

ONLINE APPENDIX

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*

American Economic Journal: Applied Economics

Appendix 1: Timeline

Appendix 2: Additional Details on Listing, Sample Selection and Qualitative Sample

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 and longer-term impacts

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 of Spillover Analysis

Appendix 11: Sharpened Q-values for Mechanism Impacts

Appendix 12: No Impact on Market Prices or Purchase Costs

Appendix 13: Heterogeneity by Prior Access to Credit

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 – Aug 2017

Appendix 2: Additional Details on Listing, Sample Selection, and Qualitative Sample

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.

Qualitative Sample and Methodology

ICRW conducted two rounds of qualitative work. The first took place in April 2014, corresponding to 6 months post-training. Its purpose was to focus on process-oriented research to understand how the training had been implemented, and to note any unexpected issues that we could ask about in the first-year follow-up survey. This work consisted of eight focus-group discussions with women entrepreneurs who had taken part in the training (N = 77) from the 4 counties, and 21 key informant interviews with organizations that implemented the training (n = 13), with the ILO (n = 1) and IPA (n = 1) staff, and professional training staff (n = 6). The second round of qualitative work took place in October and November 2014 (following the first-year follow-up survey), and was intended to understand the longer-term influence of the GET Ahead Training on women entrepreneurs' businesses and personal lives 12 – 16 months after the training. This qualitative work consisted of 42 in-depth interviews with women entrepreneurs (30 from the training treatment group and 12 from the spillover group of women who were in markets where others had received training). These respondents were randomly sampled after stratifying by location, age (above or below median age), education (above and below primary levels of education), and sector in order to provide a range of different types of women. They provide views of a random sample of the participants, and did not select on how much treatment had impacted the women.

The in-depth interviews provide the source of quotes used in this paper. These interviews asked women about their household, their business, their experience with the training, the challenges their business faced, and the nature of competition in their market. They were asked how the training may have affected their business, but the interviewer was careful to note that changes could be positive, negative, or no changes at all. ICRW (2015) then describes further their methodology for coding and interpreting the findings using Atlas-ti.

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

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.

Day 1	Day 2	Day 3	Day 4	Day 5
<ul style="list-style-type: none"> • Opening • Gender equality promotion: life cycle of people and enterprises • The business woman: she can do it 	<ul style="list-style-type: none"> • The business environment: she is not alone • Business ideas • Marketing 	<ul style="list-style-type: none"> • Production, services and technology • Marketing 	<ul style="list-style-type: none"> • Finance • Manage of self and others • Business support and networking 	<ul style="list-style-type: none"> • Management or networking • Action Planning • Closing

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

Appendix Table 4: 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 and Longer-term Impacts

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 – 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	<ul style="list-style-type: none"> • Introduction of mentor mentees • Definition of objectives • Understanding the mentoring agreement 	<ul style="list-style-type: none"> • 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 <ul style="list-style-type: none"> • 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 <ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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's outcomes • Agreeing on a sustainability plan after the end formal mentorship 	2 hrs

Appendix Table 5 uses the four-year market census, taken approximately two years after the mentoring intervention, to examine whether mentoring had differential effects from training alone over this longer time horizon. Column 1 shows that training alone increased the likelihood of firms surviving and being found in this market census, whereas there is no significant effect of mentoring. This is consistent with the three-year results. Then, examining last week's sales and profits, which condition on survival, we see the results depend on whether levels or logs are used. Using logs, we cannot reject equality of impacts of training alone and mentoring, with the coefficients on profits being very similar in magnitude to one another. In contrast, when we examine levels, the impacts are higher for mentoring than training alone, and significantly so in the case of sales.

