Do Gender and Business Trainings Affect Business Outcomes? Experimental Evidence from Vietnam

Bulte, Erwin; Lensink, Robert; Nhung Vu, [No Value]

Published in:
Management Science

DOI:
10.1287/mnsc.2016.2472

Publication date:
2017

Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Do Gender and Business Trainings Affect Business Outcomes? Experimental Evidence from Vietnam

Erwin Bulte, Robert Lensink, Nhung Vu

To cite this article:
Erwin Bulte, Robert Lensink, Nhung Vu (2017) Do Gender and Business Trainings Affect Business Outcomes? Experimental Evidence from Vietnam. Management Science 63(9):2885-2902. https://doi.org/10.1287/mnsc.2016.2472

http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 63, No. 9, September 2017, pp. 2885–2902
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Do Gender and Business Trainings Affect Business Outcomes? Experimental Evidence from Vietnam

Erwin Bulte,[a] **Robert Lensink,**[b,a] **Nhung Vu**[c]

[a] Development Economics Group, Wageningen University, 6706 KN Wageningen, Netherlands; [b] Department of Economics, Econometrics and Finance, University of Groningen, 9700 AV Groningen, Netherlands; [c] School of Economics and Business Administration, Can Tho University, Can Tho, Vietnam

Contact: erwin.bulte@wur.nl (EB); b.w.lensink@rug.nl (RL); vthnhung@ctu.edu.vn (NV)

Received: May 1, 2015
Revised: October 13, 2015; December 15, 2015
Accepted: January 6, 2016
Published Online in Articles in Advance: August 11, 2016

https://doi.org/10.1287/mnsc.2016.2472

Abstract. We use a randomized control trial to evaluate the impact of a business training for female clients of a microfinance institution in northern Vietnam, and we consider the impact on (i) business knowledge, (ii) practices, and (ii) outcomes, as well as (iv) firm entry and exit decisions. In addition, we vary the nature of the intervention by inviting husbands to participate in the trainings for a subsample of our respondents. To gauge both short-term and medium-term effects, we combine data from two separate postintervention surveys. We find evidence of economically substantive impacts on knowledge, practices, and outcomes, and on the extensive margin (entry and exit). We also document that it takes time for the "downstream" outcomes of the trainings to materialize; although we find evidence of medium-term effects, no such evidence exists for the short term. Inviting husbands to participate in the trainings does not affect any of our knowledge or practice measures, but we document weak evidence for differential impact on (agricultural) sales and profits.

History: Accepted by Uri Gneezy, behavioral economics.
Funding: The authors thank the International Initiative for Impact Evaluation (3ie) [OW3.1132: Business training services] and the Netherlands Organisation for Scientific Research [N.W.O. Grant 453-10-001] for financial support, and Rosine van Velzen for excellent research assistance.
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mnsc.2016.2472.

Keywords: microfinance • finance-plus • evaluation of training interventions

1. Introduction

Not so long ago, microfinance was celebrated as one of the more promising development tools. This situation has changed in recent years. New theories question the usefulness of microcredit to raise incomes of poor people (e.g., Banerjee 2013), and empirical evidence tends to reject the hypothesis that simply providing access to capital to the poor will lift them out of poverty. For example, Banerjee et al. (2015) summarize the results from randomized controlled trials (RCTs) in six countries on four continents (Bosnia, Ethiopia, India, Mexico, Morocco and Mongolia). While they document some (mixed and generally weak) evidence of positive effects of microcredit on various relevant outcomes—including entrepreneurial activity, business size, and female empowerment—they conclude microcredit does not have "transformative effects" and generally fails to help the poor to raise their incomes above subsistence levels. The conclusion that lack of access to capital may not be the key impediment to firm growth is consistent with evidence from other countries, including Sri Lanka (de Mel et al. 2008, 2009), Ghana (Fafchamps et. al. 2011), and Tanzania (Berge et al. 2014). Interestingly, the impact of access to microfinance appears to be heterogeneous. Various researchers observe that impact varies with gender (Berge et al. 2014, de Mel et al. 2009, Giné and Mansuri 2014, Bruhn and Zia 2013), and others have pointed to the importance of human capital as a key complement of financial capital (e.g., Bloom et al. 2010, Bruhn et al. 2010, Sayinzoga et al. 2016).

In response, many microfinance institutions (MFIs) have embraced business development and financial literacy trainings as a key component of their expansion strategy. This is occasionally referred to as the finance-plus strategy. Evaluating the impact of such trainings is complicated by selection effects. Demand-side bias emerges when better (or worse) entrepreneurs self-select into treatment, or when individuals who know they stand to gain most from participating in a training choose to participate. Supply-side bias arises when training agencies target the best entrepreneurs, focus on the most promising sectors, or choose to operate in regions with the most suitable geographical conditions. Likewise, bias emerges when they target the least promising individuals, for distributional reasons. While the evidentiary base is still thin, several researchers have adopted an RCT approach to

tackle these concerns. The mounting evidence, however, remains ambiguous and mixed (e.g., Berge et al. 2014, Bjorvatn and Tungodden 2010, Giné and Mansuri 2014, Karlan and Valdivia 2011).[1] Most studies conclude that especially women do not gain much from attending business trainings.

This emerging literature is summarized by McKenzie and Woodruff (2014). While most studies confirm that training programs affect knowledge levels and/or business practices, there is much weaker evidence for the hypothesis that it also affects business outcomes. Berge et al. (2014), for instance, find that business outcomes only improve for men (and not for women),[2] and only if a training is combined with a grant. Some evaluations, but not all, find that trainings affect firm survivorship or start-up rates. Evidence for the hypothesis that trainings affect firm profits or entrepreneurial income is weakest. However, McKenzie and Woodruff (2014) also point to various shortcomings of existing studies, compromising the ability of these studies to find impact. For example, most studies tend to suffer from low statistical power due to small sample sizes combined with highly variable outcome variables and heterogeneous firms. Most studies also focus on impacts in the "short term"; often endline data are collected within months after completing the training. Arguably, many "downstream effects" will materialize later. They also argue in favor of analyzing improved outcome measures, and they propose that future analyses should seek to test which elements of content matter most for transforming the lives of micro-entrepreneurs.

In this paper we evaluate the impact of a gender and business training provided to female clients of the Tao Yeu May (TYM) Fund, a large microfinance institution in the north of Vietnam. We use an RCT design to measure effects on business knowledge, business practices, and business results. We also estimate the impact of the intervention on business entry and business exit. Minding the various challenges to the interpretation of results in earlier studies, we tried to use a sample that is sufficiently large to detect relatively modest effect sizes (but see concerns about noncompliance discussed in Section 3), and we measured impact twice after completing the training (to gauge the "dynamics" of impact; see also Berge et al. 2012).[3] We use sales and profits as outcome measures, and we consider both farm and nonfarm economic activities. We also try to probe the importance of varying the content of the training by introducing an innovation. Specifically, and responding to recent suggestions in the literature, for a random subsample of centres we asked women to bring their husbands along to participate in the training (e.g., Rahman et al. 2011, World Bank 2011). This allows us to test the auxiliary hypothesis that the gender composition of the group matters.

We incentivized men to participate in the trainings, but did not expect all men to actually participate. Nevertheless, the participation of (some) men was expected to raise the quality of the trainings, because they could share their knowledge with female clients, changing the nature and depth of discussions about the training material. But of course impact could materialize via additional channels as well. For example, participating men may be affected by the training themselves (e.g., if the gender module alters their outlook on gender inequality). Many businesswomen are held back by time constraints and have limited freedom to move around, partly because they are still responsible for the bulk of the housework (Giné and Mansuri 2014, Berge et al. 2014). Many businesswomen also mention that key (household and business) decisions are made by their spouses. When men participate in trainings, such external constraints may be relaxed—perhaps improving female business outcomes. However, men may also dominate the training, decreasing female attendance or opportunities for learning, in which case the attendance of husbands could negatively affect impact for women. Ultimately, the net impact is an empirical matter that we seek to gauge.

The paper's contribution is fourfold. First, we document positive impacts of a training intervention for female respondents. We compute both intention to treat (ITT) and local average treatment effects (LATE)[4] and find that participating in the training has large and robust effects on knowledge and business practices of women. "Downstream" impacts on profits, profit margins, and sales are weaker, not surprisingly, but we do document some positive effects. Notably, we find some evidence that medium-term profits increase. This is encouraging because women are often disproportionately affected by the hardship of poverty, and unlike earlier work we find that entrepreneurship trainings might help them to alleviate their plight. Existing evidence for Tanzania (Berge et al. 2014) and Pakistan (Giné and Mansuri 2014), for example, reveals that benefits are gender dependent and concentrated among male clients.[5] While we cannot identify why outcomes are more positive for our sample of female clients from Vietnam than for women in other settings, a number of possible explanations are easily provided: the context and culture are different (specifically, Vietnam has a dynamic business climate); the trainings were spread across a prolonged period of nine months and included opportunities for repetition; the trainings were supported by microcredit; and finally, husbands were involved in the trainings (for a subsample of the clients).

This brings us to the second contribution of the paper. We integrated husbands in the intervention and ask whether the presence of men helps to increase

the impact of the training. This speaks to a literature on the gender composition of teams and group performance.[6] While many development interventions, including efforts to increase financial service delivery, exclusively target women, existing work on the (economic) performance of teams suggests that mixed-gender teams on average outperform teams dominated by a single sex (e.g., Fenwick and Neal 2001, Hoogendoorn et al. 2013). In the context of a large business game, Apesteguia et al. (2012) also find that women-only groups are outperformed by any other gender combination. Therefore, it seems sensible to explore whether mixed-gender groups are also more effective vehicles for providing an entrepreneurship training. Our support for this hypothesis, however, is weak. The presence of husbands does not seem to affect the accumulation of knowledge or adoption of practices. However, we do present some evidence suggesting differential outcomes in terms of downstream profits.

