The Direct and Spillover Impacts of a Business Training Program for Female Entrepreneurs in Kenya[*]

Preliminary

David McKenzie, *World Bank*

Susana Puerto, *ILO*

Executive Summary/Extended Abstract

This paper contains the interim findings of a randomized evaluation of the ILO's Get Ahead business training program for female microenterprise owners in four counties of Kenya. The study involves 3,537 women operating in 157 markets. Randomization took place first at the market level (93 treated markets, 64 control markets), and then at the individual level within the treated markets (1,172 treated women, 988 spillover controls, and 1,377 pure controls). This design enables us to measure any spillover effects of receiving training on other women operating in the same markets.

Training took place between June and November 2013. A long-form follow-up survey was conducted 12 months after training, and had a 90 percent response rate. This was followed by a shorter follow-up survey 3 months later which also had 90 percent response. Pooling the two surveys allows for follow-up data on 95 percent of the sample to give impacts 12-15 months after training.

The key results over this time frame are:

- There was high demand for training among those invited, with 77.7 percent of the treatment group attending at least one day, and 95 percent of those who attended one day staying for all 5 days of the training.
- Training had no impact on business knowledge questions requiring calculations, but led to a modest, but statistically significant, increase in the business practices used. The treatment effect was 5 percentage points, equivalent to the average trained firm using 1.3 more practices out of the 26 we measure. Training appears equally effective in terms of increasing business practices for women with low and high initial business skills.
- Almost 11 percent of firms close down over the period, and training has no impact on business survival.
- When we allow for spillover effects, we estimate that the receipt of training results in a 7-8 percent increase in weekly sales and profits, and a 0.06 standard deviation increase in an aggregate index of a variety of different sales and profits measures. This aggregate impact is statistically significant at the 10 percent level. There is some suggestive evidence that this is offset by a negative spillover on the sales and profits of untreated women operating in the same markets – sales per week for these women fall and are statistically significant at the 10 percent level, however we cannot reject that there are no spillovers when we look at the aggregate index.
- If we assume that there are no spillover effects (as has been standard in most of the literature to date), then the impacts on sales and profits are larger, and are both statistically significant at the 5 percent level. The impact is now a 14 percent increase in profits, and 18 percent increase in sales, or a 0.08 standard deviation increase in the aggregate index.
- This impact on sales and profits is equivalent to a US$1.30 to $2.47 per week increase in profits.

- Training has modest positive impacts on subjective well-being and mental health, results in modest increases in marketing and customer retention, but has no significant impacts on access to finance, costing, owner labor supply, self-efficacy, attitudes, female empowerment, or discussing work with other women. There is a small increase in membership in merry-go-rounds (roscas).
- An expectations elicitation exercise reveals these impacts to be substantially smaller than anticipated by the Get Ahead trainers, but for the most part in line with the expectations of the ILO and IPA project team collaborating on this study.

1. Introduction

Business training is one of the most common forms of active support provided by Governments, NGOs, and some international organizations, to small firms around the world. However, until recently there has been very little rigorous evidence as to the impacts of these programmes. A recent literature summarized in McKenzie and Woodruff (2014) has begun to measure these impacts, but have often suffered from low statistical power, and have been unable to measure the extent to which gains to trained firms come at the expense of other firms.

This impact evaluation aims to measure the causal impact of the International Labour Organization (ILO)'s *Gender and Enterprise Together* (GET ahead) business training program on 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. To do so, the evaluation uses a randomized control trial (RCT) methodology with a large sample, and with a two-level randomized experiment: randomized selection of villages, and of individuals within villages.

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

2.1 Selecting a Sample

The selection of the study areas was the result of a participatory process that involved the Technical Committee of the ILO Women Entrepreneurship and Economic Empowerment (WEDEE) project as well as other relevant stakeholders[1]. A Stakeholder retreat in October 2012 was used to pre-select 10 counties from the 47 counties in Kenya as possible locations for the study. A more detailed review of these 10 counties and consultations with the stakeholders were then used to select 4 counties in which to provide the ILO Gender and Entrepreneurship Together (GET Ahead) training: Kakamega and Kisii in the Western region, and Embu and Kitui in the Eastern region.

---

[1] Department of Micro and Small Enterprise Development (DMSED) of 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, Inoorero University.

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. The listing operation took place one county at a time between June 3, 2013 and November 1, 2013. 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 as detailed in the table below. 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. These 3,537 individuals were located in 157 separate markets.

2.2 Randomization Procedure

These individuals were then assigned to treatment and control in a two-stage process.

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. In Kakamega and Kisii the size strata were 26 or under, 27 to 30, 31+; in Embu the

size strata were <20 or 20+, and in Kitui the size strata varied within smaller geographic clusters between the two cuts used elsewhere. 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 (see Table 1 below).

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 the individuals within each strata to training. When the number of individuals in the strata was odd, the odd unit was also randomly assigned to training. This resulted in 1172 of the 2160 individuals in treated markets being assigned to treatment, and 988 to control (Table 1).

Table 1: Summary of Random Assignment

County	# of Strata	Total	Treatment	Control	Total	Treatment in Treated Market	Control in Treated Market	Control in Control Market
		Markets				**Individuals**		
Kakamega	9	**31**	19	12	**782**	257	220	305
Kisii	9	**34**	20	14	**844**	274	237	333
Embu	6	**36**	20	16	**715**	231	189	295
Kitui	11	**56**	34	22	**1196**	410	342	444
Total	**35**	**157**	**93**	**64**	**3537**	**1172**	**988**	**1377**

2.3 Characteristics of the Sample and Verification of Random Assignment

Table 2 provides some basic characteristics of firms at the market level for treatment and control markets. We see that randomization has succeeded in generating markets with comparable characteristics, with a test of joint orthogonality being unable to reject the null hypothesis that average characteristics of the markets are unrelated to treatment assignment. The typical marketplace in our study has 22 firms surveyed in it, and the average firm earns 1100 KSH (US$13) per week in profits, on sales of 4500 KSH. The typical market contains 75 percent of the study firms in retail, and 25 percent in services. The average market has 46 percent of the firms in it claiming to have a business license, and only 8.5% saying they have previously taken any form of training to help their business.

Table 2: Verification of Randomization at the Market Level

Means of	Treated Markets	Control Markets	Test of Equality p-value
Number of individuals in baseline	23.2	21.5	0.162
Mean weekly profits from census for baseline respondents	1133	1103	0.299
Mean weekly sales from census for baseline respondents	4648	4307	0.055
Share of firms in retail	0.75	0.75	0.934
Share of firms in services	0.25	0.25	0.959
Share of firms registered with city council	0.47	0.45	0.368
Share of firms with any form of previous business training	0.09	0.08	0.481
Test of joint orthogonality			0.403
Number of markets	93	64	

Table 3 provides some key characteristics of the women selected in our sample and their firms. Again the sample looks similar across the three groups on most characteristics. We can't reject the joint orthogonality of baseline characteristics when comparing the treatment group to the spillover group, or the treatment group to the pure control. Baseline sales are lower in the pure control group than the spillover group, and a joint orthogonality test has a p-value of 0.09 for this bivariate comparison. Since we are making three bivariate comparisons in testing joint orthogonality, a multiple testing adjustment would require this p-value to be lower for us to be concerned about imbalance.

