value (1) vs (3)	P-value (2) vs (3)
	Treatment Group (1)	Spillover Group (2)	Pure Controls (3)				
Age	36.0	35.6	35.7	9.05	0.482	0.454	0.690
Years of Education	8.92	8.91	9.09	2.92	0.910	0.569	0.515
Married	0.67	0.66	0.67	0.47	0.404	0.638	0.518
Household Size	4.97	4.85	4.85	2.10	0.188	0.262	0.499
Age of Firm	6.39	6.57	6.27	6.92	0.574	0.741	0.403
Number of Employees	0.27	0.27	0.27	0.63	0.747	0.989	0.953
Weekly Profits	1128	1140	1091	834	0.987	0.395	0.322
Weekly Sales	5220	5885	5401	8048	0.065	0.266	0.288
Capital Stock	30571	34092	29370	60530	0.248	0.863	0.101
Ever Received Bank/MFI Loan	0.24	0.25	0.24	0.43	0.891	0.482	0.934
Keeps Records	0.37	0.34	0.34	0.47	0.235	0.094	0.974
Business Practices Score	0.53	0.53	0.52	0.19	0.934	0.487	0.598
Retail Firm	0.77	0.76	0.75	0.43	0.446	0.121	0.474
Registered with City Council	0.44	0.45	0.42	0.49	0.328	0.447	0.100
Joint orthogonality test p-value					0.597	0.435	0.691
Sample Size	1172	988	1377				

Notes:

Tests of treatment versus spillover group control for individual-level randomization strata and are based on robust standard errors. Tests of the treatment or spillover group compared to the pure control group control for market level randomization strata and use standard errors clustered at the market level.

Table 2: Impacts on Primary Outcomes

	Firm Survival	Daily Sales	Weekly Sales	Main Product Sales	Weekly Profits	Main Product Profits	Photo Inventories	Aggregate Index
Panel A: Impact of Assignment to Training, Allowing for Spillovers								
Assigned to Training* 1 Year	0.006 (0.010)	161** (68)	284 (255)	150 (297)	78 (69)	52 (91)	482 (343)	0.049* (0.026)
Assigned to Training* 3 Years	0.030** (0.013)	171** (85)	1038*** (349)	454* (268)	221** (88)	155 (97)	818 (879)	0.088*** (0.032)
Spillover Group * 1 Year	0.002 (0.011)	32 (70)	-476* (270)	157 (328)	-65 (66)	-16 (93)	337 (345)	-0.011 (0.026)
Spillover Group * 3 Years	0.013 (0.014)	1 (85)	27 (330)	182 (270)	-29 (79)	5 (95)	671 (874)	0.002 (0.030)
Mean of Pure Control Group	0.852	1173	5763	3368	1439	1137	8567	0.005
Sample Size	13508	12943	12909	12064	12881	11985	5598	12923
P-value: Training 1 year=3 years	0.056	0.889	0.006	0.380	0.051	0.399	0.679	0.148
P-value: Spillover 1 year=3 years	0.403	0.726	0.103	0.943	0.650	0.866	0.688	0.647
P-value: Training effect zero both years	0.068	0.047	0.008	0.239	0.043	0.269	0.344	0.024
Panel B: Impact of Mentoring Compared to Training Alone, Allowing for Spillovers								
Assigned to Mentoring	0.011 (0.018)	172 (115)	1216** (486)	586 (366)	279** (116)	133 (129)	1909 (1339)	0.098** (0.041)
Assigned to Training Alone	0.052*** (0.016)	209* (110)	954** (427)	523 (357)	300*** (114)	236* (132)	256 (941)	0.104** (0.042)
Spillover Group to Mentoring	0.024 (0.018)	-75 (112)	138 (451)	391 (340)	-16 (99)	27 (120)	1972 (1299)	-0.001 (0.037)
Spillover Group to Training Alone	0.003 (0.020)	122 (105)	-3 (401)	231 (380)	53 (97)	69 (129)	-192 (1095)	0.034 (0.037)
Sample Size	5822	5625	5606	5535	5591	5530	2655	5608
P-value: Mentoring = Training Alone	0.030	0.765	0.605	0.886	0.877	0.522	0.184	0.901

Notes: Robust standard errors in parentheses, clustered at the market level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects, for the baseline value of the outcome, and for survey round fixed effects.

Panel B uses only survey rounds 4 and 5, since mentoring intervention was carried out after round 3.

Aggregate index is the average of standardized z-scores of the other variables. See data appendix for variable definitions.

Table 3: Impacts on Secondary Outcomes

	Employed for pay	All work income	Empowerment Index	Life Ladder Today	Life Ladder 5 Years	Mental Health	Household Durables Index
Panel A: Impact of Assignment to Training, Allowing for Spillovers							
Assigned to Training* 1 Year	0.009 (0.008)	177* (96)	0.109 (0.110)	0.254*** (0.050)	0.253*** (0.064)	0.265* (0.151)	0.108* (0.058)
Assigned to Training* 3 Years	0.020* (0.011)	389*** (141)	-0.037 (0.094)	0.221*** (0.049)	0.322*** (0.054)	0.356** (0.162)	0.091 (0.065)
Spillover Group * 1 Year	0.003 (0.009)	-112 (93)	0.013 (0.113)	-0.011 (0.056)	0.000 (0.068)	0.115 (0.143)	0.048 (0.059)
Spillover Group * 3 Years	0.011 (0.012)	24 (133)	0.070 (0.105)	0.062 (0.046)	0.029 (0.058)	0.272 (0.174)	0.073 (0.070)
Mean of Pure Control Group	0.885	2144	7.034	5.015	7.982	17.707	-0.063
Sample Size	13508	12881	5873	12609	12608	5873	5823
P-value: Training 1 year=3 years	0.323	0.114	0.289	0.581	0.420	0.659	0.813
P-value: Spillover 1 year=3 years	0.533	0.283	0.689	0.297	0.740	0.460	0.721
P-value: Training effect zero both years	0.169	0.019	0.543	0.000	0.000	0.034	0.137
Panel B: Impact of Mentoring Compared to Training Alone, Allowing for Spillovers							
Assigned to Mentoring	0.005 (0.015)	488** (187)	-0.062 (0.133)	0.283*** (0.063)	0.417*** (0.057)	0.477** (0.206)	0.226* (0.121)
Assigned to Training Alone	0.035** (0.014)	477*** (181)	0.133 (0.115)	0.204*** (0.056)	0.271*** (0.064)	0.169 (0.217)	0.132 (0.106)
Spillover Group to Mentoring	0.017 (0.015)	30 (180)	-0.019 (0.142)	0.192*** (0.052)	0.093 (0.067)	0.236 (0.232)	0.306*** (0.117)
Spillover Group to Training Alone	0.005 (0.016)	178 (153)	0.293** (0.128)	0.042 (0.052)	0.064 (0.062)	0.049 (0.221)	-0.000 (0.128)
Sample Size	5822	5591	2470	5511	5511	2470	2465
P-value: Mentoring = Training Alone	0.062	0.958	0.174	0.270	0.037	0.237	0.450

Notes: Robust standard errors in parentheses, clustered at the market level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects, for the baseline value of the outcome where available, and for survey fixed effects. Panel B uses round 4 and 5 data only.