The third contribution of the paper follows Berge et al. (2014) and documents that the impact of the intervention evolves over time—there is no positive profit or sales response shortly after finishing the intervention. Hence, the training's impact seems to become stronger over time, presumably because firms have been able to incorporate some of the new insights into practice. Fourth, and finally, we have access to a relatively large sample of female clients, which enables us to pick up even modest effects. Some of the earlier studies produce estimates of impact on economic outcomes that are statistically insignificant for women but nevertheless "point in the right direction." For example, Berge et al. (2014) report a positive but insignificant coefficient for the training dummy in models explaining long-term and average profits of female clients. But their sample consists of only several hundred respondents.

This paper is organized as follows. In Section 2 we outline the background (context), introduce the intervention, and briefly discuss the (implicit) theory of change applied by the MFI. In Section 3 we will introduce and discuss our data, and show that randomization "worked" in the sense that we obtained a data set that appears balanced on observables across treatments and control. We also present, in Section 4, our simple identification strategy, based on crosssection ordinary least squares (OLS) estimators, and panel (difference-in-differences) estimator. Results are presented in Section 5, where we report short-term and medium-term effects of the intervention separately. We report posttreatment intention to treat effects in the main text, and report supportive panel results and local average treatment effects in the online appendix. These panel and LATE findings are consistent with the results reported in the text. Section 6 concludes.

2. Context and Intervention

We conduct our study in northern Vietnam. This is a region, like others in Vietnam, characterized by rapid economic growth and the emergence of a (middle) class of entrepreneurs. Vietnam is characterized as a deeply patriarchal society, with traditional gender norms based on Confucianism and Buddhism (Duvvury et al. 2012). While, according to the Vietnam Country Gender Assessment (VCGA, World Bank 2011), Vietnam has made progress in terms of gender equality, a serious gender gap remains. According to the 2009 Labor Force Survey (LFS), the employment status of no less than 69% of women is "vulnerable," compared to only 54% of men (International Labour Organization 2010). Some 36% of men and 43% of women are classified as unskilled workers.[7]

We collaborate with TYM fund to evaluate the impact of a gender and business training to poor female clients. The TYM fund is the largest microfinance organization in northern Vietnam. It started as a microfinance project of the Vietnam Woman Union in 1989 and has been in operation in its current form since 1992. Its main mission is to improve the quality of life and the status of poor women and their families by providing them with access to financial and nonfinancial services. This is achieved by three financial services: loans, savings opportunities, and mutual assistance funds. As of September 2011, TYM is active in 10 poor areas in northern Vietnam, working through 43 branches. It established 1,450 training centers, each serving 30–40 female clients, for a total of approximately 48,000 female clients. In return for receiving financial and nonfinancial services, women must become members of a lending center. All the services are provided at weekly or monthly center meetings in which loan officers assess loan application forms and collect repayments and savings. Center meetings also allow TYM members to exchange experiences and information about production and business, as well as enable TYM staff and external experts to disseminate knowledge on family, gender, and other issues. Finally, the centers host a range of social activities.

We conducted our study in four TYM branches: one in Hanoi and three in nearby Vinh Phuc. In terms of many observables (e.g., share of population dependent on agriculture-related activities) our sample is comparable to the country average. The women in our sample are engaged in agricultural and nonfarm business activities. The major agricultural activity is rice cultivation (this is indicated by 90% of the women engaged in farming). Other agricultural activities include growing fruits, and raising pigs, cows or poultry (chicken, ducks, doves). In terms of business activities, the women in our sample are mainly involved in retail trade, manufacturing, and services (not further specified). Secondary business activities are vendor trade

and wholesale trade. During the study period, no other trainings were offered to our sample population—neither by TYM nor by other organizations.

We evaluate the impact of a gender and business (G&B) training intervention offered to female clients of TYM. The trainings provided through TYM fund are based on the Gender and Entrepreneurship Together (GET) Ahead for Women in Enterprise Training Package and Resource Kit, designed by the International Labour Organization (ILO). The standard content has been slightly modified to better fit the Vietnamese context. The first module of the program covered basics on gender and entrepreneurship, promotion of equality between men and women, and the life cycle of people and enterprises. The second module considered the businesswoman and her self-confidence. The third module focused on the businesswoman and her environment, self-development, and business mapping. The fourth module discussed business projects, including business ideas, opportunities, and challenges. The fifth module covered marketing and sales. The sixth module covered calculations and aspects of financial literacy. The seventh module focused on managing cash. The eighth module discussed how to record accounts receivable and accounts payable. The ninth and final module covered how to calculate cost of production and cost of goods sold. Whereas the content of these training sessions appears comparable to that of interventions studied by others, the monthly spacing of training sessions is rather different. Specifically, the interventions analyzed by Berge et al. (2014) and Karlan and Valdivia (2011) were based on weekly sessions. The training studied by Giné and Mansuri (2014) involved an intensive successive six-day (hands-on) training based on role play and case studies, and a visit to a local market.

The theory of change, as envisaged by TYM, is very simple: By increasing business knowledge, the training is expected to change business practices; women are expected to implement "innovative" practices such as record keeping and separating domestic from firm finance. They are also expected to engage in marketing activities and professionalize their relations with suppliers and clients. In turn, these changes should improve business outcomes (sales and profits). In addition, the invitation of husbands to participate in the intervention (in one of the treatment arms) was motivated by the expectation that the presence of men would enhance the quality of the discussions about the content of modules because of their experience in business matters. Moreover, exposing men to ideas about gender equality may help improve the bargaining position of women in the household and help them to effectively run their businesses.[8] The ultimate downstream impacts of improved business outcomes and enhanced gender equality are improved outcomes for men and women in terms of such indicators as nutritional status, health, and (eventually) education. However, in this paper we do not focus on these long-term and distant impacts.

As mentioned, we use two treatment arms to probe whether inviting husbands to participate in the training is a (scalable) approach to enhance the impact of the training. Importantly, we did not expect every woman to bring her husband along (even if this would have been welcomed). Some husbands are unlikely to participate, even when incentivized, and 18% of the clients in our sample are single, widowed, divorced, or separated. Hence, our analysis captures the effect of the presence of (some) husbands—not necessarily one's own husband—on business knowledge, practices, and outcomes. To encourage male participation in the trainings, we offered a fee to participating men. Since we were also interested in better understanding how fees affect male participation, we varied the size of the fee across modules. Specifically, and for simplicity, the fee was diminished "linearly" over time. For the first module we offered 100,000 VND (USD 5), and this amount was lowered by 10,000 VND for all successive modules (so that the fee for participating in the sixth module was 50,000 VND).

Trainings took place during nine monthly center meetings, and each module took 45–60 minutes. Because most TYM clients lacked a strong educational background, the trainers used many support tools, including role play, color cards, and pictures, to help clients understand and remember the content. In addition to the monthly training module, the trainers organized the possibility of individual sessions to discuss client-specific problems every week. These consultations generally took 15–30 minutes. Some of the staff members at TYM headquarters were trained by ILO about the GET training package. Before the training started, all loan officers in treatment groups attended "training of trainers" courses taught by the TYM headquarters staff. The training was free of charge and voluntary; clients could leave after they made their loan payment and before the training began.

3. Data

We randomly assign credit centers, each with an average of 30 female clients, to one of the two treatment arms or to the comparison group.[9] Since we randomize the G&B training at the credit center level (to limit spillover effects) we use a cluster sampling approach. The four branches included in this study contain 187 credit centers. We stratify randomization by lending branch, taking similar proportions of treatment and control groups across branches. Because of concerns about noncompliance among husbands (reducing power), we oversampled the treatment groups where husbands were invited. Our design includes

70 credit centers where male partners were invited to join the G&B training (T1), and 31 credit centers for which only female clients were invited (T2). The control groups C include 86 credit centers. We randomly selected 23 members per center for the interviews. A few centers had fewer than 23 clients, in which case we interviewed all borrowers.

At the baseline, midline, and endline we interviewed 4,041, 3,826 and 4,350 female microfinance borrowers, respectively. To increase the sample size for T2, we interviewed all members per center (30 instead of 23) during the midline. Hence, at the midline, 316 "new women" were interviewed. For the three groups, attrition rates between baseline and midline were 13%, 16%, and 13%, for T1, T2, and C, respectively. We interviewed 3,412 female borrowers across all three waves (where T1 = 1,311; T2 = 549; and C = 1,552). During the endline, we also interviewed (almost all) female borrowers who dropped out as members from TYM, so that there is almost zero attrition between midline and endline (0.1%, 0.0%, and 0.2%, for T1, T2, and C, respectively). These women are included in the analyses. Overall attrition rates are low.

In the posttreatment ITT estimates, we use the entire sample. Hence, the most parsimonious specifications for the midline and endline estimates are based on samples of 3,826 and 4,350 clients, respectively (these tables are included in the online appendix). For some variables—notably the baseline controls and some outcome variables—there are missing values, so we also estimate models with a slightly smaller sample size. As mentioned, for the midline and endline we surveyed some additional women, for which we lack information about baseline controls. For the double difference estimates (see the appendix), we estimated unbalanced panels. As a robustness check we also estimated all posttreatment models by only including women surveyed in the baseline and all double difference models with balanced panels. The results of these robustness checks are similar to the results presented in the paper (and available on request).