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 baseline data reveals that women are operating a rather limited variety of businesses. The most common business types are selling fruit and vegetables, selling household goods, dressmaking, selling grains and cereals, and operating a food kiosk or small restaurant. The mean firm has capital stock of 31,000 KSH (US$370) and the median 10,500 KSH (US$124). One quarter have ever received financing from a bank or microfinance organization. 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.

Table 3: Individual Characteristics and Verification of Randomization at the Individual Level

	Treatment Group (1)	Spillover Group (2)	Pure Controls (3)	P-value (1) vs (2)	P-value (1) vs (3)	P-value (2) vs (3)
Age	36.0	35.6	35.7	0.482	0.454	0.690
Years of Education	8.92	8.91	9.09	0.910	0.569	0.515
Married	0.67	0.66	0.67	0.404	0.638	0.518
Household Size	4.97	4.85	4.85	0.188	0.262	0.499
Empowerment Index	6.93	6.94	7.01	0.907	0.348	0.569
Age of Firm	6.39	6.57	6.27	0.574	0.741	0.403
Number of Employees	0.27	0.27	0.27	0.747	0.989	0.953
Weekly Profits	1128	1140	1091	0.987	0.395	0.322
Weekly Sales	4517	4859	4182	0.148	0.085	0.001
Capital Stock	30571	34092	29370	0.248	0.863	0.101
Ever Received Bank or Microfinance Loan	0.24	0.25	0.24	0.891	0.482	0.934
Keeps Records	0.37	0.34	0.34	0.235	0.094	0.974
Business Practices Score	0.53	0.53	0.52	0.934	0.487	0.598
Joint orthogonality test p-value				0.810	0.620	0.090
Sample Size	1172	988	1377			

Notes:

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

3. The Get Ahead Training Program and TrainingAttendance

3.1 The Get Ahead Program

The training provided is the ILO's Gender and Entrepreneurship Together – GET Ahead for Women in Enterprise program. This program "differs from conventional business training materials by highlighting essential entrepreneurial skills from a gender perspective, whether applied to starting or improving an individual, family or group business. 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 and obtain support through groups, networks and institutions dealing with enterprise development" (Bauer et al, 2004). The program began in Thailand in 2001, and has now been used in X different countries.

An objective of the program is to create a "business mind" among low-income women engaged in small-scale businesses. The training methodology is participatory, with practical exercises to teach concepts. For example, women learn about the different types of costs involved in production, and how to account for their own costs through making lemonade; 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.

Topics covered included several gender concepts that tend not be emphasized in general business training programs such as: the difference between sex and gender, and the role of cultural constraints in shaping women in business; dividing household and business tasks; and how to network with other women and the role of women's associations. In addition, it covers a number of topics more typical of standard programs 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 course is a five-day course. 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. The median marketplace had a straight-line distance of 14.3 kilometers from the training location, with a 25-75 range of 8.2 to 23.2 kilometers.

The workshop lasts 5 days.

3.2 Training Costs

Based on the previous workshops, each workshop cost is KSH 566,415 equivalent to US$ 6,663. Since a workshop caters for 20-30 women, the cost per woman trained is therefore between US$222 and US$333.

3.3. Training Time Frame and Attendance

Training took place during the following dates:

- Kakamega: June 25-July 20, 2013
- Kisii: July 19-September 6, 2013
- Embu: September 16-October 11, 2013
- Kitui: October 28-November 29, 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 5 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 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.

4. Follow-up Surveys and Estimation Approach

4.1 Follow-up Surveys

Two rounds of follow-up surveys were conducted. The first was a comprehensive long-form survey that collected data on a wide range of business outcomes. This was fielded between June and October 2014, approximately one year after training had taken place in each county. A much shorter second follow-up was conducted between November 2014 and February 2015 in order to provide a second observation on volatile business outcomes like sales and profits.

Overall we were able to interview 90.0 percent of the sample in the long follow-up survey, and 89.7 percent in the short follow-up survey, with 95.0 percent of the sample getting interviewed in at least one of the two follow-ups. 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 99.2 percent of the sample in at least one of the two rounds.

Table 4 examines how data availability varies with treatment status. We see that those in the treatment group were 2 to 3 percentage points more likely to be interviewed than the pure control group or the spillover group. This is a relatively small level of differential reporting, and we will examine the robustness of our results to this difference through the construction of Lee bounds.

Table 4: Data Availability By Treatment Status

	Interviewed			Data on survival available		
	R2	R3	Either	R2	R3	Either
Assigned to Receive Training	0.034***	0.030***	0.023***	0.011	0.016**	0.003
	(0.012)	(0.011)	(0.008)	(0.007)	(0.006)	(0.004)
Spillover Group	0.014	-0.013	-0.003	0.014**	-0.003	0.001
	(0.013)	(0.014)	(0.010)	(0.007)	(0.009)	(0.004)
Pure Control Group Mean	0.886	0.889	0.943	0.967	0.960	0.990
Sample Size	3537	3537	3537	3537	3537	3537

Notes:
Robust standard errors in parentheses, clustered at the market level.
*, **, and *** denote significance at the 10, 5, and 1 percent levels respectively
R2 and R3 denote the long and short follow-up survey rounds respectively.

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 first follow-up. We had two independent field staff 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 at baseline and follow-up. Secondly, intensive qualitative work for this study was carried out by ICRW (2015), and we compare and contrast the quantitative results with their findings.

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.[2] This plan pre-

---

[2] http://www.socialscienceregistry.org/trials/287

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 to filing this plan, additional funding was obtained from PEDL, which enabled us to conduct short follow-up surveys to collect additional data on profits and sales immediately following the long-form follow-up surveys. This was originally planned in our original 3ie proposal, but had been cut back due to budget costs being higher than originally anticipated. Following McKenzie (2012) we pool this short follow-up data as an additional round together with the long-term data in order to increase power.

Our base specification is then to examine outcomes at the individual firm level by estimating the following ANCOVA specification for firm I in market j at time t:

$$Y_{i,j,,t} = \beta_0 + \beta_1 T_{i,j} + \beta_2 S_{i,j} + \pi Y_{i,j,t=0} + \gamma M_{ij,,t=0} + X'_{k,i,j}\theta + \delta_t + \varepsilon_{i,j,t=1} \quad (1)$$

Where $Y_{i,j,t}$ is the given outcome variable measured post-treatment, $Y_{i,j,t=0}$ is its baseline value and $M_{i,j,t=0}$ 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; we follow Bruhn and McKenzie (2009) in using X_k as a vector of randomization strata dummy variables (geographic region*market size*profit range in market census), δ_t is a survey round dummy, and $\varepsilon_{i,j,t}$ is the error term. β_1 provides the intent-to-treat effect, which is the effect of being assigned to treatment relative to being a firm in the control markets, while β_2 measures the spillover effect by comparing control firms in treated markets to control firms in control markets. The standard errors are then clustered at the market level to account for the market level random assignment.

In addition to the intention-to-treat effect, which measures the effect of being invited to training, we also estimate the local average treatment effect (LATE) of receiving training, by instrumenting training attendance with training assignment in equation (1). This LATE measure gives the effect of training for the individuals who would attend training when invited. None of the control group attended training, so the LATE is the same as the average treatment effect on the treated.