Table 4: Impacts on Market Level for women

	Total Market Profits	Total Market Sales	Total Market Customer Transactions	# New Entrants	Total Other Firms	Other Firm Profits
Market Assigned to Training*1 Year	792 (1378)	2672 (5211)	131 (104)	0.193 (0.509)	1.473 (1.375)	4439 (2940)
Market Assigned to Training *3 Years	2554 (1568)	15353** (6613)	318** (124)	0.833 (0.785)	1.290 (1.784)	2662 (3685)
Mean of Control Markets	27651	110944	1916	8.59	22.48	34870
Sample Size	628	628	628	301	301	301
P-value: 1 Year=3 Years	0.201	0.021	0.038	0.517	0.933	0.684

Notes:

Standard errors in parentheses, clustered at the market level. Unit of observation is market-survey round.

*, **, and *** denote significance at the 10, 5, and 1 percent levels respectively.

Regressions include controls for randomization strata, survey round, and baseline value of outcome where available.

Columns 1 to 3 are for the experimental sample and aggregate over the four follow-up rounds. Columns 4 to 6 are based on market census conducted in round 2 (one year) and round 4 (three years) and are for firms operated by women that are not in our experimental sample.

Table 5: Market Level Impacts in 2017 including males

	Operating and Interviewed	Log Sales	Log Sales	Log Sales	Log Profits	Log Profits	Log Profits
Assigned to Training	0.012 (0.017)	0.141** (0.064)	0.141** (0.064)		0.151** (0.061)	0.152** (0.061)	
Female Spillover Group	0.015 (0.019)	0.060 (0.063)	0.060 (0.063)		0.009 (0.060)	0.009 (0.060)	
Male Spillover Group		0.067 (0.062)			0.032 (0.061)		
Male Spillover in Study Sector			0.056 (0.078)			-0.026 (0.077)	
Male Spillover in Male-Dominated Sector			0.076 (0.071)			0.083 (0.069)	
Owner is Male		0.234*** (0.075)	0.234*** (0.075)	0.218*** (0.037)	0.297*** (0.068)	0.297*** (0.068)	0.268*** (0.038)
Market Assigned to Treatment				0.096** (0.044)			0.082* (0.045)
Mean of Female Pure Control Group	0.836	8.363	8.363	8.363	7.149	7.149	7.149
Sample Size	2984	5184	5184	5184	5113	5113	5113

Notes: Robust standard errors in parentheses, clustered at the market level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects.

First column is only for experimental sample. Remaining columns are for all firms surveyed in a market census taken from June-August 2017 that also included male-operated businesses.

Table 6: Impact on Business Knowledge, Business Practices, and Use of a Mentor

	Business Knowledge	Business Practices	Worked with Mentor
Panel A: Impact of Assignment to Training, Allowing for Spillovers			
Assigned to Training*1 Year	-0.018 (0.089)	0.049*** (0.008)	
Assigned to Training* 3 Years		0.068*** (0.009)	0.335*** (0.037)
Spillover Group * 1 Year	0.124 (0.091)	0.014* (0.008)	
Spillover Group * 3 Years		0.006 (0.009)	0.018 (0.020)
Mean of Pure Control Group	2.000	0.533	0.106
Sample Size	3059	5404	2520
P-value: Training 1 year=3 years		0.067	
P-value: Spillover 1 year=3 years		0.389	
P-value: Training effect zero both years		0.000	
Panel B: Impact of Mentoring Compared to Training Alone, Allowing for Spillovers			
Assigned to Mentoring		0.096*** (0.012)	0.666*** (0.028)
Assigned to Training Alone		0.033*** (0.010)	0.108*** (0.020)
Spillover Group to Mentoring		-0.006 (0.011)	0.009 (0.023)
Spillover Group to Training Alone		0.004 (0.012)	0.047* (0.026)
Sample Size		2236	2212
P-value: Mentoring = Training Alone		0.000	0.000

Notes: Robust standard errors in parentheses, clustered at the market level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects.

Business knowledge only collected in round 2, mentoring only collected in round 4.

Business practices collected in round 2 and 4, with regression controlling for survey round and for baseline business practices.

Outcomes are conditional on operating a firm

Table 7: Impact on Dealings with Customers

	Weekly Customers	Gained new Customer	Lost regular Customer	Sales per Customer	Open Business at set time	Introduced New Product
Panel A: Impact of Assignment to Training, Allowing for Spillovers						
Assigned to Training* 1 Year	1.7 (3.8)	0.023* (0.013)	-0.038** (0.017)	25.5 (30.0)		0.116*** (0.022)
Assigned to Training* 3 Years	14.5*** (4.4)	0.034** (0.016)	-0.041*** (0.015)	38.7 (25.9)	0.062*** (0.024)	0.089*** (0.021)
Spillover Group * 1 Year	-2.2 (3.9)	0.009 (0.014)	-0.020 (0.018)	0.7 (20.6)		0.007 (0.021)
Spillover Group * 3 Years	-0.7 (4.6)	0.008 (0.018)	-0.009 (0.016)	2.2 (17.1)	0.030 (0.023)	0.008 (0.021)
Mean of Pure Control Group	100	0.767	0.804	134	0.342	0.269
Sample Size	12867	12977	12978	11186	2398	6216
P-value: Training 1 year=3 years	0.004	0.586	0.878	0.734		0.327
P-value: Spillover 1 year=3 years	0.753	0.964	0.666	0.951		0.970
P-value: Training effect zero both years	0.003	0.049	0.003	0.247		0.000
Panel B: Impact of Mentoring Compared to Training Alone, Allowing for Spillovers						
Assigned to Mentoring	12.1* (6.1)	0.004 (0.021)	-0.048*** (0.017)	50.3 (38.0)	0.103*** (0.033)	0.102*** (0.028)
Assigned to Training Alone	12.0** (5.2)	0.054*** (0.020)	-0.054*** (0.017)	24.6 (24.6)	0.030 (0.031)	0.084*** (0.028)
Spillover Group to Mentoring	0.8 (6.4)	0.000 (0.022)	-0.035* (0.020)	-5.3 (19.1)	0.026 (0.030)	0.021 (0.026)
Spillover Group to Training Alone	-4.3 (6.0)	0.009 (0.023)	-0.002 (0.019)	-0.0 (20.5)	0.039 (0.032)	0.022 (0.029)
Sample Size	5602	5636	5636	4801	2090	2636
P-value: Mentoring = Training Alone	0.992	0.020	0.758	0.558	0.071	0.597

Notes: Robust standard errors in parentheses, clustered at the market level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects.