We use a nonlinear (logit) probability model to study differences in attrition across groups (qualitatively similar results emerge when we estimate a linear model). A dummy for attrition is equal to one if households are not interviewed in at least one of the two follow-up surveys. Since attrition between midline to endline was almost zero, the attrition analyses predominantly refer to attrition between baseline and midline. Regression results are reported in Table 1. In column (1) we regress attrition status on baseline variables and find that only our region dummy enters (marginally) significantly. In column (2) we include the treatment variables and find for both treatment arms attrition is the same as in the control group. In column (3) we include interaction terms between observables and treatment arm and now find that some interaction terms enter significantly. Larger households are associated with reduced attrition risk for both treatment arms (relative to the control group), but the effect is small. Interestingly, married women are more likely to drop out if their

Table 1. Test for Random Attrition

Variables	(1)	(2)	(3)
*T*1		−0.18	0.50
		(0.156)	(0.886)
*T*2		0.24	−1.81
		(0.215)	(1.127)
*Knowledge*1	0.02	0.02	−0.03
	(0.033)	(0.033)	(0.044)
Age	−0.01	−0.01	−0.01
	(0.006)*	(0.006)*	(0.009)
Household size	−0.06	−0.06	0.04
	(0.039)	(0.039)	(0.061)
Married	−0.11	−0.12	−0.18
	(0.148)	(0.147)	(0.220)
Region	−0.40	−0.41	−0.54
	(0.236)*	(0.232)*	(0.359)
Age × *T*1			−0.01
			(0.013)
Hhsize × *T*1			−0.19
			(0.084)**
Married × *T*1			−0.29
			(0.302)
Region × *T*1			0.19
			(0.483)
*Knowledge*1 × *T*1			0.09
			(0.073)
Age × *T*2			0.02
			(0.015)
Hhsize × *T*2			−0.19
			(0.104)*
Married × *T*2			1.17
			(0.430)***
Region × *T*2			0.42
			(0.667)
*Knowledge*1 × *T*2			0.11
			(0.090)
Treatment	−0.04		
	(0.147)		
Constant	−1.34	−1.33	−1.26
	(0.422)***	(0.418)***	(0.602)**
Observations	3,941	3,941	3,941
Test T1 = T2 (*p*-value)		0.06	0.05

Notes. Logit results. Robust standard errors adjusted for clusters in centers are in parentheses. Variables: *T*1 refers to the treatment group where male partners were invited to join the training; *T*2 refers to the treatment group where only female members were invited; *Knowledge*1 refers to business knowledge index 1; *Age* refers to the age of the female member; *Hhsize* refers to household size; *Married* refers to a binary dummy with a one if married; and *Region* refers to a region dummy. A T1 (or T2) behind a variable refers to an interaction term. For example *Married* × *T*1 indicates an interaction between *Married* and *T*1. Column (1) is based on Table 3 in Vu et al. (2015).

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

husbands are not invited to the trainings (compared to both the control group and T1). To attenuate potential concerns about nonrandom attrition, we will compute (lower) bounds on our estimates of treatment effects. These results are presented in an online appendix and discussed below.

Data collection started in October and November of 2011, with a baseline survey. Selected women received the training between February and October 2012. Midline data, intended to capture "short-term effects," were collected in March and April 2013 or some six months after completing the training (and 15 months after completing the first module). One full year after completing the intervention (or some 21 months after completing the first module) we organized an additional endline survey to pick up "medium-term effects." Hence, these data were collected in October and November 2013. We compare short- and medium-term effects to learn about the dynamics of impact—whether the impact tapers off over time (as individuals abandon practices and revert back to pre-training practices) or increases. The impact might increase because trained individuals will have had more time to adopt new practices. In February 2013, we also organized six focus group discussions, as well as in depth interviews with two women from each group to learn about the perceived relevance of the training, and gauge overall satisfaction. In December 2014 some additional focus group discussions and in-depth interviews were conducted to discuss our main findings with different stakeholders.

Baseline values of our dependent and control variables are summarized in Table 2. We draw attention to the fact that 33% of our respondents are involved in nonfarm economic activities, and 78% of our respondents are involved in agricultural activities (production and sales). This means that 884 respondents, or 22% of our sample, are involved in both types. In a further analysis we will zoom in on the subsamples of nonfarm business owners and agricultural business owners and examine whether the impact of the training is different across groups. Recall that the training program did not contain specific modules about agricultural production, to observe that some of the business practices promoted during the training (e.g., "decorating the place to entice customers to the shop") may be less relevant for women specialized in agriculture.

We construct two business knowledge indices, called *Knowledge*1 and *Knowledge*2. Following Karlan and Valdivia (2011), these indices are the sum of correct answers to a series of business-related questions. *Knowledge*1 is based on 16 such questions, asked at the baseline, midline, and endline. *Knowledge*2 is based on 23 (other) questions, and these questions were only asked during the midline and endline. This obviously implies that we cannot use the difference-in-differences

Table 2. Descriptive Statistics: Variables at the Baseline

	N	Mean	Std. dev.	Min	Max
Control variables					
Age (years old)	4,035	43.77	10.33	19.00	72.00
Schooling (years)	4,030	6.82	2.91	0.00	18.00
Married	4,041	0.82	0.39	0.00	1.00
Ethnic group (Kinh)	4,041	0.94	0.23	0.00	1.00
Household size (Hhsize)	3,943	4.74	1.56	1.00	15.00
Region (Hanoi)	4,041	0.26	0.44	0.00	1.00
Credit access TYM	4,037	1.10	0.70	0.00	2.00
Interest in training	4,037	0.76	0.43	0.00	1.00
Monthly income	4,037	6,064.50	3,418.23	0.00	50,000.00
Land size	4,041	1,439.44	1,116.41	0.00	7,200.00
Agricultural activity	4,036	0.78	0.41	0.00	1.00
Nonfarm business activity	4,035	0.33	0.47	0.00	1.00
Dependent variables					
*Knowledge*1	4,041	8.94	1.72	0.00	14.00
General practices	4,036	0.00	1.42	−2.50	2.60
Innovation	4,036	0.00	1.12	−0.37	14.42
Agri_sales (1,000 dong)	4,041	2,311	9,709	0.00	300,000
Agri_profit (1,000 dong)	4,041	405	4,897	−66,667	166,333
Sales (1,000 dong)	4,041	18,267	76,280	0.00	2,400,000
Profit (1,000 dong)	4,041	3,034	17,625	−420,000.00	434,524
Sales_total (1,000 dong)	4,041	20,578	76,678	0	2,400,000
Profit_total (1,000 dong)	4,041	3,438	18,281	−614,500	434,224

Note. This table is based on Table H1 in Vu et al. (2015).

estimator for this variable. Details of the questions are available in Online Appendix 1.

We also use survey data to construct four business practice indices: *General practices*, *Innovation*, *Marketing*, and *Record keeping and planning*. Data for the first two indices were collected three times: at the baseline, midline, and endline. For the other indices we only have midline and the endline values. *General practices* and *Innovation* are based on the outcomes of a principal component analysis (PCA) on seven business practices questions. *General practices* is the first component, and *Innovation* is the second component. The "weights" obtained by applying the PCA on baseline questions are used to construct indices for *General practices* and *Innovation* at midline and endline (see the online appendix for details). The *Marketing* and *Record keeping and planning* indices are calculated similarly. However, in the absence of baseline data, *Marketing* simply refers to the first principal component from a PCA on 13 business questions at midline, and *Record keeping and planning* is the second component. The weights obtained from PCA at midline are used to construct *Marketing* and *Record keeping and planning* indices at endline (again, see the online appendix for details).[10]

We use survey questions to create indicators for profits and sales in the previous month. In the baseline, midline, and endline surveys, we asked respondents to report profits and sales of their three main nonfarm and farming activities. To avoid having differences in the incidence of zero profits or sales across treatment arms affect our results, we deal with zeros by using the inverse hyperbolic sine transformation in some models (rather than taking logs, but we obtain qualitatively similar results when taking logs).[11]

A legitimate concern is that the training itself might affect the accuracy of reporting sales and profits. If so, this could generate nonclassical measurement error in these variables with implications discussed by McKenzie and Woodruff (2014). Empirical results from the literature suggest a mixed picture: whereas de Mel et al. (2014) do not find that trainings affect reporting, Drexler et al. (2014) find that trainings reduce reporting errors, and Berge et al. (2014) find that trainings seem to increase reporting errors. Another potential concern is experimenter demand effects, which would upwards bias our estimates of impact if treated respondents are inclined to produce more favorable numbers than respondents from the control group. Such concerns are attenuated, however, by the simple observation that our "control group" also consists of TYM clients benefitting from access to financial services. Unfortunately, we lack data from other sources to cross-check the self-reported sales and profit outcomes. Hence, we note the important caveat that our impact estimates will overestimate the true impact if trained respondents systematically report higher sales and profits than subjects from the control group. With that caveat in mind, the variables *Agri_sales*, *Agri_profit*, *Sales*, and *Profit* refer to summed sales and profits for these activities in the month prior to the surveys. The variables *Sales_total* and *Profit_total* refer to total sales and profits of the respondent's portfolio of economic activities. As alternative outcomes variables, we also measured sales and profits in a so-called "normal month." To economize on space we do not report results based on these variables, but they are qualitatively similar to the ones presented below and are available on request from the authors. Online Appendix 2 contains a table summarizing our outcome variables (Table A5).

To probe whether randomization resulted in "comparable groups" of female clients, we use baseline data and regress our household variables on the treatment dummies. Results are also reported in Online Appendix 2. Tables A6 and A7 show that demographic and other baseline controls for the treatment groups are not significantly different from the control group at the baseline and that outcome measures of business also tend to be statistically indistinguishable across groups.