Appendix Table 5: 4-Year Market-Census Impacts for Mentoring vs Training Alone

	Firm Operating and Interviewed	Last week's Sales		Last week's profits	
		Levels	Logs	Levels	Logs
Assigned to Mentoring	0.006 (0.024)	1381* (781)	0.265*** (0.085)	358* (195)	0.218*** (0.082)
Spillover Group to Mentoring	0.017 (0.027)	66 (681)	0.100 (0.075)	-13 (189)	0.042 (0.079)
Assigned to Training Alone	0.060*** (0.021)	-387 (629)	0.147* (0.078)	157 (169)	0.222*** (0.074)
Spillover Group to Training Alone	0.007 (0.025)	-229 (698)	0.136 (0.086)	-78 (172)	0.058 (0.077)
Mean of Pure Control Group	0.71	7757	8.32	2177	7.12
Sample Size	3104	2234	2128	2233	2113
P-value: Mentoring = Training Alone	0.044	0.022	0.204	0.335	0.961

Notes:

Regressions control for randomization strata fixed effects and are for the set of markets eligible for the mentoring intervention. Robust standard errors in parentheses, clustered at the market level.

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

Appendix Figure 5 plots the densities of log sales and log profits, which helps show where this difference is arising. We see the profits distributions are very similar, with the exception of the very top tail, where there is more mass among firms assigned to mentoring. This same increase in mass at the top tail for firms assigned to mentoring is also there for sales, along with a slight increase in where the peak of the distribution is. These results then suggest that mentoring may have helped a few firms to grow profits and sales at the top of the distribution, but had similar impacts for most firms.

Appendix Figure 5: Density Functions of Log Profits and Log Sales for Treatment Group by Assignment to Mentoring Status



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. Firms which move from one market to another are still coded as surviving.

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)¹, asked about the last week as a reference period since pre-testing found a weekly rather than monthly recall was easier for business owners to answer: “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 paid yourself. That is, what were the profits of your business during last week?”. This is 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.

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 their firms’ money (coded as 1 if only they have access); has some money which they have sole control over and can spend how they like; do not need anyone’s permission to visit a friend, to travel to sell a business asset, 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.

¹ De Mel, Suresh., McKenzie, David., Woodruff, Christopher., (2009). “Measuring microenterprise profits: Must we ask how the sausage is made?”, *Journal of Development Economics* 88(1): 19-31

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, 2017). 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 themselves as “very confident” in their ability to do (only asked in round 2). This includes entrepreneurial tasks like “estimate customer demand for a new product”, “persuade a bank to lend you money” and “identify good employees”.

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. Examples include “Even when my business is going well, I keep my eyes open in case I find a way to improve it”, “I don’t worry about where my business will be in the future – I just plan week to week based on what comes up” (negatively coded), and “My business provides about the same as others/is doing about the same as others, so there’s no need to make it better.” (negatively coded).

Appendix 7: Data Availability

Appendix Table 7 shows data availability by treatment status.

Appendix Table 7: 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 8:: 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

9.1 Robustness of Direct Effects to Conditioning on Survival

Appendix Table 9.1 shows we continue to find direct treatment effects on weekly sales, profits, and the aggregate index after three years even when we condition on survival, and that the spillover effects remain statistically insignificant.

Appendix Table 9.1: 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 year	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.

9.2 Sector-by-Sector Treatment Effects for Main Sectors

Appendix Table 9.2 shows treatment effects separately by sector for the main sectors, and shows that we cannot reject equality of treatment effects by sector, nor equality of spillover effects by sector.

Appendix Table 9.2: Impacts and Spillovers within the Experimental Sample at the Sector Level