Opening at set time only asked in round 4 survey and is conditional on business operating.

Table 8: Multiplier Effects - Impact on Percent of Customers that are other businesses

	Percent of Customers that are other businesses (ordered)	Percent of Customers that are other businesses (discrete)
Assigned to Training	0.202*** (0.075)	3.855*** (1.431)
Female Spillover Group	0.100 (0.075)	1.749 (1.403)
Male Spillover in Study Sector	0.137 (0.089)	2.595 (1.703)
Male Spillover in Male-dominated Sector	0.169** (0.083)	3.377** (1.586)
Owner is Male	0.122 (0.082)	2.501 (1.592)
Mean of Female Pure Control Group	2.442	29.230
Sample Size	5384	5384

Notes: Robust standard errors in parentheses, clustered at the market level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects.

Table 9: Impacts on Financial Control, Access to Finance, and Capital Levels

	Ratio of Worst/Current Profit	Monitors Sales Trends	Fraction Stock Lost to spoilage	Received a bulk discount	Production Cost Change (%)	Received Loan from Bank/MFI	Has Business Bank Account	Inventory Value	Capital Stock Value
Panel A: Impact of Assignment to Training, Allowing for Spillovers									
Assigned to Training*1 Year	0.014 (0.009)	0.112*** (0.023)	-0.007 (0.014)	0.009 (0.023)	15.536 (15.561)	0.017 (0.017)		2475 (2125)	-52 (882)
Assigned to Training* 3 Years	-0.007 (0.011)	0.141*** (0.022)	-0.014 (0.021)	0.034 (0.023)	13.954 (19.084)	0.013 (0.019)	0.058** (0.022)	5889** (2978)	3219*** (1085)
Spillover Group * 1 Year	-0.004 (0.009)	0.025 (0.024)	0.037** (0.017)	-0.028 (0.025)	7.825 (17.121)	0.014 (0.018)		-339 (1966)	-828 (917)
Spillover Group * 3 Years	-0.011 (0.012)	0.000 (0.023)	-0.005 (0.024)	0.013 (0.024)	-5.558 (20.019)	-0.017 (0.021)	-0.003 (0.023)	3350 (2829)	672 (1065)
Mean of Pure Control Group	0.450	0.269	0.194	0.423	160.4	0.172	0.354	19918	13410
Sample Size	10609	6214	4989	5879	5541	5404	3005	6183	6203
P-value: Training 1 year=3 years	0.136	0.263	0.766	0.447	0.935	0.850		0.107	0.002
P-value: Spillover 1 year=3 years	0.622	0.355	0.102	0.255	0.526	0.139		0.093	0.174
P-value: Training effect zero both years	0.252	0.000	0.767	0.308	0.569	0.583		0.137	0.005
Panel B: Impact of Mentoring Compared to Training Alone, Allowing for Spillovers									
Assigned to Mentoring	-0.024* (0.012)	0.194*** (0.029)	-0.038 (0.026)	0.029 (0.030)	-8.074 (23.297)	0.004 (0.026)	0.048 (0.034)	15371*** (4523)	3180** (1504)
Assigned to Training Alone	-0.015 (0.013)	0.070*** (0.024)	-0.009 (0.028)	0.002 (0.030)	-15.671 (25.575)	0.044* (0.026)	0.099*** (0.025)	4705* (2793)	2561* (1330)
Spillover Group to Mentoring	-0.010 (0.014)	-0.012 (0.031)	0.002 (0.032)	-0.012 (0.030)	-3.198 (25.752)	-0.023 (0.027)	0.010 (0.035)	9707** (3770)	628 (1365)
Spillover Group to Training Alone	-0.029** (0.014)	-0.014 (0.032)	-0.029 (0.034)	0.010 (0.032)	-41.846 (26.112)	0.006 (0.028)	0.023 (0.025)	2649 (3198)	-133 (1451)
Sample Size	4563	2636	2078	2497	2389	2236	2636	2634	2635
P-value: Mentoring = Training Alone	0.515	0.000	0.353	0.432	0.791	0.157	0.137	0.021	0.720

Notes: Robust standard errors in parentheses, clustered at the market level. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

All regressions control for randomization strata fixed effects. Mentoring regressions only for three year follow-up.

Business bank account not asked in round 2 follow-up survey.

Table 10: Attitudes and Associations

	Own Labor Hours	Entrepreneurial Self-Efficacy	GET Ahead Attitudes	Women's Association	# Discuss Business	Works together with other women
Panel A: Impact of Assignment to Training, Allowing for Spillovers						
Assigned to Training*1 Year	0.059 (1.106)	-0.037 (0.181)	-0.018 (0.021)	0.003 (0.017)	0.181 (0.210)	0.035 (0.024)
Assigned to Training* 3 Years	4.596*** (1.274)			0.057*** (0.018)	0.584** (0.277)	0.033 (0.026)
Spillover Group * 1 Year	-2.060* (1.138)	-0.180 (0.171)	-0.047** (0.022)	0.026 (0.019)	0.032 (0.224)	-0.000 (0.025)
Spillover Group * 3 Years	-0.192 (1.356)			-0.013 (0.017)	0.066 (0.308)	0.010 (0.024)
Mean of Pure Control Group	42.6	4.457	3.802	0.115	4.672	0.490
Sample Size	6215	3059	3059	5400	5398	5403
P-value: Training 1 year=3 years	0.001			0.042	0.259	0.959
P-value: Spillover 1 year=3 years	0.189			0.155	0.934	0.731
P-value: Training effect zero both years	0.001			0.005	0.072	0.242
Panel B: Impact of Mentoring Compared to Training Alone, Allowing for Spillovers						
Assigned to Mentoring	4.777*** (1.642)			0.055** (0.023)	0.920** (0.362)	0.027 (0.039)
Assigned to Training Alone	4.300** (1.677)			0.062*** (0.023)	0.303 (0.329)	0.030 (0.032)
Spillover Group to Mentoring	-0.429 (1.758)			-0.029 (0.020)	0.040 (0.417)	-0.016 (0.032)
Spillover Group to Training Alone	0.101 (1.659)			0.002 (0.025)	-0.003 (0.390)	0.041 (0.031)
Sample Size	2635			2234	2235	2236
P-value: Mentoring = Training Alone	0.814			0.795	0.135	0.940

Notes: Robust standard errors in parentheses, clustered at the market level. *, **, and *** denote significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects. Mentoring regressions only for three year follow-up. Self-efficacy and Get Ahead attitudes only asked in round 2.