One of the main questions of interest in this study is whether there are spillovers from the treated individuals to other individuals 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). If there are no spillovers, we can also ask what is the impact of being assigned to training relative to not being selected for training within the same village? This will use the sample of treated firms in the treatment markets, and control firms in treatment markets, and estimate equation (1) without the spillover term. Since training was randomized at the individual level within villages, this does not require clustering of standard errors, and Huber-White standard errors will be used after controlling for the individual-level stratification dummies. This specification offers greater power, but requires the assumption (tested above) of no spillovers. Likewise we can also estimate the no-spillover LATE treatment effect, which is the effect of actually receiving training relative to not being selected for training within the same marketplace.

5. Results

5.1 Impacts on Survival

McKenzie and Woodruff (2015) find across a range of countries that small firms which have better business practices have higher likelihoods of survival. It is therefore of interest to see whether the training changed firm survival rates at all. The survival rate of the firms in the pure control markets is 89 percent, meaning that 11 percent of firms failed over the course of one year. Column 1 shows we find a rather precise zero effect of business training on survival rates. The ITT is 0.006, reflecting a statistically insignificant 0.6 percentage point higher survival rate for firms assigned to treatment. Likewise we find an insignificant spillover effect on survival: having other firms in your market receive business training does not affect your firm's survival over the one year period. The qualitative analysis contains no information about survival effects.

Our pre-analysis plan also hypothesized that there may be heterogeneous impacts on survival if training leads less skilled and less profitable business owners to realize they may be better off closing their businesses. To examine this, we test whether there is any treatment interaction with initial business practices or initial profits being below the median level. While firms with higher

baseline profits and better business practices are more likely to survive, we find no significant evidence of treatment heterogeneity in survival along either dimension (Appendix 1).

5.2 Impacts on Primary Outcomes: Firm Sales and Profits

Our primary outcomes are whether the business training succeeds in increasing 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.

We measure sales in the last day, as well as the last week, and also ask about sales of the main product. Profits in the last week are elicited via a direct recall question following the wording used in de Mel et al. (2009). In addition we also measure mark-up profits on the main good, by asking the sale price, unit cost, and number of units sold of the main product. As per our pre-analysis plan, we truncate all these variables at the 99th percentile to reduce the influence of outliers. Profits and sales are reported in terms of nominal Kenyan shillings. The inflation rate was 6.02 percent in 2014.[3]

Table 5 provides the results. In each column we begin by reporting the ITT and spillover effect from estimating equation (1). We then report the LATE or TOT effect of actually receiving training under this same assumption of potential spillovers. Panel B of the table then provides results estimated just from comparing treated to control firms within treated markets, under the assumption of no spillovers. We report both the ITT and TOT effects for this specification. Finally the foot of the table reports the implied percentage changes relative to the control mean implied by these treatment effects.

---

[3] Consumer price inflation from Kenya National Bureau of Statistics.

Table 5: Impacts on Primary Outcomes of Surival, Sales, and Profitability

	Firm Survival	Last Day Sales	Last Week Sales	Main Product Sales	Weekly Profits	Main Product Profits	Photo Value Inventories	Aggregate Index
Panel A: Results allowing for the possibility of Spillovers								
ITT effect of being assigned to training	0.006	182.024**	360.847	176.247	90.453	52.577	537.030	0.050*
	(0.009)	(75.247)	(282.554)	(311.445)	(74.944)	(94.843)	(340.532)	(0.026)
Spillover effect	0.003	36.031	-507.334*	175.013	-64.975	-27.098	402.309	-0.011
	(0.011)	(74.862)	(277.356)	(351.189)	(69.944)	(99.203)	(363.240)	(0.026)
TOT effect of actually receiving training	0.008	228.057**	451.966	220.298	113.185	65.845	651.686	0.062*
	(0.011)	(93.596)	(351.155)	(386.654)	(93.108)	(117.862)	(408.139)	(0.033)
Pure Control Mean	0.891	1232.042	6265.314	4344.540	1483.820	1343.171	4883.558	-0.009
Sample Size	6859	6532	6522	5755	6511	5682	2602	6532
Panel B: Results assuming no spillovers								
ITT effect of being assigned to training	0.003	154.963**	879.923***	-33.948	167.400**	81.676	194.167	0.062**
	(0.011)	(70.244)	(281.366)	(353.572)	(66.993)	(103.115)	(365.836)	(0.026)
TOT effect of actually receiving training	0.004	194.435**	1103.511***	-42.433	209.931**	102.386	235.888	0.078**
	(0.014)	(85.915)	(343.336)	(429.511)	(81.818)	(125.463)	(417.075)	(0.031)
Control Group Mean	0.898	1288.264	6048.651	4624.517	1481.987	1350.184	5571.446	0.003
Sample Size	4206	4017	4008	3535	4005	3501	1617	4014
Implied percentage Increase relatie to control mean from LATE								
Spillovers allowed		18.5	7.2	5.1	7.6	4.9	13.3	
Assuming no spillovers		15.1	18.2	-0.9	14.2	7.6	4.2	

Notes:

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

Standard errors clustered at the market level in panel A, and individual level in panel B.

All regressions include controls for the baseline outcome where available, for randomization strata, and for the survey round.

Regressions pool together the short and long follow-up surveys (except the photo value of inventories)

Aggregate index is the average of the standardized z-scores of columns 2 through 7.

Consider first the results for the effect of training in panel A, where we allow for the possibility of spillover effects. All of the point estimates are positive. The impact on sales in the last day is significant at the 5 percent level, but none of the other measures are individually significant. The main product sales and profits measures are noisier than the aggregate measures, and have more missing values due to some firms not being able to say how many units of their main product they had sold. Likewise we have less power for the photo-valued measure, since this was only measured in a single-follow-up. Aggregating all these measures together into an aggregate index yields a marginally significant increase in primary outcomes. The coefficient implies that being assigned to receive training leads to a 0.05 standard deviation increase in business performance. Thus while there appears to be an effect, this effect is relatively small.

The local average treatment effects imply percentage increases in the different measures that range from 4.9 to 18.5 percent, and are in the order of 7 percent for our two preferred measures: weekly sales and weekly profits. The LATE effect for the aggregate index is for a 0.06 standard deviation increase.

Consider next the evidence for spillover effects. Our two preferred measures of weekly profits and weekly sales both suggest negative spillover effects, with this effect significant at the 10 percent level for weekly sales. However, we find positive and insignificant spillover estimates for several of our other profits and sales measures. As a result, the aggregate index is small, with the point estimate being a 0.01 standard deviation reduction, which is not statistically significant. Therefore we are unable to reject that there are no spillover effects. 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 training, 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.

Since we cannot reject the null hypothesis of no spillover effects for most outcomes, we can also use the individual-level randomization within treated markets to obtain alternative measures of the treatment effects. These are shown in panel B of Table 5. Since randomization is then at the individual level, there is more statistical power to detect treatment effects. Although the point estimates are of similar magnitudes to those in panel A in most cases, we see more statistically significant results due to this increased power. In particular, the impact on weekly sales is now

significant at the 1 percent level, and the impacts on daily sales, weekly profits, and the overall aggregate index are all significant at the 5 percent level. The point estimates are slightly larger for our key outcomes (reflecting the negative point estimates on the spillovers), and imply a TOT effect of receiving training of a 14 to 18 percent increase in sales and profits, and a 0.078 standard deviation increase in the aggregate index.