We monitored attendance at the training sessions by asking loan officers to keep attendance lists. The average participation rate was about 80% for women. To address issues associated with noncompliance, we will compute both intention to treat (ITT) and local average treatment effects (LATE). Overall, the training content was much appreciated by participating women, and the average score was close to 9 (on a scale from 1 to 10).[12] The participation rate was much lower for husbands (in treatment 1) and declined from 40% to slightly under 10%. Approximately 40% of the husbands did not attend any training session, 60% attended one session or more, and 2% attended all sessions. However, there were always some men present in the sessions of the T1 treatment arm. The reason provided by most men for not participating was high opportunity costs (but of course we cannot exclude that some men found the focus on gender issues uninteresting).[13] In spite of the low participation rate by husbands, both our survey evidence and the focus group discussions confirmed that women appreciated the attendance of husbands. Indeed, female attendance was slightly higher in the treatment arm with husbands present (some 3%–4% higher compliance, on average, which is significant at the $p = 0.1$ level).[14] Finally, our data suggest that male attendance responds to financial incentives, so in future interventions the NGO can increase husband participation by providing sufficient financial compensation.[15]

4. Identification Strategy

We use multiple estimators to evaluate the impact of the training (with and without husbands) on business

knowledge, practices, and outcomes. Specifically, we have computed intention to treat (ITT) effects based on posttreatment data, difference-in-differences estimates, and local average treatment effects (using random assignment to the group receiving an invitation to participate in the training as an instrumental variable for participation). The results are robust across estimators, and to economize on space we only report ex post ITT estimation results in the main text. The other estimation results, and details of the estimation strategy, are relegated to Online Appendix 3 (Tables A8–A10). Random assignment occurs at the center level, so we cluster standard errors at the center level in all models. We include controls (measured at the baseline) in all specifications to improve precision of our estimates, but in the appendix we also report similar results for a series of parsimonious models without baseline controls.

4.1. Posttreatment Analysis (ITT)

First, we use intention-to-treat (ITT) estimators. We estimate impact at both the midline and endline for both treatment arms separately. The posttreatment analysis simply regresses (midline and endline) outcomes Y_{ijt} on treatment dummies groups and controls:

$$Y_{ijt} = \beta_0 + \beta_1 \mathrm{T1}_{ij} + \beta_2 \mathrm{T2}_{ij} + \beta_3 \mathbf{X}_{ij0} + \varepsilon_{ijt}, \quad (1)$$

where Y_{ijt} refers to an outcome variable for an individual i in centre j at the midline or endline survey t; β_0 is a constant, $\mathrm{T1}_{ij}$ is a dummy equal to one if training is offered to a woman as well as to her husband (group T1); $\mathrm{T2}_{ij}$ is a dummy variable that takes the value of one if the woman is selected to receive business training individually; $\mathbf{X}_{ij0}$ is a vector of baseline controls: age, household size, marital status, and region; and ε_{ijt} is an error term. Coefficients of interest are β_1 and β_2; The latter estimates the effect of participating in the standard training at the midline (or endline) survey, and the former estimates the effect of training when husbands are invited. The additional effect of inviting husbands is obtained by subtracting β_2 from β_1.

4.2. Logit Estimation of Business Entry and Exit (ITT)

Next, we gauge the impact of training on the extensive margin and estimate the effect of the training (with and without husbands) on the probability of women to start new business activities or stop their main business activity. For this purpose we use a logistic regression:

$$D_{ijt} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \mathrm{T2}_{ij} + \beta_2 \mathrm{T1}_{ij} + \beta_3 \mathbf{X}_{ijt})}} + \varepsilon_{ijt}, \quad (2)$$

where D_{ijt} is a dummy for business start-up (S_{ij}) or business failure (F_{ij}). The coefficients of interest are β_1 and β_2: the former estimates the effect of training on business start-up and business failure, and the latter estimates the effect of the presence of husbands on women's business entry and exit.

4.3. Probing the Theory of Change

We also seek to assess the plausibility of the (implicit) theory of change of TYM, namely that trainings build knowledge, which affects practices and business outcomes. To do this, we estimate an instrumental variable model where we regress our first business knowledge index on the treatment dummies, and then use predicted knowledge levels to explain variation in business practices adopted, and variation in outcomes (profits as well as sales):

$$\begin{aligned} &\textit{Business knowledge}_{ij} \\ &\quad = \alpha_0 + \alpha_1 \mathrm{T1}_{ij} + \alpha_2 \mathrm{T2}_{ij} + \alpha_3 \mathbf{X}_{ij0} + \varepsilon_{ij}, \quad (3a) \\ &\textit{Practices}_{ij} \\ &\quad = \beta_0 + \beta_1 \widehat{\textit{Business knowledge}}_{ij} + \beta_2 \mathbf{X}_{ij0} + \varepsilon_{ij}, \quad (3b) \\ &\textit{Outcomes}_{ij} \\ &\quad = \beta_0 + \beta_1 \widehat{\textit{Business knowledge}}_{ij} + \beta_2 \mathbf{X}_{ij0} + \varepsilon_{ij}. \quad (3c) \end{aligned}$$

Although useful to evaluate the intervention logic of TYM, the regression results of this model do not necessarily produce an unbiased estimate of the causal effect of knowledge on practices and outcomes. The reason is that the exclusion restriction of the instrument may be violated or, in other words, it is possible that the training affects practices and outcomes via other channels than merely accumulated knowledge. For example, the training could induce women to work more hours, or it could reduce the husbands' hostility. If so, the estimates of β_1 should be interpreted as upper bounds of the impact of knowledge on outcomes. In case the stable unit treatment value assumption is also violated, or the noninterference assumption that the treatment status of any client does not affect the potential outcomes of other clients, then assigning the direction of the bias in instrumental variable estimates is less straightforward.

4.4. Attrition: Lee Bounds Analysis

Whereas our data set suffers only from mild attrition, our impact estimates will be biased when attrition is correlated with potential outcomes. To gauge the sensitivity of our results with respect to such bias, we compute extreme (lower) bounds for the estimated treatment effects by trimming our data set (Lee 2009, Gerber and Green 2012). We compute so-called Lee bounds to create a credible counterfactual for the subsample of MFI clients that always participates in a survey (regardless of treatment status). More specifically, our control group has the lowest attrition rate (13%) and the T2 group the highest attrition rate (16%).

Therefore, we drop, for each outcome variable, the bottom 3% of the (posttreatment) observations of the control group, and redo the analyses using the restricted sample. For example, for *Knowledge*1 we dropped all observations of the control group with an index score below 6, and for *Knowledge*2 we dropped all observations with index scores below 7. This trimming of the control group depresses the difference between outcomes for the treatment and control group, hence the interpretation of the new impact estimate as a lower bound.

5. Results

Tables 3–7 summarize our estimates of the causal effect of participating in the G&B training on business knowledge, practices, and outcomes. As mentioned, these tables present regression results of the ITT estimator, based on midline and endline data. Supportive panel (difference-in-differences) and LATE estimation results are reported in Online Appendix 3. Table 8 contains the logit results for the extensive margin, Table 9 considers the theory of change, and Table 10 reports heterogeneous treatment effects.

Table 3. Training and Business Knowledge

Variables	(1) *Knowledge*1	(2) *Knowledge*1	(3) *Knowledge*2	(4) *Knowledge*2
	Midline	Endline	Midline	Endline
*T*1	2.23 (0.231)***	2.52 (0.281)***	2.73 (0.415)***	3.03 (0.433)***
*T*2	2.05 (0.278)***	2.30 (0.361)***	2.71 (0.539)***	2.63 (0.539)***
Age	−0.00 (0.004)	−0.00 (0.005)	−0.00 (0.009)	−0.00 (0.008)
Household size	0.01 (0.031)	0.05 (0.032)*	0.04 (0.051)	0.10 (0.046)**
Married	0.17 (0.110)	−0.19 (0.116)	0.06 (0.224)	−0.45 (0.158)***
Region	−0.68 (0.228)***	−0.98 (0.306)***	−0.15 (0.386)	−0.75 (0.479)
Constant	10.40 (0.332)***	10.47 (0.358)***	18.51 (0.585)***	14.48 (0.536)***
Observations	3,503	4,234	3,503	4,234
R-squared	0.229	0.195	0.115	0.150
Test T1 = T2 (*p*-value)	0.47	0.55	0.97	0.55

Notes. Linear regression results. Robust standard errors adjusted for clusters in centers in parentheses. Variables: *T*1 refers to the treatment group where male partners were invited to join the training; *T*2 refers to the treatment group where only female members were invited; *Knowledge*1 refers to business knowledge index 1; *Knowledge*2 refers to business knowledge index 2; *Age* refers to the age of the female member; *Married* is to a binary dummy with a value of one if married; and *Region* refers to a region dummy. All control variables refer to "baseline" values (this also holds for the other tables with impact results). This table is based on Table I21 in Vu et al. (2015).

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

5.1. Effects of the G&B Training on Business Knowledge

Table 3 demonstrates that the training has a large impact on our measures of business knowledge. For both of our knowledge indices we find that participating in the training increased performance. All results are statistically significant at the 1% level and are also economically significant, corresponding to approximately 25% of the mean knowledge score (at the baseline) and with more than one full standard deviation of this variable. With respect to the control variables, there is no evidence of robustly significant correlations. Observe that extending the impact period from six to 12 months does not have a significant effect on the depreciation or accumulation of knowledge. Estimated coefficients are rather similar and not different at conventional significance levels.