ONLINE APPENDICES

Appendix 1: Timeline

Appendix 2: Additional Details on Listing and Sample Selection

Appendix 3: Model for a GET Ahead 5-day workshop for entrepreneurs

Appendix 4: Sample Means by Assignment to Mentoring

Appendix 5: Mentoring design and implementation

Appendix 6: Measurement of Key Outcomes

Appendix 7: Data Availability

Appendix 8: LATE Impacts for Primary Outcomes

Appendix 9: Robustness Checks on Primary Outcomes

Appendix 10: Robustness to Excluding Nearby Markets

Appendix 11: Sharpened Q-values for Mechanism Impacts

Appendix 12: No Impact on Market Prices or Purchase Costs

Appendix 1: Timeline

Listing and Baseline Survey: June-November 2013

Business Training Intervention: June-November 2013

Round 2 Follow-up and Round 2 Market Census: June-October 2014 (One-Year Post-Training)

Round 3 Short Follow-up: November 2014-February 2015

Mentoring Intervention: July-November 2015

Round 4 Follow-up and Round 4 Market Census: February-July 2016 (Three-Years Post-Training; 6-10 Months Post-Mentoring)

Round 5 Follow-up: May-October 2016.

Market Census Including Male-Owned Firms and Customer Survey: June ó Aug 2017

Appendix 2: Additional Details on Listing and Sample Selection

Altogether 6,296 female-owned businesses in 161 markets were listed. After the census, three markets in Kakamega county were dropped because the number of women in these markets was too few. We then applied an eligibility filter to determine which women to include in the baseline survey. This filter required the women to have reported profits, and not to have reported profits that exceeded sales; to have a phone number that could be used to invite them for training; to be 55 years or younger in age; to not be running a business that only dealt with phone cards or m-pesa, or that was a school; that the person responding not be an employee; that the business not have more than 3 employees; that the business have profits in the past week between 0 and 4000 KSH; that sales in the past week be less than or equal to 50,000 KSH; and that the individual had at least one year of schooling. These criteria were chosen to reduce the amount of heterogeneity in the sample (thereby increasing our ability to detect treatment effects), and to increase the odds of being able to contact and find individuals again. Applying this eligibility filter reduced the 6,296 individuals to 4,037 individuals (64%). Baseline surveys took place soon after the listing surveys in each county, between June and November 2013. Out of a target of 4,037 individuals, we were able to interview 3,537 (87.6%) in time to consider them for inviting to training.

Appendix 3: Model for a GET Ahead 5-day workshop for entrepreneurs

Day 1	Day 2	Day 3	Day 4	Day 5
<p>Opening</p> <p>Gender equality promotion: life cycle of people and enterprises</p> <p>The business woman: she can do it</p>	<p>The business environment: she is not alone</p> <p>Business ideas</p> <p>Marketing</p>	<p>Production, services and technology</p> <p>Marketing</p>	<p>Finance</p> <p>Manage of self and others</p> <p>Business support and networking</p>	<p>Management or networking</p> <p>Action</p> <p>Planning</p> <p>Closing</p>

Appendix 4: Sample Means by Assignment to Mentoring or Training Only

Appendix Table 1: Means by Treatment Status for Mentoring Intervention

	Assigned Mentor (1)	Training Only (2)	Spillover for Mentor (3)	Spillover for Training Only (4)	Pure Control (5)	P-value (1) vs (2)	P-value (3) vs (4)
Age	35.73	36.39	35.38	35.70	35.80	0.262	0.807
Years of Education	9.08	8.76	9.00	8.89	9.08	0.019	0.321
Married	0.69	0.65	0.68	0.64	0.68	0.166	0.999
Household Size	4.96	4.99	4.93	4.83	4.86	0.962	0.841
Age of Firm	6.36	6.82	6.53	6.73	6.31	0.310	0.491
Number of Employees	0.27	0.26	0.27	0.27	0.26	0.985	0.422
Weekly Profits	1124	1115	1132	1155	1085	0.782	0.643
Weekly Sales	5657	5054	6028	6016	5247	0.596	0.558
Capital Stock	37859	24222	39042	27324	26733	0.001	0.005
Ever Received Bank/MFI Loan	0.26	0.26	0.23	0.27	0.23	0.652	0.061
Keeps Records	0.36	0.37	0.33	0.36	0.34	0.967	0.318
Business Practices Score	0.53	0.52	0.53	0.53	0.52	0.153	0.920
Retail Firm	0.77	0.81	0.76	0.79	0.75	0.223	0.115
Registered with City Council	0.48	0.36	0.49	0.39	0.40	0.001	0.023
Sample Size	524	521	459	442	1158		

Notes: Means and Sample Sizes are shown for Sample in Non-Small Markets, for which mentoring intervention applied.

Appendix 5: Mentoring Design and Implementation

The mentoring intervention provided personalized, hands-on problem-solving support and peer learning to women who had previously received the Get Ahead program with the goal of reinforcing intended business training outcomes ó from improved management skills to business growth.

The intervention targeted 446 women who had been exposed to the business training in 2013 and expressed interest in further support through mentorship. Two local, public service providers, Kenya Industrial Estate (KIE) and the Women Enterprise Fund (WEF), were identified as partners in the delivery of the mentoring services. KIE implemented the mentoring in Kakamega and Kisii, while WEF did it in Embu and Kitui.

The design of the mentoring intervention combined group and individual sessions for a period of five months (July to November 2015). Each female-owned firm (or mentee) received 15 mentoring sessions: 10 through group sessions and five through one-to-one meetings with the mentor. Group sessions occurred twice a month, every two weeks, while individual sessions took place once a month. Each mentor was assigned a group of five mentees. The table below summarizes the topics covered during the 10 group mentoring sessions. Individual sessions deepened discussions on the above topics based on the needs of the mentee and her business.