The effect of receiving training on the level of weekly profits is estimated at 113 KSH if there are spillovers, and 210 KSH if there are not. At an exchange rate of 85 KSH per USD, this equates to an increase of $1.30 to $2.47 per week. At an estimated cost of $222-333 per person trained, these gains would need to last for at least 90 to 256 weeks (2 to 5 years) for the benefits of the program to start exceeding costs.

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. Appendix 2 shows the results, with the first two columns allowing for spillovers and the second two columns excluding them. We find no evidence of spillover effects, and ITT effects on profits and sales which suggest approximately a 10 percent increase in sales and profits. However, while these point estimates suggest similar magnitudes to our levels specifications, they are not statistically significant.

Figures 1 and 2 compare the cumulative distribution functions of the three groups for weekly profits and weekly sales. We see that there is not a lot of separation among the three groups, which is consistent with our modest treatment effects. However, the treatment group's CDF always lies on top of, or to the right of, the other two groups, suggesting equal or slightly larger profits and sales across the distribution. For approximately the bottom 70 percent of the distribution the treatment group and spillover group have very similar CDFs, with both separated from the pure control group. This suggests the possibility of positive spillovers at the bottom of the distribution. In contrast, in the top one-third of the distribution, the spillover group's CDF lies to the left of both the pure control group and the treatment group, suggesting possible negative spillovers at the top of the distribution.

Figure 1: Comparing the Follow-up Profits Distributions

Cumulative Distribution of Profits by Treatment Group

Weekly profits (Kenyan Shillings)

— Treatment
---- Pure Controls
.... Spillover Controls
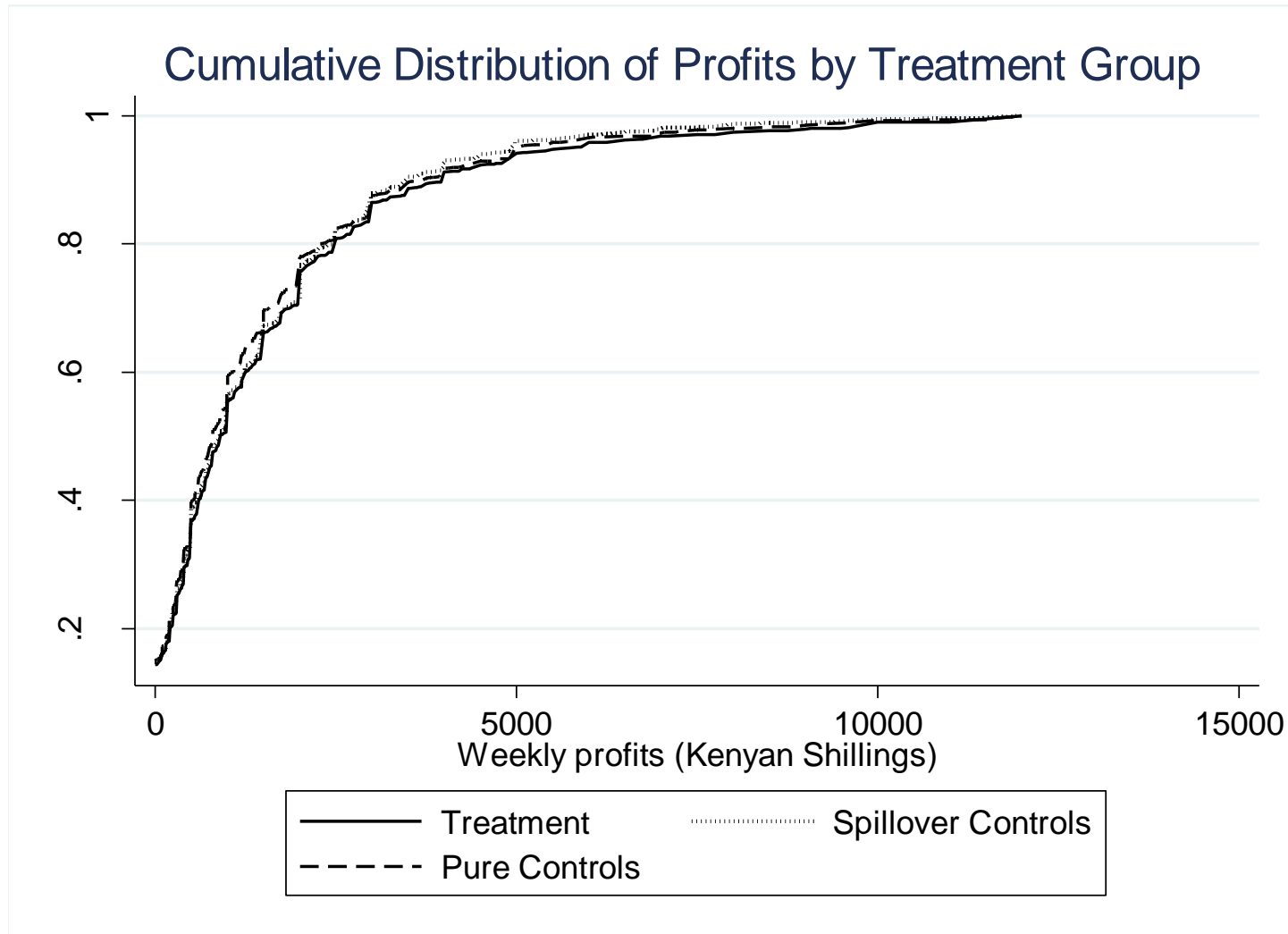
Figure 2: Comparing the Follow-up Sales Distributions

Cumulative Distribution of Sales by Treatment Group

Weekly sales (Kenyan Shillings)