	Fruit and Vegetables	Grain Sellers	Tailors	Grocery Stores	Restaurants	Other	p-value for testing equality across sectors
Panel A: Profits and Sales Index							
Assigned to Training* 3 Years	0.101** (0.042)	0.121 (0.097)	0.118 (0.073)	0.129 (0.106)	0.118 (0.089)	0.053 (0.060)	0.986
Spillover Group*3 Years	0.044 (0.043)	-0.157 (0.097)	0.110 (0.081)	-0.004 (0.107)	0.058 (0.068)	-0.007 (0.061)	0.459
Percent of Sample in Sector	32.5	11.0	10.2	10.1	9.6	27.2	
Control Group Mean	-0.109	0.158	-0.103	0.153	0.041	0.050	
Sample Size	4248	1444	1321	1294	1260	3445	
Panel B: Profits in the last week							
Assigned to Training* 3 Years	351*** (131)	126 (221)	312 (196)	-14 (330)	260 (229)	173 (160)	0.761
Spillover Group*3 Years	134 (128)	-351* (210)	-41 (190)	-399 (312)	297 (260)	-1 (140)	0.116
Control Group Mean	1199	1583	1233	1926	1503	1529	
Sample Size	4235	1438	1318	1284	1259	3434	
Panel C: Sales in the last week							
Assigned to Training* 3 Years	1260*** (457)	614 (1155)	887 (604)	1070 (1466)	674 (771)	783 (594)	0.960
Spillover Group*3 Years	599 (480)	-1605 (1158)	27 (533)	-372 (1411)	957 (867)	-198 (617)	0.252
Control Group Mean	4753	7849	3429	9378	5499	5795	
Sample Size	4245	1443	1317	1292	1259	3442	

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

Sector is defined as of baseline.

9.3 Were ineligible women in the same markets affected?

Appendix Table 9.3 uses our market censuses to measure whether having women trained in a market caused a change in the number of other women starting businesses in the market, the total number of other firms run by women in these markets, or the total profits and sales earned by these firms. The spillover effects are all positive in sign, suggesting these other women were not hurt by having some women trained.

Appendix Table 9.3: Spillover Impacts on Ineligible Women

	Ineligible Women			
	# New Entrants	Total Other Firms	Other Firm Profits	Other Firm Sales
Market Assigned to Training*1 Year	0.193 (0.509)	1.473 (1.375)	4439 (2940)	24509** (12197)
Market Assigned to Training *3 Years	0.833 (0.785)	1.290 (1.784)	2662 (3685)	9608 (13896)
Mean of Control Markets	8.59	22.48	34870	127470
Sample Size	301	301	301	301
P-value: 1 Year=3 Years	0.517	0.933	0.684	0.404

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 and survey round.

Data from market census conducted in round 2 (one year) and round 4 (three years) are for firms operated by women that are not in our experimental sample.

Appendix 10: Robustness of Spillover Analysis

Appendix Table 10.1 uses the four year market census to examine how our spillover analysis for women-only carried out in panels A and B of Table 8 would change if estimated on this longer period. In Panel A, our 3-year impact combined our round 4 and 5 follow-up rounds, and so had more data with which to estimate the impact than we have for the 4-year impact. We see treatment estimates which are similar in magnitude, but which are less precise. This is consistent with the continued treatment effect seen for treated firms found in Table 7. In Table B, the percentage of female-run firms that are trained falls from 28.1 percent at 3 years to 23.4 percent in our four-year follow-up, as some experimental firms close and other non-experimental firms enter. If there is no impact of treatment on entry, we would expect the market-level treatment effect to fall, and our point estimates are indeed lower and no longer significant. This highlights that the lack of significant impact in panel D of Table 8 when we include men does not just reflect the impact of including men, but also the difficulty of detecting an impact even among all women using just this single round four-year follow-up.

Appendix Table 10.1: Impacts on Market-Level Profits and Sales with 2017 Census

	Total Profits		Total Sales	
	Levels	Logs	Levels	Logs
Panel A: The Market of Women Running Comparable Businesses				
3-Year Impact shown in Table 8	2476*	0.079	13502**	0.165***
	(1495)	(0.060)	(6280)	(0.058)
Market Assigned to Training*4 Years	5290*	0.089	12013	0.065
	(3078)	(0.085)	(11865)	(0.095)
Control Market Mean at 4 Years	33953	10.27	124314	11.54
Control Market S.D. at 4 Years	20614	0.66	87971	0.70
Sample Size (Markets)	157	156	157	156
Percent of Volume Going to Trained Firms	58	58	57	57
Percent of Firms in Market Trained	55	55	55	55
Panel B: The Market of All Women-Run Businesses including New Entrants and Ineligibles				
3-Year Impact shown in Table 8	7093	0.114	31963	0.164**
	(5167)	(0.070)	(21417)	(0.073)
Market Assigned to Training*4 Years	5295	0.026	15033	0.032
	(6433)	(0.073)	(23526)	(0.078)
Control Market Mean at 4 Years	79795	11.15	291463	12.43
Control Market S.D. at 4 Years	44897	0.52	171207	0.55
Sample Size (Markets*Rounds)	157	157	157	157
Percent of Volume Going to Trained Firms	27	27	26	26
Percent of Firms in Market Trained	23	23	23	23