The ILO contracted a mentoring expert for program development and curriculum design, which was subsequently discussed and delivered to mentors from Kenya Industrial Estate and the Women Enterprise Fund.

110 mentors were recruited, only 100 were subsequently trained, and 89 selected for the program.

The program reached out to 446 women who had participated in the Get Ahead Program, 392 signed up, were inducted and received training. However, only 361 women stayed in the program throughout the five months.

Therefore, while originally 89 mentors were linked to 392 mentees, by the end of the program the number of mentees per mentor ranged from 3 to 6.

Drop out reasons included (i) in most cases lack of interest after realizing there were no grants involved in the offer, and (ii) in few cases considerations about the mentoring program not being helpful in enhancing skills and business growth.

Monitoring and evaluation of the mentoring was done through (i) monthly meetings with mentors and implementing partners, (ii) monthly reports per mentor, (iii) tracking of journals filled by mentors after each session, (iv) visits and phone calls with mentors for follow up and support, (v) two meetings between the ILO and the implementing partners throughout the duration of the program, and (vi) one closing forum in each county to gather feedback from all parties.

Mentoring approach: group sessions

Session	Objective/Module	Outcome	Time
Session 1: Introduction, objectives and agreement	Introduction of mentor mentees Definition of objectives Understanding the mentoring agreement	Knowing each other Being clear on the mentorship objective and expectations Understanding and signing the mentoring agreement	2 hrs
Session 2: GROW goal	Goals review	Mentees goals established How the mentees will look like when they are successful How will the business look like after growing, as regards to sourcing, production, marketing, packaging, sales, networking, access to finance and general business management	2 hrs
Session 3: GROW reality	Reality analysis	Current status established Where is the mentee right now in view of her personal and business goals Business performance established in terms of; sourcing, production, marketing, packaging, sales, networking, access to finance and general business management	2 hrs
Session 4: GROW gap analysis	Identifying key gaps (goals vis-à-vis current status)	Gaps identified. This will be done by exploring the difference between the goal and the current reality.	2 hrs
Session 5: GROW exploring options	Exploring options to addressed identified gaps	A list of options to close each of the gaps identified (in terms of capabilities, skills and assets).	2 hrs
Session 6: Feasibility assessment	Financial analysis	Exploring the feasibility of filling in the gaps What is the cost of implementing the options What will be the source of finance required to close the gaps	2 hrs
Session 7: GROW way forward /action points	Plan of action to address identified gaps	Analysing options and agree on way forward Drawing an action plan with clear timelines and budget	2 hrs
Session 8: Implementing the way forward	Evaluating progress	Exploring what the mentee has done on the agreed Action Points Identifying successes, challenges, way forward	2 hrs
Session 9: Implementing the way forward	Evaluating progress	Exploring what the mentee has done on the agreed action points Identifying successes, challenges, way forward	2 hrs
Session 10: Evaluation, review of action plan and sustainability	End of program evaluation, review of action plan and agreement on sustainability plan	Documenting the impact of the mentorship, i.e. how has the mentee and business benefited from the mentoring sessions Reviewing the action plan based on sessions 8 and 9 Agreeing on a sustainability plan after the end formal mentorship	2 hrs

Appendix 6: Measurement of Key Outcomes

All nominal values were converted into real (August 2013) Kenyan Shillings using the consumer price index for the midpoint of each survey round.

Firm Survival is measured as whether the owner still operates a business, regardless of whether or not they have changed the business line. For individuals who could not be interviewed, survival was measured by asking family members and neighbors the status of the owner.

Daily sales are sales of the business in the last day, coded as zero if the business was closed that day, or is closed for good. It is truncated at the 99th percentile.

Weekly sales are total sales of the business in the last week, coded as zero if the business was closed that week, or is closed for good. It is truncated at the 99th percentile. When aggregated across all firms in the market, this forms *Total market sales*.

Main product sales are obtained by multiplying the number of units of the main product sold in the last week by the unit price, truncated at the 99th percentile and coded as zero if the business is closed.

Business Profits are measured using the direct question of de Mel et al. (2009)¹⁵, asked about the last week as a reference period since pre-testing found a weekly rather than monthly recall was more accurate. The question was: "What was the total income the business earned during last week after paying all expenses including wages of employees, but not including any income you received from other sources?" The response was coded as zero if the business is closed, and truncated at the 99th percentile. When aggregated across all firms in the market, this forms *Total market profits*.

Main product profits are obtained by multiplying the mark-up on the main product sold by the number of units sold of this product in the past week, truncated at the 1st and 99th percentile, and coded as zero if the business is closed.

Photo inventories are the value of inventories as assessed by valuing a photograph of the business inventories. A common set of market prices are used to aggregate products. Two independent enumerators would count the number of each product they see in the photo (e.g. 53 tomatoes) and then aggregate by the price per product. If the two valuations differed by more than 5000 KSH, they would iterate again until they agreed on a valuation.

Aggregate index of profits and sales is the average of standardized z-scores of the primary profits and sales measures.

Employed for pay is coded as one if they are self-employed or worked for wages in the past week.

Income from work is the sum of weekly profits and income from wage work in the past week.

¹⁵ F g'O gn'Uwtgij 0'O eMgp| kg.'F cxkf 0'Y qqf twth'Ej tkvqr j gt0*422; +06O gcwtlpi 'b letqgpvgr tkug'r tqhku'0 wuv'y g' cunfj qy 'y j g'ucwuci g'ku'b cf gAö." *Journal of Development Economics* 88(1): 19-31

Empowerment index is the sum of the following outcomes: Compelled to spend money on husband or family (coded 1 if they answer no); not the only person with access to a car (coded as 1 if only they have access); has some money which they have sole control over and can spend as they wish; to travel to a new location to work, to stay overnight in a different town, to work later than usual hours, to take out a loan, or to spend money on an investment for their business.

Life Ladder Today and *Life Ladder 5 Years* are measured by a standard 10-step Cantril ladder, where individuals are asked to imagine the best (step 10) and worst (step 1) possible lives for themselves, and then say which step represents their current position and where they will be in 5 years.

Mental health is measured by the MHI-5 index of Veit and Ware (1983), coded so higher scores denote better mental health.