Treatment   Spillover Controls
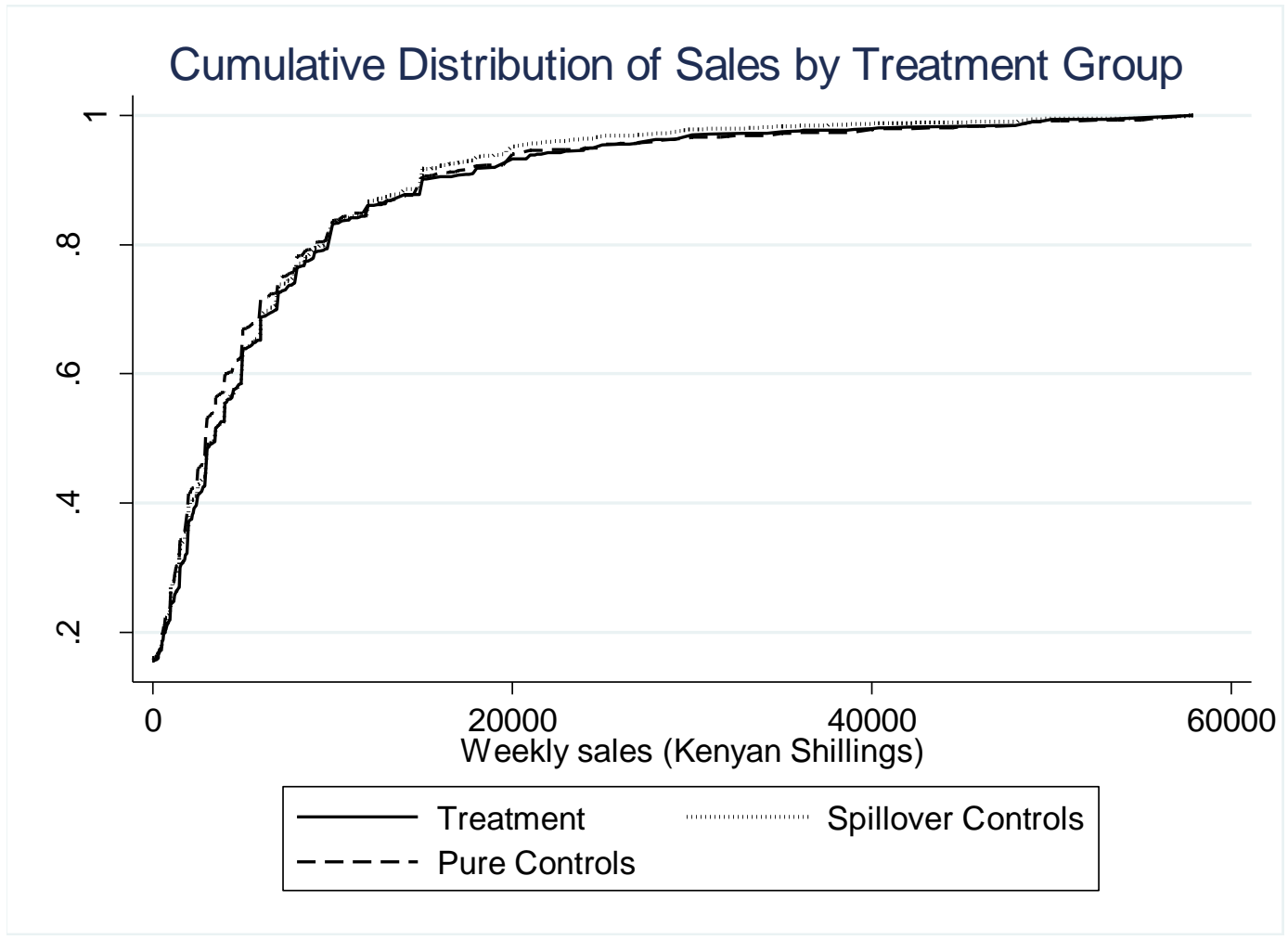Pure Controls

Table 6 explores this possibility of heterogeneity in the treatment effects further. Our pre-analysis plan hypothesized that the effect of training may differ with the initial skill level of the firm owner, with stronger effects for the more skilled if business training was a complement to existing skills and negative effects if it was a substitute for existing skills. Columns 1 to 3 test this with regard to three measures of baseline skills: having above median business practices, having at least 12 years of education, or having an above median digitspan recall. The treatment effect interaction is positive in all three cases, and is marginally significant for those with above median business practices at baseline. This provides some weak evidence that the training is more effective for those who already have a certain level of skills. However, when we consider the spillover effects, these are all negative, but not significant, for those with initially higher skills. This is consistent with the pattern shown in the CDFs, in which any spillovers seem to be more negative at the top of the distribution. One possible reason for this would be that control group women with very low levels of initial skills might be able to learn some very basic skills by observing and copying the treated women, whereas more skilled women would be already using these skills and would only experience spillovers through competition effects. However, these effects are not large in magnitude, and we cannot reject that there are no spillovers even for the more skilled group.

The last column of Table 6 was not pre-specified, but is motivated by the CDFs. It examines heterogeneity by whether or not the individual was in the top third of profits (1500 KSH or more per week) at baseline. Similar to the skills results, we see a positive interaction of treatment with high initial profits, and a negative spillover effect. However, again the results are not statistically significant. We also hypothesized that the treatment effects may vary with the level of competition faced by firms. Appendix 3 shows that there is no significant heterogeneity in either the treatment effects or spillover effects along these dimensions, with the point estimates varying in sign depending on the measure of competition used.[4]

---

[4] The final dimension of heterogeneity specified in the pre-analysis plan was with regard to baseline empowerment levels. We likewise find no significant heterogeneity in this dimension.

Table 6: Testing for Treatment Heterogeneity with Initial Skills

Dependent Variable: Weekly profits

Assigned to Training	-26.009	23.400	53.692	80.605
	(92.303)	(75.084)	(81.086)	(87.220)
Assigned to Training*High Baseline Business Practices	210.487*			
	(116.656)			
Assigned to Training*12 or more years schooling		223.441		
		(173.178)		
Assigned to Training*High Digitspan Recall			122.664	
			(146.000)	
Assigned to Training*Baseline Profits in Top Third				58.344
				(157.970)
Spillover Group	-44.097	-24.790	-64.209	-33.683
	(93.677)	(72.300)	(77.876)	(78.370)
Spillover Group*High Baseline Business Practices	-41.576			
	(108.377)			
Spillover Group*12 or more years schooling		-132.154		
		(153.513)		
Spillover Group*High Digitspan Recall			-23.852	
			(133.630)	
Spillover Group*Baseline Profits in Top Third				-47.316
				(149.485)
High Baseline Business Practices	314.815***			
	(73.442)			
12 or more years schooling		319.555**		
		(127.779)		
High Digitspan Recall			208.339*	
			(108.903)	
Baseline Profits in the Top Third				1032.613***
				(119.539)
Pure Control Group Mean	1484	1486	1486	1484
P-value for testing spillover effect zero for interacted group	0.308	0.274	0.463	0.552
Sample Size	6511	6497	6464	6511

Notes:

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

Standard errors clustered at the market level.

All regressions include controls for the baseline outcome, for randomization strata, the survey round and the survey round interacted with the heterogeneity variable of interest.

An alternative approach to testing whether the program has improved outcomes inclusive of any spillovers is to examine average outcomes at the level of the market. We do this by estimating the following equation at the level of the 157 markets:

Average profits in market j = a + b*Market j assigned to Training + d'X_j + e_j (2)

Where X_j are dummies for the market randomization strata (geographic region*market size), and average profits average the profits of all treatment and control firms in the treated markets, and of all control firms in the control markets.

Table 7 shows the results of estimating equation (2). We find that the overall effect on average profits in the markets is 6 KSH per week, which is not statistically different from zero. The average effect on market sales is negative and not statistically significant. These results are then consistent with any gains to the treated being offset by negative spillovers for those not treated.