Notes:

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

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

Regressions control for randomization strata.

Appendix Table 10.2 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 10.2: 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.² Appendix Table 11 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 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.

² 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.

Appendix Table 11: P-values and Sharpened Q-values for Mechanisms

Appendix Table 11: Sharpened Q-values for Mechanisms

Table	Measure	1 Year		3 Years	
		P-value	Sharpened Q	P-value	Sharpened Q
Table 9	Business Knowledge	0.839	0.628		
Table 9	Business Practices	0.000	0.001	0.000	0.001
Table 9	Own labor	0.957	0.716	0.000	0.001
Table 9	Self-efficacy	0.838	0.628		
Table 9	Get Ahead attitudes	0.396	0.380		
Table 9	Number Discuss Business	0.388	0.380	0.037	0.077
Table 9	Works with others	0.154	0.203	0.195	0.243
Table 11	Weekly Customers	0.648	0.523	0.001	0.005
Table 11	Gained New Customer	0.082	0.131	0.040	0.078
Table 11	Lost New Customer	0.025	0.055	0.006	0.019
Table 11	Sales per customer	0.396	0.380	0.138	0.194
Table 11	Open set time			0.010	0.025
Table 11	Introduce new product	0.000	0.001	0.000	0.001
Table 12	Profit ratio	0.131	0.194	0.530	0.425
Table 12	Monitors Sales Trends	0.000	0.001	0.000	0.001
Table 12	Fraction stock spoiled	0.620	0.511	0.521	0.425
Table 12	Received bulk discount	0.700	0.529	0.141	0.194
Table 12	Production cost change	0.320	0.380	0.466	0.407
Table 12	Received loan	0.323	0.380	0.506	0.425
Table 12	Business bank account			0.010	0.025
Table 12	Inventory value	0.246	0.308	0.050	0.089
Table 12	Capital stock	0.953	0.716	0.003	0.011

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 products³, 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 12 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

³ Note, this differs from the main product sold asked at baseline. Firms were only asked whether they sell any of a list of common products in the three year follow-up survey. As a result, this analysis is conditional on an outcome which is itself potentially affected by treatment (products sold). It should therefore be considered suggestive only.

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. Nevertheless, the confidence intervals do allow for the possibility of modest price reductions (3 to 7 percent) for maize, tomatoes and beans; and are wider still for sugar (allowing up to a 11 percent drop) and kale (allowing up to a 21 percent drop).

Appendix Table 12: 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.

Appendix 13: Heterogeneity by Prior Access to Credit

Appendix Table 13 splits the sample according to whether the firm had access to credit at baseline or not. We see the treatment impacts emerge sooner for firms with this prior access to credit, and take time to emerge for firms without prior access to credit.

Appendix Table 13: Impacts by Prior Loan Access

	Had Bank or MFI Loan		Never had Loan	
	Weekly Profits	Weekly Sales	Weekly Profits	Weekly Sales
Assigned to Training*1 Year	311** (143)	1022* (613)	12 (74)	127 (269)
Assigned to Training*3 Years	413** (176)	1702** (753)	165* (95)	898** (346)
Mean of Pure Control Group	1719	7179	1348	5302
Sample Size	3204	3213	9677	9696
P-value: Training 1 Year = 3 Years	0.509	0.281	0.059	0.013

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, for the baseline of the outcome, survey round fixed effects, and for spillover groups.