Household durables index is the first principal component of dummy variables for ownership of 10 household assets (iron and heaters, fridge or freezer, fan, sewing machine, radio or CD player, TV or DVD player, Motorcycle or scooter, Car or Van, oven, and gas cooker) and of the number of cows owned and number of goats owned.

Number of new entrants is the number of female-owned firms operating in the market outside of our experimental sample that have opened since the baseline survey and training intervention.

Total other firms is the total number of other female-owned firms operating in the market outside of our experimental sample (new entrants plus those pre-existing firms that were not included due to being absent from the market at the time of listing, or being dropped by our eligibility filters).

Other firm profits is the total profits in the market of the other female-operated firms outside the survey sample.

Weekly customers is the number of customers the firm has in the past week, truncated at the 99th percentile. This is coded as zero for firms which are closed. When aggregated to the market level, this forms *Total market customer transactions*.

Business knowledge is the number correct out of 7 questions intended to measure whether the firm owner can calculate sales, expenses, and profits.

Business practices is the proportion of 26 practices in marketing, record-keeping, buying and stock control, and financial planning used by the firm (McKenzie and Woodruff, 2015). This is only measured for firms which are surviving at the time of the survey.

Worked with a mentor is a dummy variable for whether they have worked with a mentor to try to improve their business in the past year (only asked in round 4).

Entrepreneurial self-efficacy is the number out of 10 of business activities that the owner rates as high confidence (only asked in round 2). This includes: (1) starting a business, (2) hiring employees, (3) negotiating with suppliers, (4) negotiating with customers, (5) raising capital, (6) developing a business plan, (7) marketing, (8) managing cash flow, (9) managing inventory, and (10) managing risk.

Get Ahead Attitudes is the sum of scores on 11 questions designed to measure attitudes Get Ahead training is meant to encourage. These are scored 1 through 5, where 1=strongly disagree, 5 = strongly agree. Questions will be coded so that higher scores indicate better entrepreneurial attitudes

Even when my business is going well, I keep my eyes open in case I
I just plan week to week based on
My business provides about
(negatively coded).

Appendix 7: Data Availability

Appendix Table 2 shows data availability by treatment status.

Appendix Table 2: Data Availability by Treatment Status

	Round 2	Round 3	R2 or R3	Round 4	Round 5	R4 or R5
Panel A: Interviewed						
Assigned to Training	0.034*** (0.012)	0.030*** (0.011)	0.023*** (0.008)	0.009 (0.013)	0.023* (0.013)	0.010 (0.011)
Spillover Group	0.014 (0.013)	-0.013 (0.014)	-0.003 (0.010)	-0.026* (0.014)	0.002 (0.015)	-0.009 (0.012)
Pure Control Mean	0.886	0.889	0.943	0.894	0.876	0.923
Panel B: Data on Survival Available						
Assigned to Training	0.011* (0.006)	0.016** (0.006)	0.002 (0.003)	0.005 (0.009)	0.010 (0.010)	0.003 (0.007)
Spillover Group	0.015** (0.006)	-0.003 (0.008)	0.002 (0.004)	0.000 (0.010)	0.008 (0.010)	0.004 (0.007)
Pure Control Mean	0.968	0.962	0.991	0.947	0.924	0.970
Panel C: Data on Weekly Sales and Profits Available						
Assigned to Training	0.027** (0.012)	0.031*** (0.009)	0.016** (0.006)	0.013 (0.013)	0.031** (0.013)	0.011 (0.010)
Spillover Group	0.015 (0.013)	-0.006 (0.012)	-0.003 (0.008)	-0.007 (0.014)	0.011 (0.014)	0.006 (0.011)
Pure Control Mean	0.907	0.913	0.964	0.903	0.881	0.939
Sample Size	3537	3537	3537	3537	3537	3537

Notes: Robust Standard Errors in Parentheses, Clustered at the Market Level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

Data on weekly sales and profits availability codes data as available if firm is known to be closed (since then sales and profits are known to be zero).

Appendix 8: LATE Impacts for Primary Outcomes

Appendix Table 3: LATE Impacts on Primary Outcomes

	Firm Survival	Daily Sales	Weekly Sales	Main Product Sales	Weekly Profits	Main Product Profits	Photo Inventories	Aggregate Index
Impact of Receiving Training, Allowing for Spillovers								
Received Training*1 Year	0.007 (0.012)	202** (85)	355 (318)	187 (370)	97 (86)	65 (113)	597 (421)	0.061* (0.032)
Received Training* 3 Years	0.038** (0.016)	212** (104)	1283*** (428)	562* (329)	273** (108)	192 (119)	1017 (1087)	0.108*** (0.039)
Spillover Group * 1 Year	0.002 (0.011)	32 (70)	-476* (268)	157 (327)	-65 (66)	-16 (93)	336 (343)	-0.011 (0.026)
Spillover Group * 3 Years	0.013 (0.014)	1 (85)	25 (328)	181 (269)	-30 (79)	5 (94)	670 (867)	0.002 (0.029)
Sample Size	13508	12943	12909	12064	12881	11985	5598	12923

Notes: Robust standard errors in parentheses, clustered at the market level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects, the baseline value of the outcome, and survey round fixed effects.

Receipt of training instrumented with assignment to training.

Appendix 9: Robustness Checks on Primary Outcomes

We carry out three robustness checks on our primary outcomes. First, we use the objective visual photographic evidence as to whether the business looks bigger or not than it was at baseline (Appendix Table 4). Second, we report results conditional on survival (Appendix Table 5). Third, we examine transforms of profits and sales which are less susceptible to outliers (Appendix Table 6). Since our initial screening was intended to create relatively homogeneous firms, our pre-analysis plan specified that our primary analysis would focus on the outcome variables measured in terms of levels. However, it indicated that as a robustness check, we would also examine the impact of treatment on the inverse hyperbolic sine transformation of total profits in the last month, and total sales in the last week: $\log(y+(y^2+1)^{1/2})$. This is similar to the log transformation, but allows for zeros and negative values. We also consider the log transformation, which conditions on survival. The first column of Appendix Table 6 also considers an alternative definition of survival, in which we assume which firms which were observed closed in round T, and then not observed in subsequent waves, remain closed. Our results are robust to this alternative.