Table 7: Market-Level Regressions

	Average Profits	Average Sales
Treatment Market	6.274	-173.560
	(84.080)	(364.238)
Short Follow-up Survey	-187.502***	-1083.273***
	(47.446)	(199.247)
Control Market Mean	1534.009	6436.600
Sample Size	314	314

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

There is always a concern with any business training program that the treatment affects how profits and sales are reported, regardless of any genuine effect. There are two sources of concern here. The first is that training may affect the accuracy of reporting. If businesses systematically under-estimate or over-estimate sales or profits, then any change in reporting would lead to bias in the measured treatment effects. In Appendix Table 4, we investigate whether training leads to more accurate reporting through measuring the treatment impact on the number of reporting errors made. We find no significant treatment effect, suggesting that training is not affecting reporting accuracy. A second concern is that participants may want to show gratitude for being given training and exhibit desirability bias, overstating how well their business is doing. If this were the case, we should expect to see more of a deviation between the photo-estimated level of inventories and what they report their inventory levels to be. Although the point estimate is positive, we cannot reject that there is no significant treatment effect on this difference (p=0.246). We therefore view these potential reporting concerns as unlikely to be major factors driving our results.

5.3 Impact on Secondary Outcomes: Employment, Empowerment, Subjective Well-Being, and Asset Ownership

Our pre-analysis plan specified four domains of secondary outcomes that were of interest for potentially showing impacts from the Get Ahead training. We estimate these impacts using equations (1) and (2), although for variables that were not measured in the short follow-up survey, we use only the one round of follow-up data. Table 8 provides the results.

The first set of secondary outcomes concerns employment, whether in self-employment or wage work. Column 1 of Table 8 shows that 90.3 percent of the pure control group are engaged in some form of employment for pay at the time of the follow-up surveys, and that there is no impact of treatment on this. Column 2 looks at all income from work, which combines profits from self-employment with any earnings from wage labor. We see a marginally significant increase in total labor income for those going through training, with a negative, but insignificant, spillover effect. If we then rely on the within-market randomization under the no spillovers assumption, panel B shows a highly significant impact on individual income. The increase is larger than the increase in profits alone seen in Table 5, suggesting the training may also have helped by increasing incomes of those going into wage work.

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 they need someone's permission to travel to sell a business asset, do they have money they have 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. 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' time 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 4.8 out of 10 on the life ladder currently, but expecting to be on step 8.1 in five years' time. Training increases both current and future subjective well-being by 0.3 steps, which is approximately 0.18 of a standard deviation. The impact on mental health is positive,

but only significant if we do not allow for spillovers, due to a positive but insignificant point estimate on the spillover variable.

Finally we examine the impact on ownership of household durable assets. There is a positive, but small and marginally significant impact when we allow for spillovers, and a positive, but statistically insignificant impact when we assume no spillovers. The standard deviation of the asset index is 1.5, so the estimated impact is 0.05 to 0.08 standard deviations, which is a small effect.

5.4 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.

5.4.1. Changes in Business Knowledge and the Use of Business Practices

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 is 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 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 9 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. (2011) 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 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 54 percent of these practices.

25

Table 8: Impacts on Secondary Outcomes

	Employed for pay	Income from all work	Empowerment Index	Ladder of Life Today	Ladder of Live 5 years	Mental Health	Household Durable Assets
Panel A: Results allowing for the possibility of Spillovers							
ITT effect of being assigned to training	0.009	178.104*	0.093	0.241***	0.246***	0.277*	0.097*
	(0.008)	(93.928)	(0.108)	(0.051)	(0.061)	(0.142)	(0.055)
Spillover effect	0.003	-109.441	-0.001	-0.027	-0.006	0.125	0.036
	(0.009)	(90.361)	(0.112)	(0.054)	(0.063)	(0.139)	(0.058)
TOT effect of actually receiving training	0.011	222.945*	0.114	0.300***	0.307***	0.340**	0.119*
	(0.010)	(116.587)	(0.131)	(0.063)	(0.075)	(0.172)	(0.067)
Mean of Pure Control Group	0.903	1665.846	6.979	4.793	8.132	18.909	-0.055
Sample Size	6859	6511	3059	6345	6344	3059	3014
Panel B: Results assuming no spillovers							
ITT effect of being assigned to training	0.004	317.302***	0.106	0.260***	0.248***	0.128	0.067
	(0.010)	(84.467)	(0.097)	(0.054)	(0.055)	(0.163)	(0.061)
TOT effect of actually receiving training	0.006	397.898***	0.131	0.325***	0.309***	0.157	0.082
	(0.013)	(103.080)	(0.114)	(0.066)	(0.066)	(0.190)	(0.071)
Control Group Mean	0.910	1590.949	7.004	4.790	8.140	19.056	0.011
Sample Size	4206	4005	1888	3907	3906	1888	1862

Notes:

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

Standard errors clustered at the market level in panel A, and individual level in panel B.

All regressions include controls for the baseline outcome where available, for randomization strata, and for the survey round.

Empowerment index is the sum of 10 indicators, with a higher score indicating higher empowerment.

Ladder of life today and in five years are Cantril Ladder questions on subjective well-being, where 10 is the highest well-being.

Mental health is measured by the MHI-5 index, coded so a higher score represents better mental health.

Household Durable assets is the first principal component of ownership of 12 durable assets.

Column 2 shows that the impact of being invited to training is an increase in the use of business practices. However, although the treatment effect is statistically significant at the 1 percent level, the magnitude of the effect is not that large: treated firms have a 0.05 increase in the proportion of practices used. This is approximately a 10 percent increase on the control mean, or equivalently that treated firms are using 1.3 more practices out of the 26 than the pure control group. The LATE estimate of the effect of actually receiving training is 0.06. So the training lead to only a relatively small increase in business practices. If we take this increase in business practices and the implied associations between business practices and business outcomes in McKenzie and Woodruff (2015), this would lead us to predict a 4 to 6 percent LATE for the increase in sales and profits – which is very similar to the treatment effects seen at the foot of Table 5.

Table 9 also provides marginally significant evidence of a positive spillover in business practices from the treatment group to untreated firms in the same marketplaces. However, the magnitude of the spillover is very small: 0.01 is only one-fifth the size of the treatment effect, and is equivalent to one in four control firms in treated markets utilizing one practice more out of 26.

In the remainder of Table 9 we examine heterogeneity in the business practice impacts according to baseline level of business practices and the education and digitspan recall of the firm owners. Although individuals with higher baseline practices and more schooling are likely to use better business practices at follow-up, we see no differential effect of either the treatment or the spillover with these human capital variables. This suggests that the treatment is having similar effects for both those with low initial business skills as it does for those with higher initial business skills.

Table 9: Impacts on Business Knowledge and Practices

	Business Knowledge	Business Practices	Business Practices	Business Practices	Business Practices
ITT effect of being assigned to training	-0.018	0.050***	0.054***	0.051***	0.047***
	(0.089)	(0.008)	(0.012)	(0.009)	(0.009)
Spillover effect	0.124	0.015*	0.008	0.016	0.015
	(0.091)	(0.008)	(0.012)	(0.010)	(0.009)
Baseline Business Practices Score		0.312***	0.321***	0.290***	0.307***
		(0.017)	(0.025)	(0.017)	(0.017)
High Baseline Business Practices*Treatment			-0.008		
			(0.016)		
High Baseline Business Practices*Spillover			0.012		
			(0.015)		
High Baseline Business Practices			-0.005		
			(0.013)		
At least 12 years schooling*Treatment				-0.006	
				(0.015)	
At least 12 years schooling*Spillovers				-0.002	
				(0.017)	
At least 12 years schooling				0.049***	
				(0.010)	
High Digitspan Recall*Treatment					0.009
					(0.015)
High Digitspan Recall*Spillover					0.000
					(0.017)
High Digitspan Recall					0.012
					(0.011)
Mean of Control Markets	2.000	0.536	0.536	0.536	0.536
Sample Size	3059	2860	2860	2853	2840

Notes:
*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.
Standard errors clustered at the market level.
All regressions include controls for the baseline outcome where available, and for randomization strata.
Business Knowledge is the Number of Questions out of 7 answered correctly
Business Practices is the Proportion of 26 business practices used by the firm
Round 2 data only used. Data on business practices are only for firms which survive.