The table also reveals that inviting husbands does not matter for knowledge accumulation: the coefficients of the T1 treatment arm are statistically indistinguishable from the coefficients of the T2 treatment arm (this is evident from the *p*-values reported in the bottom row of the table). The lack of a significant difference does not necessarily reflect low power, which would translate into large standard errors and imprecisely estimated effects. The coefficients of the two treatment arms are also very similar, differing only by some 2.5% of the mean knowledge score at the baseline. Of course it is an open question to what extent these coefficients would have been (more) different in case of higher response rates among invited husbands.[16]

Extreme (lower) bounds for the ITT estimator as well as the LATE estimator are reported in Online Appendix 4 (Tables A11 and A12). Whereas the estimated coefficients are somewhat smaller than the ones reported in the main text, all estimated treatment effects remain significant at the 1% level.

5.2. Effects of the G&B Training on Business Practices

Table 4 summarizes our impact estimates for the adoption of key business practices: record keeping, marketing efforts, innovation behavior, and general business practices. We again find economically meaningful and statistically significant effects across all models (and similarly for alternative estimators—see Online Appendix 3). Moreover, these positive results emerge for both our midline and endline data sets. All estimated coefficients are significant at the 1% level, and all impacts exceed one standard deviation of the dependent variable of interest (at the baseline).

Unlike the results for the knowledge indices, we now find an apparent additional effect of extending the impact analysis from 6 to 12 months. Our endline estimates of impact on business practices, reported in columns (5)–(8) of Table 4, are significantly *greater*

Table 4. Training and Business Practices

Variables	(1) *General practices*	(2) *Innovation*	(3) *Marketing*	(4) *Record keeping and planning*	(5) *General practices*	(6) *Innovation*	(7) *Marketing*	(8) *Record keeping and planning*
	Midline	Midline	Midline	Midline	Endline	Endline	Endline	Endline
*T*1	1.25	2.96	1.69	1.93	1.82	5.68	3.11	2.74
	(0.128)***	(0.422)***	(0.196)***	(0.180)***	(0.104)***	(0.438)***	(0.173)***	(0.173)***
*T*2	1.26	3.17	2.00	2.04	1.74	6.07	2.94	2.62
	(0.167)***	(0.592)***	(0.222)***	(0.224)***	(0.135)***	(0.596)***	(0.230)***	(0.209)***
Age	−0.01	−0.03	−0.02	−0.01	0.00	−0.00	−0.00	−0.00
	(0.003)***	(0.010)***	(0.004)***	(0.004)***	(0.002)	(0.009)	(0.004)	(0.003)
Household size	0.03	0.01	0.02	−0.00	0.02	−0.01	0.04	0.04
	(0.015)**	(0.052)	(0.022)	(0.022)	(0.013)	(0.053)	(0.023)*	(0.020)**
Married	0.34	0.18	0.22	0.22	0.26	0.36	0.22	0.20
	(0.078)***	(0.226)	(0.099)**	(0.088)**	(0.065)***	(0.204)*	(0.085)**	(0.085)**
Region	−0.35	0.58	0.03	−0.23	−0.17	0.57	−0.00	0.07
	(0.146)**	(0.415)	(0.172)	(0.180)	(0.113)	(0.452)	(0.159)	(0.156)
Constant	0.29	2.03	−0.51	−0.73	−0.11	1.69	−1.24	−1.17
	(0.205)	(0.596)***	(0.252)**	(0.234)***	(0.153)	(0.603)***	(0.220)***	(0.212)***
Observations	3,492	3,492	3,487	3,487	4,180	4,180	4,145	4,145
R-squared	0.204	0.122	0.219	0.288	0.382	0.329	0.445	0.440
Test T1 = T2 (*p*-value)	0.93	0.78	0.22	0.67	0.55	0.53	0.43	0.55

Notes. Linear regression results. Robust standard errors adjusted for clusters in centers in parentheses. Variables: *T*1 refers to the treatment group where male partners were invited to join the training; *T*2 refers to the treatment group where only female members were invited; *Age* refers to the age of the female member; *Married* refers to a binary dummy with a one if married; *Region* refers to a region dummy. *General practices* refers to the general business practices index; *Innovation* refers to the innovation business practices index; *Marketing* refers to the marketing skills business practices index; and *Record keeping and planning* refers to the record and planning business practices index. This table is based on Table I22 in Vu et al. (2015).

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

than our estimates of impact at the midline reported in columns (1)–(4) of Table 4 ($p < 0.05$ for all measures). This reflects that adopting new practices takes time. For example, consider our general practices measure. An additional six months delay in impact measurement results in an increase in the index score of approximately 0.6, or about 45% of a standard deviation (measured at the baseline). The impact on innovation is even larger. Hence, these results suggest that limiting the analysis to short-term data, collected within six months of the training, would underestimate the true effect of the training.

As above, *p*-values in the bottom row of Table 4 reveal inviting husbands does *not* significantly change the effect of training on business practices. Again, estimated coefficients are very similar across treatment arms. In Online Appendix 4 we present lower bounds for the ITT and LATE for all business practices; all treatment effects remain significant at the 1% level (Tables A13–A17).

5.3. Effects of the G&B Training on Nonfarm and Farming Outcomes

In Table 5 we summarize the estimation results for our variables measuring aggregate sales and profits. Table 6 contains our estimates of the impact of participating in the training on sales and profits for subjects engaged in nonfarm business activities. We report outcomes for the summation across the three main economic activities, which should internalize almost all effects because very few subjects were engaged in more than three activities. Whereas the training did not focus on agricultural production per se, some of the lessons are likely relevant for activities in this domain as well, so we report impacts in terms of sales and profits for agricultural activities separately in Table 7. Since several respondents in the sample reported zero sales or profits, we also report outcomes based on the inverse hyperbolic sine transformation, which can handle zeros and can be interpreted in the same way as a standard logarithmic variable (see Burbidge et al. 1988).

Not surprisingly, these "downstream" results are a lot more variable than the earlier ones. Nevertheless, while more tentatively, we believe the findings tell a compelling story. Although the training does not affect short-term sales or profits, we do observe significant impacts on profits at the endline (columns (7) and (8) of Table 5). The gain in total profits equal about 30% of the mean value of our measure of total profits at the baseline. In absolute terms, the average increase in profits equals about USD 70 per year. Moreover, when using the inverse hyperbolic sine transformation, our results also suggest that aggregate sales increased in response to the training (column (6) of Table 5). The latter effect on sales is significantly more pronounced in

Table 5. Training and Total Outcomes

Variables	(1) *Sales_total*	(2) *ihst* (*Sales_total*)	(3) *Profit_total*	(4) *ihst* (*Profit_total*)	(5) *Sales_total*	(6) *ihst* (*Sales_total*)	(7) *Profit_total*	(8) *ihst* (*Profit_total*)
	Midline	Midline	Midline	Midline	Endline	Endline	Endline	Endline
*T*1	−1,097.08	0.02	636.46	0.21	2,598.29	1.53	1,646.33	2.14
	(2,490.874)	(0.243)	(844.112)	(0.480)	(3,350.347)	(0.265)***	(460.294)***	(0.531)***
*T*2	−2,277.15	0.42	282.28	0.57	−572.09	0.86	1,136.44	1.60
	(2,424.642)	(0.257)	(652.686)	(0.620)	(3,404.545)	(0.340)**	(699.969)	(0.651)**
Age	−444.40	−0.04	−55.13	−0.05	−561.25	−0.02	−77.57	−0.04
	(79.392)***	(0.009)***	(27.091)**	(0.014)***	(106.382)***	(0.007)**	(14.270)***	(0.011)***
Household size	713.35	0.13	238.25	0.03	786.22	0.07	−39.95	−0.07
	(508.563)	(0.048)***	(148.544)	(0.076)	(502.288)	(0.045)	(99.597)	(0.084)
Married	2,411.04	1.16	265.60	0.84	4,093.46	0.92	436.31	0.79
	(1,632.687)	(0.219)***	(433.262)	(0.318)***	(1,515.511)***	(0.201)***	(323.514)	(0.310)**
Region	−7,796.60	−1.54	−1,574.85	−3.17	−9,501.50	−0.87	−982.20	−1.58
	(2,039.088)***	(0.257)***	(531.051)***	(0.571)***	(2,555.608)***	(0.283)***	(461.056)**	(0.563)***
Constant	32,419.31	7.47	3,578.41	4.74	37,597.66	6.97	5,952.80	5.03
	(5,636.665)***	(0.529)***	(1,956.972)*	(0.870)***	(6,021.649)***	(0.435)***	(826.259)***	(0.705)***
Observations	3,503	3,503	3,503	3,503	4,234	4,234	4,234	4,234
R-squared	0.017	0.051	0.004	0.050	0.021	0.052	0.020	0.043
Test T1 = T2 (*p*-value)	0.66	0.14	0.70	0.56	0.30	0.04	0.50	0.40

Notes. Linear regression results. Posttreatment analyses. Robust standard errors adjusted for clusters in centers in parentheses. Variables: *T*1 refers to the treatment group where male partners were invited to join the training; *T*2 refers to the treatment group where only female members were invited; *Age* refers to the age of the female member; *Married* refers to a binary dummy with a one if married; *Region* refers to a region dummy. *Sales_total* refers to previous month sales; *Profit_total* refers to previous month profits; and *ihst* $(\cdot)$ refers to the inverse hyperbolic sine transformation $(\cdot)$. Control variables are measured at baseline.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 6. Training and Business Outcomes