In the last two columns of Appendix Table 6 we examine whether training leads to more accurate reporting. Column 6 measures the treatment impact on the number of reporting errors made. We consider six errors: reporting sales less than profits, daily sales less than weekly sales, worst week profits and sales, and unit costs higher than prices for their most profitable product. We find no significant treatment effect on the number of errors made, suggesting that training does not affect reporting accuracy. Column 7 considers the difference (in log terms) between the value of inventories reported by the firm, and the value assessed by our team based on their photo. If treated participants were deliberately reporting their businesses to be bigger, we should expect a systematic bias in this difference with treatment status. However, we cannot reject that the difference does not vary with treatment status, suggesting reporting accuracy does not change with treatment. Together with the physical evidence in Appendix Table 4, this suggests the greater size of the treatment firms is a real, and not just a reporting, effect.

Appendix Table 4: Is the Business Visually Bigger After Three Years?

	All firms	Survivors	All Firms	Survivors
Assigned to Training	0.042** (0.021)	0.037 (0.023)		
Spillover Group	0.019 (0.022)	0.019 (0.023)		
Assigned to Mentoring			0.059** (0.028)	0.058* (0.031)
Spillover Group for Mentoring			0.022 (0.030)	0.017 (0.032)
Assigned to Training Only			0.034 (0.028)	0.022 (0.029)
Spillover Group for Training			0.028 (0.026)	0.023 (0.027)
Sample Size	2864	2571	2527	2265
P-value: Mentoring=Training Only			0.440	0.286

Notes:

Robust standard errors in parentheses, clustered at the market level.

*, **, and *** denote significance at the 10, 5, and 1 percent levels respectively.

Outcome is whether photograph of business inventories in Round 4 survey is shows business to be larger than baseline photograph shows.

Columns 2 and 4 condition on the firm surviving, columns 1 and 3 code closed firms as not being bigger.

Columns 3 and 4 are for the "non-small" markets for which mentoring was randomized.

Appendix 5: Impacts on Primary Outcomes Conditional on Survival

	Daily Sales	Weekly Sales	Main Product Sales	Weekly Profits	Main Product Profits	Photo Inventories	Aggregate Index
Panel A: Impact of Assignment to Training, Allowing for Spillovers							
Assigned to Training* 1 Year	165** (77)	294 (287)	126 (329)	69 (75)	46 (102)	604 (392)	0.051* (0.029)
Assigned to Training* 3 Years	145 (94)	967** (382)	380 (294)	182* (98)	124 (107)	487 (1017)	0.080** (0.035)
Spillover Group * 1 Year	30 (78)	-557* (290)	168 (368)	-79 (70)	-23 (105)	357 (404)	-0.012 (0.028)
Spillover Group * 3 Years	-24 (95)	-46 (354)	169 (303)	-61 (86)	-17 (108)	607 (982)	-0.001 (0.033)
Mean of Pure Control Group	1386	6818	3993	1702	1348	10694	0.107
Sample Size	11339	11305	10460	11277	10381	4775	11319
P-value: Training 1 year=3 years	0.804	0.027	0.509	0.177	0.567	0.902	0.324
P-value: Spillover 1 year=3 years	0.574	0.138	0.998	0.840	0.964	0.797	0.733
P-value: Training effect zero both years	0.097	0.034	0.435	0.180	0.494	0.307	0.068
Panel B: Impact of Mentoring Compared to Training Alone, Allowing for Spillovers							
Assigned to Mentoring	171 (132)	1289** (529)	628 (395)	295** (132)	132 (142)	2081 (1560)	0.107** (0.045)
Assigned to Training Alone	165 (118)	792* (456)	360 (383)	233* (119)	183 (140)	-658 (1093)	0.087* (0.044)
Spillover Group to Mentoring	-124 (120)	-28 (467)	355 (367)	-68 (102)	1 (133)	2020 (1414)	-0.015 (0.039)
Spillover Group to Training Alone	139 (118)	26 (438)	243 (441)	62 (110)	66 (150)	-460 (1220)	0.046 (0.042)
Sample Size	4862	4843	4772	4828	4767	2193	4845
P-value: Mentoring = Training Alone	0.967	0.354	0.559	0.674	0.761	0.058	0.693

Notes: Robust standard errors in parentheses, clustered at the market level

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects, for the baseline value of the outcome and for survey round fixed effects. Panel B uses only survey rounds 4 and 5, since mentoring intervention was carried out between round 3 and round 4. Aggregate index is the average of standardized z-scores of the other variables. See data appendix for description of the different outcome variables.

Appendix Table 6: Robustness of Primary Outcome Results

	Alternative Survival	Inverse Hyperbolic Sales	Log Profits	Log Sales	Log Profits	Number Errors	Inventory Reporting
Assigned to Training*1 Year	0.006 (0.010)	0.124 (0.116)	0.098 (0.101)	0.094** (0.037)	0.068 (0.047)	-0.014 (0.026)	0.060 (0.063)
Assigned to Training* 3 Years	0.031** (0.014)	0.569*** (0.136)	0.428*** (0.119)	0.172*** (0.048)	0.138*** (0.051)	0.006 (0.030)	0.078 (0.063)
Spillover Group * 1 Year	0.001 (0.011)	-0.015 (0.120)	-0.008 (0.103)	-0.027 (0.040)	0.010 (0.045)	0.004 (0.029)	-0.053 (0.066)
Spillover Group * 3 Years	0.013 (0.014)	0.237 (0.151)	0.182 (0.130)	0.009 (0.047)	0.043 (0.047)	-0.014 (0.029)	0.028 (0.074)
Mean of Pure Control Group	0.835	7.270	6.240	8.265	6.889	0.542	0.849
Sample Size	13666	12909	12881	10420	10790	11541	4391
P-value: Training 1 year=3 years	0.046	0.001	0.003	0.036	0.112	0.607	0.836
P-value: Spillover 1 year=3 years	0.374	0.069	0.110	0.385	0.463	0.638	0.375
P-value: Training effect zero both years	0.074	0.000	0.001	0.002	0.027	0.847	0.344

Notes: Robust standard errors in parentheses, clustered at the market level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects, the baseline value of the outcome where available, and for survey round fixed effects. See test for variable definitions

Appendix 10: Robustness to Excluding Nearby Markets

Appendix Table 7 shows that our main results of a strong treatment effect, and a small and statistically insignificant spillover effect (repeated in column 1) continue to hold when we exclude markets in which consumers have the greatest ability to switch from shopping at other markets. We consider two definitions of having another market nearby. The first is based on our 2017 customer surveys. Columns 2 and 3 exclude markets in which customers say they spend less than 50% on average of their food expenditure and less than 70% of food expenditure in that market. This measure has the advantage of excluding markets where customers also go to the largest urban center/administrative capital in a county to shop (such as Embu town or Kitui market), but has the disadvantage that it classifies markets based on post-treatment customer behavior. The second measure comes from calculating the (straight-line) distance from each market in our study to every other market, and then dropping markets which are within 2km of another study market (column 4) or within 5km (column 5). We see our results are robust to excluding markets with nearby shopping options.