5.4.2 Impact on Intermediate Channels of Customers and Reaction to Down Times

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 10. The women in our sample typically serve a lot of customers, with a median of 40 and mean of 80 customers per week in the control markets (after

top-coding at the 99th percentile to reduce the influence of outliers). Column 1 shows a positive treatment effect and negative spillover effect, but neither are statistically significant. Columns 2 and 3 show the treatment does lead firms to be 2 percentage points more likely to gain a new customer during the last 3 months, and 4 percentage points less likely to lose one, with no significant spillover effects. There is thus some evidence for the treatment allowing firms to access more customers. The qualitative surveys found mention of this channel, as indicated by the quote ""There are those who have improved…There are those who didn't know how to attract customers, but now I can see they have been able to attract customers…They are talking to them nicely unlike before where they would talk rudely" (ICRW, 2015).

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 4 and 5 of Table 10. We see that treatment does have a positive effect on the ratio of business profits in the worst week of the year to the current week, and does lead businesses to be more likely to regularly use business records to know if sales of a particular product are increasing or decreasing. Both impacts are statistically significant, with the impact on using records to monitor changes moderately large relative to the control group mean (a 0.11 effect relative to a base level of 0.30).

Table 10: Impact on Intermediate Channels: Customers and Reaction to Down Periods

	Number of Customers in past week	Gained a new customer in last 3 months	Lost a regular customer in last 3 months	Ratio of worst to current week profits	Uses records to see if sales increasing or decreasing
ITT effect of being assigned to training	1.650	0.021*	-0.042***	0.026**	0.111***
	(3.634)	(0.013)	(0.016)	(0.012)	(0.022)
Spillover effect	-2.179	0.007	-0.024	-0.004	0.029
	(3.713)	(0.013)	(0.017)	(0.010)	(0.022)
Mean of Control Markets	79.845	0.808	0.765	0.426	0.297
Sample Size	6484	6555	6556	2752	3203

Notes:

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

Standard errors clustered at the market level.

All regressions include controls for the baseline outcome where available, and for randomization strata.

Note: columns 4 and 5 are available only for the long follow-up survey.

5.4.3 Impacts on Finance, Asset Levels, and Costing

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 1of Table 11 shows that access to finance is limited in our sample, with only 15.7 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).

Table 11: Impacts on Finance, Assets, and Costing

	Received a loan from bank or microfinance	Value of inventories	Value of Capital Stock	Fraction of stock lost due to spoilage	Received a bulk discount when buying materials	Percent change in cost of production
ITT effect of being assigned to training	0.015	3793.967**	-126.811	-0.008	0.009	20.015
	(0.015)	(1689.347)	(815.649)	(0.013)	(0.022)	(15.485)
Spillover effect	0.011	-1437.159	-857.084	0.037**	-0.028	12.153
	(0.015)	(1461.881)	(871.903)	(0.016)	(0.023)	(17.218)
Mean of Control Markets	0.157	19855.299	14018.137	0.120	0.499	128.797
Sample Size	2860	3179	3194	2628	3025	2828

Notes:
*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.
Standard errors clustered at the market level.
All regressions include controls for the baseline outcome where available, and for randomization strata.
Data are only from the long follow-up survey

Columns 2 and 3 look at whether firms have grown in size according to their inventory levels and capital stock. The value of inventories in the treatment group is 19 percent higher than the control mean, which is consistent with them having more to sell. The spillover effect is negative, but not statistically significant. Capital stock in terms of tools, machinery, equipment, furniture, and vehicles is not statistically different, which aligns with the other evidence suggesting the impact

on business size has been relatively modest, and that firms have not been borrowing to grow or change scale.

Columns 4, 5, and 6 of Table 11 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 receipt of bulk discounts in purchasing, or in the cost of producing the firm's most profitable item. The latter is measured with considerable noise, so the insignificant point estimate admits a wide confidence interval. There is a statistically significant spillover effect, whereby control women are more likely to lose stock due to spoilage. This effect does not survive any standard correction for multiple hypothesis testing across the set of outcomes examined in this table.

5.4.4 Impacts on Owner Hours, Attitudes, and Social Capital

The final set of intermediate outcomes and mechanisms are examined in Table 12. 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 the quote "Then I used to open [my business] any time I wished…I would open much later…but these days it's better since I constantly open at nine and close at night at around eight. Those days I just used 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 I am convinced that it's time to leave." (ICRW, 2015). This qualitative finding is not supported by the survey data. The average owner reports working 49 hours per week in the business, with an insignificant 0.002 treatment effect. There is a marginally significant negative spillover effect, suggesting the control women may be working slightly less in treated markets.

The training was intended to also increase the confidence of women in their ability to perform business tasks. The qualitative work contains quotes from several women saying they are more confident as a result of the training: "My confidence has changed because those things I lacked, now I have them" (ICRW, 2015). We measure entrepreneurial self-efficacy through 10 questions that measure the owner's confidence in their ability to perform key business activities such as coming up with ideas for 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.

Table 12: Impact on Owner Hours, Attitudes, and Social Capital

	Owner labor hours	Entrepreneurial Self-efficacy	Get Ahead Attitudes	Belong to a women's association	Number of other women they discuss business with	Works together with other women to get discount or purchase	Belongs to a merry-go-round
ITT effect of being assigned to training	0.002	-0.037	-0.018	0.002	0.249	0.032	0.033**
	(1.053)	(0.181)	(0.021)	(0.017)	(0.184)	(0.024)	(0.014)
Spillover effect	-2.102*	-0.180	-0.047**	0.027	0.106	-0.004	0.004
	(1.109)	(0.171)	(0.022)	(0.019)	(0.199)	(0.025)	(0.016)
Mean of Control Markets	49.353	4.457	3.802	0.151	4.157	0.506	0.845
Sample Size	3205	3059	3059	2858	2856	2859	2860

Notes:
*, **, and *** indicate significance at the 10, 5, and 1 percent levels respectively.
Standard errors clustered at the market level.
All regressions include controls for the baseline outcome where available, and for randomization strata.
Data are only from the long follow-up survey

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 of Table 12 shows no significant effect of treatment on these. We do see a negative and significant spillover effect, although this does not survive correcting for multiple hypothesis testing.

The training also emphasized and encouraging cooperating with other women in the marketplace. The remaining columns of Table 12 examine aspects of this. Columns 4, 5, and 6 are our pre-specified measures, which look at membership of women's associations, discussing business with other women in the market, and 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, nor is there evidence of spillovers. The qualitative work indicated the importance of merry-go-rounds as a forum for women to share access to finance, and exchange ideas, and suggested that several women had joined merry-go-rounds following training. We therefore also examine this outcome. Membership of these groups is very prevalent, with 84.5 percent of the pure control group belonging to at least one merry-go-round. Being invited to training leads to a 3 percentage point increase in membership of these groups.