Variables	(1) *Sales*	(2) *ihst* (*Sales*)	(3) *Profits*	(4) *ihst* (*Profits*)	(5) *Sales*	(6) *ihst* (*Sales*)	(7) *Profits*	(8) *ihst* (*Profits*)
	Midline	Midline	Midline	Midline	Endline	Endline	Endline	Endline
*T*1	−1,284.25	−0.16	605.29	0.03	819.49	0.48	975.20	0.41
	(2,518.022)	(0.357)	(843.833)	(0.358)	(3,329.905)	(0.395)	(435.129)**	(0.335)
*T*2	−2,264.14	0.27	344.58	0.57	−1,491.86	0.61	752.14	0.49
	(2,459.308)	(0.523)	(645.786)	(0.503)	(3,417.660)	(0.494)	(683.749)	(0.421)
Age	−450.30	−0.10	−54.85	−0.07	−577.76	−0.08	−83.57	−0.07
	(79.841)***	(0.009)***	(27.279)**	(0.009)***	(107.185)***	(0.009)***	(14.339)***	(0.008)***
Household size	515.14	−0.01	195.94	−0.01	544.16	−0.04	−84.01	−0.04
	(514.785)	(0.058)	(148.105)	(0.054)	(501.207)	(0.052)	(91.067)	(0.044)
Married	1,847.08	0.40	111.91	0.29	3,418.48	0.64	230.03	0.50
	(1,636.255)	(0.263)	(432.180)	(0.234)	(1,524.063)**	(0.225)***	(317.008)	(0.189)***
Region	−7,155.34	−1.51	−1,622.40	−1.38	−10,929.77	−1.71	−1,696.08	−1.45
	(2,036.662)***	(0.345)***	(521.692)***	(0.339)***	(2,429.922)***	(0.375)***	(385.342)***	(0.316)***
Constant	31,885.92	7.71	3,481.62	5.43	38,278.92	6.84	6,386.09	5.84
	(5,689.637)***	(0.575)***	(1,955.412)*	(0.593)***	(6,106.917)***	(0.550)***	(824.039)***	(0.489)***
Observations	3,503	3,503	3,503	3,503	4,234	4,234	4,234	4,234
R-squared	0.016	0.061	0.004	0.044	0.023	0.057	0.024	0.054
Test T1 = T2 (*p*-value)	0.72	0.42	0.77	0.27	0.44	0.80	0.76	0.87

Notes. Linear regression results. Posttreatment analyses at midline and endline. Robust standard errors adjusted for clusters in centers in parentheses. Variables: *T*1 refers to the treatment group where male partners were invited to join the training; *T*2 refers to the treatment group where only female members were invited; *Age* refers to the age of the female member; *Married* refers to a binary dummy with a one if married; *Region* refers to a region dummy; *Sales* refers to business sales in the previous month; *Profit* refers to profits in the previous month; and *ihst* (·) refers to the inverse hyperbolic sine transformation (·). Control variables are measured at baseline.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 7. Training and Agricultural Outcomes

Variables	(1) *Agri_sales*	(2) *ihst* (*Agri_sales*)	(3) *Agri_profit*	(4) *ihst* (*Agri_profit*)	(5) *Agri_sales*	(6) *ihst* (*Agri_sales*)	(7) *Agri_profit*	(8) *ihst* (*Agri_profit*)
	Midline	Midline	Midline	Midline	Endline	Endline	Endline	Endline
*T*1	187.17	0.04	31.17	0.05	1,778.80	1.67	671.13	2.02
	(251.003)	(0.340)	(107.840)	(0.498)	(494.889)***	(0.364)***	(210.105)***	(0.606)***
*T*2	−13.01	−0.02	−62.31	−0.22	919.76	0.57	384.30	1.06
	(313.595)	(0.421)	(135.841)	(0.622)	(579.365)	(0.533)	(255.576)	(0.713)
Age	5.90	0.05	−0.28	0.04	16.51	0.05	6.00	0.05
	(8.719)	(0.009)***	(3.519)	(0.012)***	(9.553)*	(0.009)***	(4.212)	(0.012)***
Household size	198.21	0.25	42.31	0.07	242.05	0.22	44.07	0.03
	(54.666)***	(0.052)***	(22.302)*	(0.071)	(63.073)***	(0.047)***	(31.923)	(0.083)
Married	563.96	1.04	153.70	0.55	674.98	0.85	206.28	0.45
	(214.962)***	(0.211)***	(74.186)**	(0.275)**	(221.759)***	(0.197)***	(106.287)*	(0.298)
Region	−641.27	−0.39	47.55	−1.80	1,428.27	0.24	713.88	−0.24
	(247.834)**	(0.318)	(121.041)	(0.529)***	(698.604)**	(0.366)	(303.951)**	(0.615)
Constant	533.39	0.46	96.79	−1.56	−681.26	0.91	−433.28	−1.83
	(510.740)	(0.552)	(226.841)	(0.790)**	(560.682)	(0.526)*	(226.122)*	(0.740)**
Observations	3,503	3,503	3,503	3,503	4,234	4,234	4,234	4,234
R-squared	0.010	0.036	0.002	0.023	0.028	0.061	0.025	0.025
Test T1 = T2 (*p*-value)	0.52	0.88	0.49	0.68	0.26	0.04	0.36	0.19

Notes. Linear regression results. Posttreatment analyses. Robust standard errors adjusted for clusters in centers in parentheses. Variables: *T*1 refers to the treatment group where male partners were invited to join the training; *T*2 refers to the treatment group where only female members were invited; *Age* refers to the age of the female member; *Married* refers to a binary dummy with a one if married; *Region* refers to a region dummy. *Agri_sales* refers to previous month sales of agricultural activities; *Agri_profit* refers to previous month profits of agricultural activities; and *ihst* (·) refers to the inverse hyperbolic sine transformation (·). Control variables are measured at baseline.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

case husbands were invited to participate in the training ($p = 0.04$).

When breaking down these results into business outcomes and the outcomes of agricultural activities, we observe that the impact is mainly caused by agricultural outcomes. In terms of extra nonfarm sales, the training does not seem to have a significant effect even at the endline. One coefficient is significant and another negative, perhaps suggesting that participating in the training and following up on the lessons learned during the sessions was at the expense of day-to-day running of the business. But this result is not robust, and it disappears after taking the inverse hyperbolic sine transformation. We also do not detect robust differences between the two treatment arms—inviting husbands does not significantly affect sales (p-values for differences across treatment arms are always larger than 0.10).

The results for nonfarm profits are somewhat more encouraging, even if not robust. Specifically, although we fail to document any significant impact based on the midline data, we document positive coefficients based on the endline estimates for one treatment variable in one model. When explaining untransformed profit levels, the T1 dummy enters significantly (column (7) of Table 6). However, the coefficient that we obtain after the transformation is not significantly different from zero.

The finding that profits are positively affected while sales are unaffected suggests that the training has affected profits via a reduction in costs. Unfortunately, we lack sufficiently detailed data to test this conjecture empirically. We also observe that the coefficient associated with the sales variable has a positive sign (but is imprecisely estimated). The coefficients associated with T1 and T2 are again not significantly different from each other ($p > 0.10$). The results based on the difference-in-differences estimator and the LATE estimator (in Online Appendix 3) are also suggestive of a significant impact on profits for treatment arm T1, and again we cannot reject the hypotheses that the estimated coefficients are the same across models.

Since the majority of the women in our sample are engaged in agricultural activities, we next consider the effects on sales of agricultural products and profits thereof. Results are reported in Table 7. We find rather robust evidence that household size and marital status are positively correlated in farming outcomes. More relevant for the focus of this paper, we again find that (i) the training seems to have very little impact on agricultural outcomes in the short term, but (ii) the training positively affects outcomes in the medium term. Specifically, we now find that both sales and profits are positively affected by the training for T1. The outcomes for the treatment arm without husbands are not significantly different from zero (but, again, coefficients are positive and not significantly different from the coefficients of the T1 training either; the one exception to that rule concerns the transformed *Agri_sales* presented in column (6) of Table 7).

Online Appendix 4 contains the lower bounds for the various relevant treatment effects (Table A18). These lower bounds are close to the estimates in Tables 5–7, and lower bounds for our profit measure are significantly different from zero. This suggests our results are not sensitive to attrition bias.[17] The relatively small impact of the training on profits is in line with the existing literature, surveyed by McKenzie and Woodruff (2014). Only few studies find significant positive effects of trainings on profits, partly because of low power of most studies. Interestingly, whereas our data suggest the impact of the training on profits increases over time, other studies suggest the reverse (Berge et al. 2012, de Mel et al. 2014).

5.4. Effects of the G&B Training on Business Entry and Survival

Does the training affect business activity at the extensive margin, by speeding up or delaying the start-up of new economic activities, and the exit of existing ones? Vietnam's business community is dynamic, as illustrated by the simple fact that no less than 194 women in our sample reported to have started up new business activities at the midline (and no less than an additional 170 activities started up between midline and endline).

Table 8 shows that trained women were more likely to start new activities (significant at the 1% level). This finding is not unexpected, given that modules 3 and 4 of the training focused on self-development, business mapping, and business opportunities.[18] In a separate analysis we compared the economic performance of preexisting and new business activities and, not surprisingly, found that activities existing at the baseline were, on average, more profitable than new activities. This implies that our above-mentioned estimates of the impact of the intervention on profits should be interpreted as a lower bound for the impact on existing activities because they are based on the performance of both existing and new activities (details available from the authors on request).

We also find that the effect is larger for the treatment arm where husbands were *not* invited to participate in the training ($p < 0.05$). Rather speculatively, this would be consistent with a story emphasizing that husbands prefer their women around the house, working on chores, rather than starting up new businesses. If so, it appears as if the aim to promote gender equality by inviting husbands may have backfired. Additional research on the intra-household implications of participating in gender trainings seems worthwhile (see also Armendariz et al. 2010).