Appendix Table 7: Robustness of Impact on Profits and Sales Index of Excluding Markets With Other Markets Nearby

	Full Sample	Excluding Markets where Mean Food Share of Customers is:		Excluding Markets where nearest study market is within	
		<50%	<70%	2km	5km
Assigned to Training* 1 Year	0.049* (0.026)	0.054* (0.028)	0.060* (0.033)	0.068** (0.029)	0.077* (0.041)
Assigned to Training* 3 Years	0.088*** (0.032)	0.100*** (0.034)	0.120*** (0.039)	0.084** (0.037)	0.117** (0.049)
Spillover Group * 1 Year	-0.011 (0.026)	-0.006 (0.026)	0.028 (0.031)	0.001 (0.029)	-0.004 (0.040)
Spillover Group * 3 Years	0.002 (0.030)	0.018 (0.032)	0.043 (0.037)	0.018 (0.036)	0.038 (0.043)
Mean of Pure Control Group	0.005	0.003	-0.032	0.003	-0.026
Sample Size	12923	11892	7781	10926	6806
Number of Markets	157	147	97	135	82

Notes: Robust standard errors in parentheses, clustered at the market level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects, for the baseline value of the outcome, and for survey round fixed effects.

Appendix 11: Sharpened Q-values for Mechanism Impacts

To control for multiple hypothesis testing when examining multiple mechanisms, we to construct sharpened q-values following Anderson (2008) and Benjamini et al. (2006). This process uses a two-stage procedure to control the false discovery rate when reporting results for specific outcomes.¹⁶ Appendix Table 8 reports the original p-values and corresponding sharpened q-values. We see that all three year outcomes that have p-values below 0.05 also have sharpened q-values

¹⁶ Anderson, Michael (2008), "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects", *Journal of the American Statistical Association*, 103(484), 1481-1495; and Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli (2006) "Adaptive Linear Step-Up Procedures That Control the False Discovery Rate." *Biometrika* 93 (3): 491–507.

below this level. In contrast, over the one year horizon, only the impacts on business practices, introducing new products, and monitoring sales trends are significant after this adjustment.

Appendix Table 8: P-values and Sharpened Q-values for Mechanisms

Table	Measure	1 Year		3 Years	
		P-value	Sharpened Q	P-value	Sharpened Q
Table 6	Business Knowledge	0.839	1.000	0.175	0.096
Table 6	Business Practices	0.000	0.001	0.000	0.001
Table 6	Worked with Mentor			0.000	0.001
Table 7	Weekly Customers	0.648	1.000	0.001	0.004
Table 7	Gained New Customer	0.082	0.387	0.040	0.039
Table 7	Lost New Customer	0.025	0.127	0.006	0.011
Table 7	Sales per customer	0.396	0.847	0.138	0.087
Table 7	Open set time			0.010	0.015
Table 7	Introduce new product	0.000	0.001	0.000	0.001
Table 9	Profit ratio	0.131	0.537	0.530	0.249
Table 9	Monitors Sales Trends	0.000	0.001	0.000	0.001
Table 9	Fraction stock spoiled	0.620	1.000	0.521	0.249
Table 9	Received bulk discount	0.700	1.000	0.141	0.087
Table 9	Production cost change	0.320	0.847	0.466	0.249
Table 9	Received loan	0.323	0.847	0.506	0.249
Table 9	Business bank account			0.010	0.015
Table 9	Inventory value	0.246	0.847	0.050	0.046
Table 9	Capital stock	0.953	1.000	0.003	0.007
Table 10	Own labor	0.957	1.000	0.000	0.001
Table 10	Self-efficacy	0.838	1.000	0.293	0.152
Table 10	Get Ahead attitudes	0.396	0.847	0.032	0.036
Table 10	Women's Association	0.870	1.000	0.002	0.006
Table 10	Number Discuss Business	0.388	0.847	0.037	0.039
Table 10	Works with others	0.154	0.544	0.195	0.102

Appendix 12: No Significant Impact on Market Prices or Input Costs

The businesses in our sample sell a diverse range of products, making overall analysis of price changes in the market difficult. Moreover, since many of these goods are not in standardized units, this prevents comparison of prices across firms. We therefore focus on the five most commonly sold products, which are sold by between 20.9 percent (tomatoes, price per single tomato) and 11.9 percent (price per kilogram of sugar) of firms. Firms were asked if they sell these products, and if so, the price they charge for a specified unit, and the purchase price they pay for these inputs. For example, the price for a gorogoro (2kg container) of maize. Appendix Table 9 estimates the treatment impacts on these prices. We see no significant treatment impacts or spillover impacts on these prices and unit costs. Moreover, the point estimates for the training group are positive for four out of five products. These results are consistent with firms not growing sales by lowering prices of commonly sold products, but instead by improving customer service and then increasing the variety of other products sold.

Appendix Table 9: Impacts on Key Market Prices and Unit Costs

	Maize	Kale	Sugar	Tomatoes	Beans
Panel A: Sale Price Charged					
Assigned to Training	-0.217 (0.952)	1.852 (2.509)	2.879 (3.671)	4.971 (4.766)	2.380 (3.283)
Spillover Group	0.941 (1.066)	-1.993 (1.941)	4.196 (3.297)	1.809 (4.829)	2.407 (3.636)
Mean of Pure Control Group	67.289	14.382	36.661	59.774	93.435
Sample Size	1042	1436	1171	1120	1170
Panel B: Unit Cost of Item					
Assigned to Training	-0.491 (0.788)	-0.888 (1.116)	2.664 (3.254)	2.976 (4.039)	1.807 (2.976)
Spillover Group	0.476 (0.911)	-1.584 (1.062)	3.523 (2.944)	1.255 (4.002)	2.462 (3.181)
Mean of Pure Control Group	56.179	8.547	30.053	48.368	69.409
Sample Size	1036	1420	1169	1115	1169

Notes: Robust standard errors in parentheses, clustered at the market level.

*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.

All regressions control for randomization strata fixed effects.

Maize and Beans are price per 2 kilogram container, Kale (Sukuma) is price per bunch, Sugar is price per kilogram, and tomatoes is price per single tomato. Prices and Costs truncated at the 1st and 99th percentiles.