5.5 Comparison to Expectations

To see how these estimated impacts of the program compare to the expectations of those involved in providing the program, we follow the approach of Hirshleifer et al. (2015) in eliciting the expectations of key policy participants. We asked five of the Get Ahead trainers, the ILO project team, and the IPA field associate to provide their subjective expectations of the impact of the program and of the survival rates and business practice adoption rates of firms. This was done in September 2014, after the completion of training, but prior to any of these individuals receiving any analysis of the actual program impacts.

Table 13 compares these expectations to our treatment estimates. The Get Ahead trainers are strongly over-optimistic about the likely effects of the program. They estimate that being invited to training would have large impacts on survival, business practice adoption, sales, and profits, and also result in relatively large positive spillovers for untreated individuals in the same markets. The expectations of the ILO project staff and IPA field associate are much closer to our estimated treatment effects, with the exception of their expectations about the impact on business practice adoption – they expected four times the increase on business practices that actually occurred.

Table 13: Comparison of Results to Expectations of Program Implementors

	Get Ahead Trainers	ILO & IPA Project staff	Actual Rates and Impacts
Expected closure rate for trained women	5%	11%	9%
Expected treatment impact on survival	24%	3%	0.6%
Expected business practices score for trained women	0.68	0.63	0.58
Expected treatment impact on business practices	0.33	0.21	0.05
Expected treatment impact on sales	48%	12%	6 to 15%
Expected treatment impact on profits	41%	12%	6 to 11%
Expected spillover effect on sales of untreated	23%	4%	-8%

Notes: range given for actual rates of sales and profits includes estimates with and without allowing for spillovers.

References

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.

Bruhn, Miriam and David McKenzie (2009) "In Pursuit of Balance: Randomization in Practice in Development Field Experiments", *American Economic Journal: Applied Economics*, 1(4): 200-232.

Carpena, Fenella, Shawn Cole, Jeremy Shapiro and Bilal Zia (2011) "Unpacking the Causal Chain of Financial Literacy", World Bank Policy Research Working Paper no. 5798.

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

Diwan, Faizan, Grace Makana, David McKenzie, and Silvia Paruzzolo (2014) "Invitation Choice Structure Has No Impact on Attendance in a Female Business Training Program in Kenya" *PLOS ONE*, 9(10): e109873

Hirshleifer, Sarojini, David McKenzie, Rita Almeida, and Cristobal Ridao-Cano (2015) "The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey", *Economic Journal*, forthcoming.

ICRW (2015) ""I am determined, no matter how hard it becomes, I will not give up": A qualitative assessment of the ILO's GET Ahead Business Training Program", ICRW 2015.

McKenzie, David (2012) "Beyond Baseline and Follow-up: The Case for More T in Experiments", *Journal of Development Economics* 99(2): 210-221.

McKenzie, David and Christopher Woodruff (2015) "Business Practices in Small Firms in Developing Countries", Mimeo. World Bank

McKenzie, David and Christopher Woodruff (2014) "What are we learning from business training evaluations around the developing world?", *World Bank Research Observer*, 29(1): 48-82.

Veit, C.T. & Ware, Jnr, J.E. (1983). "The structure of psychological distress and well-being in general populations", *Journal of Consulting and Clinical Psychology*, 51, 730-742

Appendix 1: No Treatment Heterogeneity in Survival by Pre-specified Variables

	Survival	Survival
Low baseline profits*Assigned to Training	-0.007	
	(0.021)	
Low baseline profits*Spillover Group	-0.000	
	(0.023)	
Low baseline profits	-0.042***	
	(0.014)	
High baseline management*Assigned to Training		-0.010
		(0.020)
High baseline management*Spillover Group		-0.017
		(0.022)
High baseline management		0.023*
		(0.014)
Mean of Pure Control Group		
Sample Size	6859	6859

Notes:

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

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

Appendix 2: Inverse-Hyperbolic Sine of Profits and Sales

	Sales	Profits	Sales	Profits
ITT effect of being assigned to training	0.128	0.103	0.120	0.095
	(0.112)	(0.098)	(0.124)	(0.105)
Spillover effect	-0.011	0.000	excluded	excluded
	(0.121)	(0.102)		
Mean of Pure Control Group	7.544	6.479	7.606	6.539
Sample Size	6522	6511	4008	4005

Notes:

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

Standard errors clustered at the market level in columns 1 and 2, and individual level in columns 3 and 4.

All regressions include controls for the baseline outcome, for randomization strata, and for the survey round.

Regressions pool together the short and long follow-up surveys.

	Weekly Sales			Weekly Profits		
Assigned to Training	287.536	141.378	261.954	129.894	47.678	43.917
	(327.261)	(417.027)	(484.047)	(81.234)	(106.655)	(128.343)
Assigned to Training*Fewer than 5 competitors	387.045			-235.180		
	(779.936)			(181.998)		
Assigned to Training*One of Named Individuals is Competitor		424.589			85.938	
		(608.463)			(140.318)	
Assigned to Training*Proportion of Sample in Their Industry			523.561			236.950
			(1782.205)			(462.905)
Spillover Group	-573.860*	-317.349	-1058.022'	-55.800	-70.300	-229.853*
	(306.585)	(374.164)	(469.895)	(75.361)	(92.305)	(125.933)
Spillover Group*Fewer than 5 competitors	353.522			-32.558		
	(796.412)			(176.043)		
Spillover Group*One of Named Individuals is Competitor		-373.057			12.223	
		(573.479)			(137.671)	
Spillover Group*Proportion of Sample in Their Industry			2397.793			725.951
			(1700.432)			(455.494)
Fewer than 5 competitors in the market	-538.651			215.764		
	(610.830)			(134.939)		
At least 1 of 4 named people was a competitor		-344.636			-60.720	
		(442.065)			(100.305)	
Proportion of Sample in Market in Their Industry			-1993.425			-461.423
			(1404.463)			(406.520)
Pure Control Mean	6265.314	6265.314	6265.314	1483.820	1483.820	1483.820
Sample Size	6522	6522	6522	6511	6511	6511

Notes:

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

Standard errors clustered at the market level.

All regressions include controls for the baseline outcome, for randomization strata, the survey round and the survey round interacted with the heterogeneity variable of interest.

Appendix 4: Impact on Reporting Errors

Dependent Variable: Total Number of Reporting Errors Out of 6 Possible

ITT effect of being assigned to training	-0.010
	(0.023)
Spillover effect	0.001
	(0.027)
Short Follow-up Survey	-0.027
	(0.020)
Mean of Pure Control Group	0.505
Sample Size	6144

Robust standard errors in parentheses, clustered at the market level

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

Note: Reporting errors are the sum of the following errors made in reporting: 1) Revenues in last week from main product exceed total revenues in the last week; 2) Profits in the last week from main product exceed total profits in the last week; 3) Total profits in the worst week of the year are higher than total profits in the last week; 4) Profits in

the last week exceed sales in the last week; 5) Total sales yesterday exceed total sales in the last week; and 6) The cost of raw materials used to produce one unit exceeds the price charged per unit for the most profitable item.