Exit of business activities is defined as business activities reported at the baseline, which were subsequently

Table 8. Training and Business Entry and Exit

Variables	(1) *Entry*	(2) *Exit*
*T*1	0.06	−0.03
	(0.019)***	(0.038)
*T*2	0.13	−0.09
	(0.035)***	(0.040)**
Age	−0.00	0.00
	(0.000)***	(0.001)
Household size	0.00	0.01
	(0.002)	(0.009)
Married	−0.02	−0.02
	(0.012)	(0.038)
Region	0.02	0.09
	(0.019)	(0.053)*
Observations	4,234	1,338
Test T1 = T2 (*p*-value)	0.03	0.17

Notes. Logit estimation results. Coefficients refer to marginal effects. Variables: *T*1 refers to the treatment group where male partners were invited to join the training; *T*2 refers to the treatment group where only female members were invited; *Age* refers to the age of the female member; *Married* refers to a binary dummy with a one if married; *Region* refers to a region dummy. *Entry*: dummy variable indicating if a household started a new business activities at the midline and/or at endline. *Exit*: business activities reported at the baseline which were subsequently abandoned at the midline or endline. Robust standard errors adjusted for clusters in centers are in parentheses.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

abandoned at the midline or endline.[19] Some 1,338 women reported to undertake one or more business activities at the baseline. Of these women, 281 stopped one or more economic activities at the midline, and an additional 252 women stopped an activity between the midline and endline. Variation in the probability of stopping with an economic activity is correlated with participating in the training, but only significantly so for treatments where husbands were not invited. Training may promote exit if the abandoned activities generated net losses for the household (see also Valdivia 2014), but we find the opposite effect—participating in the training *reduced* exit. This would be consistent with the interpretation that the incidence of loss-generating activities is reduced by the training (for which we report mixed evidence in Tables 5 and 6). However, when we interact exit with baseline measures of profitability, we do not find that loss-making activities are more likely to have terminated (details not reported). We can only speculate why this is the case. Perhaps losses are perceived as temporary, or perhaps revenues still exceed marginal costs. In both cases, staying in the business may make good economic sense.[20]

5.5. Probing the Theory of Change

We now probe the theory of change and try to establish whether the effects on the adoption of new business practices and profits are indeed caused by augmented knowledge levels. Representative results, based on *Knowledge*1, are summarized in Table 9. Columns (1)–(4) of panel A contain the results for our measures of business practices. We find strong and robust evidence that exogenous variation in knowledge levels (induced by participating in the training) affects practices measured at the endline (similar results are obtained when we consider midline data to measure practices). The estimated coefficients are consistently greater than zero ($p < 0.01$). In columns (5)–(8) of panel A we consider the downstream effects on total sales and total profits, and again we find positive coefficients.

Table 9. Probing the Theory of Change

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Knowledge, business practices, and total sales and profits at endline								
Variables	*General practices*	*Innovation*	*Marketing*	*Record keeping and planning*	*Sales_total*	*Profit_total*	*ihst* (*Sales*)	*ihst* (*Profit*)
*Knowledge*1	0.73	2.35	1.24	1.10	616.46	601.22	0.53	0.80
	(0.082)***	(0.300)***	(0.136)***	(0.112)***	(1,226.931)	(175.129)***	(0.116)***	(0.217)***
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,180	4,180	4,145	4,145	4,234	4,234	4,234	4,234
Panel B: Knowledge, business practices, and agricultural and business sales and profits at endline								
Variables	*Sales*	*ihst* (*Sales*)	*Profits*	*ihst* (*Profits*)	*Agri_sales*	*ihst* (*Agri_sales*)	*Agri_profit*	*ihst* (*Agri_profit*)
*Knowledge*1	10.16	0.21	367.23	0.18	606.31	0.53	233.99	0.69
	(1,238.471)	(0.140)	(163.895)**	(0.119)	(176.662)***	(0.159)***	(74.161)***	(0.239)***
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,234	4,234	4,234	4,234	4,234	4,234	4,234	4,234

Notes. Second stage of a 2SLS model where business knowledge was instrumented by the treatment dummies. Robust standard errors adjusted for clusters in centers in parentheses.

**$p < 0.05$; ***$p < 0.01$.

In panel B of Table 9 we report results for nonfarm and agricultural business separately. All outcome measures were collected at the endline. We find support for the thesis that enhanced knowledge boosts (agricultural and nonfarm) profits. Whereas agricultural sales are also positively affected by the training, the same appears not true for nonfarm sales.

5.6. Heterogeneous Treatment Effects

Are the effects similar across all respondents, or are they conditional on specific baseline characteristics? To probe this question, we next consider the effect on our *Knowledge*1 variable, and "split" the sample in multiple ways. For example, we distinguish between households engaged in (i) nonfarm business activities, or not, and (ii) agriculture, or not. We also differentiate between respondents based on baseline demographics (household size, marital status), and compare above and below median respondents. Results are reported in Table 10. We report very little evidence of heterogeneous treatment effects—for both the women-only and treatment as well as the treatment with husbands. Whereas the coefficients of the treatment variables are similar within and across models, the interaction terms of treatment status and baseline variables tend to be insignificant. There is one exception to that rule; the interaction term of T1 and owning a business at the bassline enters positive and significant, suggesting the training has most impact on women who already engaged in nonfarm business

Table 10. Heterogeneous Treatment Effects at Endline

Split on:	(1) *Nonfarm business activities*	(2) *Agricultural activities*	(3) *Household size*	(4) *Married*
*T*1	2.49	2.63	2.34	2.43
	(0.308)***	(0.293)***	(0.431)***	(0.341)***
*T*1 × *Business*	0.41			
	(0.239)*			
*T*2	2.02	2.70	2.40	2.67
	(0.447)***	(0.427)***	(0.431)***	(0.402)***
*T*2 × *Business*	0.60			
	(0.417)			
Age	−0.00	−0.01	−0.00	−0.00
	(0.005)	(0.005)	(0.005)	(0.005)
Household size	0.03	0.03	0.05	0.05
	(0.032)	(0.032)	(0.056)	(0.032)*
Married	−0.20	−0.17	−0.19	−0.14
	(0.121)*	(0.120)	(0.115)*	(0.205)
Region	−1.15	−1.19	−0.98	−0.98
	(0.285)***	(0.289)***	(0.305)***	(0.306)***
*T*1 × *Agriculture*		−0.02		
		(0.243)		
*T*2 × *Agriculture*		−0.59		
		(0.378)		
*T*1 × *Hhsize*			0.04	
			(0.078)	
*T*2 × *Hhsize*			−0.02	
			(0.084)	
*T*1 × *Married*				0.12
				(0.256)
*T*2 × *Married*				−0.43
				(0.261)
Constant	10.75	10.79	10.50	10.43
	(0.370)***	(0.374)***	(0.383)***	(0.385)***
Observations	3,495	3,495	4,234	4,234
Test T1 + T1* = T2 + T2 (*p*-value)	0.49	0.24	1	0.42
Test T1 = T2 (*p*-value)	0.33	0.86	0.89	0.53

Notes. Linear regression results. The dependent variable is *Knowledge*1. Robust standard errors adjusted for clusters in centers in parentheses. Variables: *T*1 refers to the treatment group where male partners were invited to join the training; *T*2 refers to the treatment group where only female members were invited; *Age* refers to the age of the female member; *Hhsize* refers to household size; *Married* is to a binary dummy with a one if married; and *Region* refers to a region dummy. Control variables are measured at baseline. *Ti* × *Business*: interaction term between *Ti* ($i = 1, 2$) and dummy for having business activities at baseline; *Ti* × *agriculture*: interaction term between *Ti* and dummy for having agricultural activities at baseline; *Ti* × *Hhsize*: interaction term between *Ti* and household size at baseline; *Ti* × *Married*: interaction term between *Ti* and dummy for being married at baseline. The last two rows in the table refer to Wald tests for testing *T*1 + *T*1*interaction term = *T*2 + *T*2*interaction term; and *T*1 = *T*2.

*$p < 0.1$; ***$p < 0.01$.

activities from the start. (Note that the coefficient for the T2 treatment is even larger, but less precisely estimated.) We have repeated this exercise for additional baseline variables (baseline knowledge and schooling levels, age) and dependent variables (several measures of practices), and we consistently find little evidence of heterogeneous treatment effects. We have also interacted the region dummy with the treatment variables to examine whether the impacts vary with the degree of urbanization, but again we did not find significant differences.

6. Conclusions and Discussion

Recent evidence suggests microfinance is *not* the panacea it was once believed to be and that augmenting enhanced access to financial capital with efforts to accumulate human capital may be a more successful strategy to reduce poverty. In response, so-called business training programs and financial literacy trainings for micro-entrepreneurs have been implemented in many countries. However, knowledge of the impact of such trainings on "downstream" outcomes of interest among the target population is sparse.

We evaluate the impact of a gender and business training that was offered to female clients of a large microfinance institution in Vietnam. We consider impacts on a range of outcomes, varying from knowledge to profits, and "unpack" profits by distinguishing between the returns to different activities. We also seek to assess whether the nature of this impact "evolves" over time—i.e., is different in the short-term than in the medium term—and whether the impact of the training is conditional on the presence of husbands at the trainings. This, we believe, is an important contribution because gender issues are at the heart of many microfinance initiatives (if only because many projects, including the one we evaluate, exclusively target female clients). The added value of including men in initiatives for women is debated in the literature, but some papers suggest that mixed-gender teams outperform teams consisting of women only.

Our results provide support for the finance-plus approach to development. We find that the gender and business trainings improve knowledge, increase the uptake of new business practices, and after some delay cause an increase in (self-reported) profits (even if the evidence for this final link is weaker). Whereas considerable volatility in sales and profit data imply we should be cautious when interpreting the impact on "deep" outcome variables, we find tentative evidence that the magnitude of the impact varies over time. If so, measuring the impact on downstream variables like profits is likely to result in underestimates of the true impact if data are collected too early after the end of the training. We also document effects at the extensive margin and find that participating in the training may increase the start-up of new economic activities and slow down the exit of existing ones. Next, we document that the general business training significantly increased the returns to agricultural practices, even if agricultural production was not part of the curriculum. We do not find evidence of heterogeneous treatment effects; the estimated coefficient does not vary with a range of household characteristics and region dummies.[21]

Not all our hypotheses were supported by the data. Most importantly, we find only weak evidence of additional impact of including husbands in the training. Women appear to obtain slightly larger sales and profits if they participated in training sessions with some men present; we consistently find that estimated treatment effects on sales and profits are larger when men are involved in the trainings, even if the difference between the coefficients is typically not significantly different from zero. Perhaps women learned more during such male-attended sessions because of more lively interaction or additional discussion, but this is not reflected in performance on the knowledge test. Alternatively, perhaps the act of inviting husbands to the training alleviates intra-household tensions and facilitates the woman to run her business. The nature of our data does not allow us to speak to this important issue, but we do recognize this is an important topic for future research. Moreover, we are careful not to dismiss the potential contribution of participating husbands too lightly. Their participation was appreciated by the women, and it is possible that positive outcomes emerge along other dimensions (i.e., beyond business-related variables). We also note that most men were not interested in participating in the trainings—especially when the financial incentive associated with participation was reduced—so that participation rates were low.

Finally, it is instructive to compare the costs and the benefits of the training intervention. Adopting a financial perspective, outcomes for the MFI are not necessarily positive. Whereas the trainings were offered at very low cost, because of the integration with regular credit center activities, it is not obvious that there are financial or commercial benefits for the MFI. The total costs of the training intervention are estimated to amount to USD 12,400 (of which nearly 70% was spent as an inducement for husbands). This amounts to less than USD 7 per trained client, which compares very favorably to cost assessments in alternative contexts. However, there is no evidence that the treatment increased savings or loans, or improved repayment rates (which are very high across all treatment arms).[22] From a financial cost-benefit perspective, therefore, the training intervention appears to be a poor investment for TYM. However, TYM's objectives are much broader than making profits on financial services; empowering women and facilitating the growth of their businesses are at the heart of the organization's strategy.

Comparing the annual gain in total profits for the beneficiaries according to the endline model (ITT of USD 60-80) to the one-off training costs for the MFI (USD 7) and the opportunity cost of time for participating women, upscaling the intervention appears worthwhile. Indeed, it has been decided to upscale the training intervention on these grounds.

Acknowledgments

The evaluation of the gender and business training is published as a 3ie report; see Vu et al. (2015). A preliminary analysis appears in van Velzen (2014).

Endnotes

[1] According to Berge et al. (2012) these estimates of impact are best viewed as upper bounds of what can be achieved when research-led interventions are "scaled up" to the (sub)national level and are no longer managed by local organizations or researchers.

[2] Interestingly, the authors use lab-in-the-field experiments to collect additional behavioral data they can use to shed light on the mechanism linking the training to outcomes.

[3] In fact, Berge et al. (2014) measure impact after one and three years, so the short term of that study corresponds with the timing of our endline.

[4] Since there is only one-sided noncompliance in our experiment (see Section 3), the estimated LATE parameter is also known as the effect of Treatment on the Treated (TOT).

[5] Some studies focusing on female entrepreneurs do find some positive effects of training on business outcomes (Calderon et al. 2013, Valdivia 2014). According to Calderon et al. (2013), business training increases profits, revenues, and the amount of clients served due to improvements in business knowledge and practices. Valdivia (2014) concludes that women with general business training are more likely to stop losing business activities. Women receiving personal assistance plan and execute more innovations, resulting in an increase in sales.

[6] Potential benefits from mixed-gender groups may derive from complementary preferences, resulting in more balanced decision making. For example, lab experiments have shown that men and women have different risk preferences, social preferences, and competitive preferences (for overview, refer to Croson and Gneezy 2009, and Charness and Gneezy 2012). Lab evidence also suggest robust differences in attitudes towards cooperation (e.g., Charness and Rustichini 2011), which may be less conducive for effective inter-gender cooperation and performance.

[7] Compared with other developing countries where the impact of entrepreneurship trainings has been analyzed, the gender situation in Vietnam is neither particularly good nor bad. This is evident from the World Bank gender equality data and statistics, but also from a composite index, called the Social Institutions and Gender Index (GISI). Vietnam has an index score of 0.18, which classifies the country as "medium" in terms of gender inequality. This score is readily compared with that of other countries for which business evaluations are available. For example, Peru has an index score of only 0.08, corresponding with "low inequality." Instead, Tanzania and Pakistan have outcomes that are worse ("high inequality") with index scores of 0.25 and 0.30, respectively.

[8] As will become clear, we cannot test whether the training affected ideas about gender equality. The reason is that participation in the training (by men) was voluntary and very incomplete. This implies that the presence of specific husbands is an endogenous variable, and documenting the differences between participating and nonparticipating husbands would imply mixing the impact of the training and (self) selection.

[9] Our study followed the standard ethical guidelines: before starting we discussed ethical issues with TYM and the Vietnam Woman Union, and obtained ethical approval from the Vietnam Woman Union. We also used "informed consent."

[10] We used principal component analyses to construct business practices indices because we wanted to reduce the number of variables related to business practices to a few, interpretable linear combinations of the data, enabling us to examine the impact of the training on different types of business practices.

[11] The inverse hyperbolic sine transformation is defined as $\log(y_i + (y_i^2 + 1)^{0.5})$. Except for very small values of y_i, this is approximately the same as $\log(2y_i)$, so it can be interpreted the same way as a standard logarithmic dependent variable. See Burbidge et al. (1988) for details.

[12] Yet, when directly asked, only 16% of the women indicated a willingness to pay for the business training.

[13] However, participating men were positive about the quality and usefulness of the training. Between 87% and 97% of those men reported that they learnt something new from the training and that they applied what they learnt in their own business.

[14] It is possible that higher attendance can have an independent effect on outcomes, If so, this factor creates a potential confound in the interpretation of the results: different impacts found across treatment arms may be due to the treatment itself (our interpretation), or possibly the fact that relatively more women attended the training.

[15] Since we varied the fee across training modules (albeit not randomly), we can explore to what extent financial compensation affects husbands' take up rates. We estimate a logit model explaining husband attendance, add husband and module dummies, and cluster standard errors at the center level. Our regression results (not shown) reveal that if the fee increases by 10,000 VND (0.5 USD), the take up rate increases by 2.7%.

[16] As mentioned in Endnote 8, we could readily zoom in on the subsample of women whose husbands *did* attend (a sizable fraction of) the trainings; however, we refrain from doing so because this subsample would suffer from selection bias (so that the treatment dummy would constitute an endogenous variable in models explaining knowledge).

[17] For example, the Lee lower bound associated with the *Last month profit* variable, measured at the endline, is still significant at the 1% level for the treatment with husbands. The Lee lower bound for *Last month main profits* is marginally insignificant ($p = 0.11$).

[18] In addition, one might expect that start-ups are more likely when women are more confident (or face less restrictions because of the gender training).

[19] Observe the difference in sample size across the two models. For the entry model we use the whole sample and consider which respondents had started up a new economic activity. For the exit model we only include respondents who owned a business at the baseline.

[20] We also asked respondents for the reasons why they terminated specific economic activities. The majority of respondents indicated to have stopped because of low profitability because of lack of demand. When comparing "reasons for stopping" across the three treatment arms, we find very few significant differences: clients in treatment arm T2 more often indicated to have stopped because of the time required for other activities than clients in other treatment arms.

[21] When we restrict the sample to business that already existed at the baseline, we find even stronger effects (not reported, but available on request from the authors). That is, since we measure impact at the midline and endline, there are both established as well as starting economic activities in the sample. Our results suggest, not surprisingly, that on average the profitability of established activities is higher than the profitability of new activities. Since the training

stimulated the starting up of new activities (so that there are more new firms in the intervention arms than in the control group), our estimates actually *underestimate* the impact of the training on sales and profits of existing business activities.

[22] We have regressed dummies for the uptake and timely repayment of short term loans on treatment dummies (lumping T1 and T2 together, or considering them separately), and find that the treatment dummies do not enter significantly (details available on request from the authors).

References

Armendariz B, Allen T, Karlan D, Mullainathan S (2010) Inviting husbands in women-only solidarity groups: Evidence from Southern Mexico. Working paper, University College London, London.

Banerjee A (2013) Microcredit under the microscope: What have we learned in the last two decades, what do we need to know? *Annual Rev. Econom.* 5:487–519.

Banerjee A, Karlan D, Zinman J (2015) Six randomized evaluations of microcredit: Introduction and further steps. *Amer. Econom. J.: Appl. Econom.* 7(1):1–21.

Berge LIO, Bjorvatn K, Tungodden B (2014) Human and financial capital for microenterprise development: Evidence from a field and lab experiment. *Management Sci.* 61(4):707–722.

Berge LIO, Bjorvatn K, Juniwaty KS, Tungodden B (2012) Business training in Tanzania: From research-driven experiment to local implementation. *J. African Economies* 21(5):808–827.

Bjorvatn K, Tungodden B (2010) Teaching business in Tanzania: Evaluating participation and performance. *J. Eur. Econom. Assoc.* 8(2/3):561–570.

Bloom N, Mahajan A, McKenzie D, Roberts J (2010) Why do firms in developing countries have low productivity? *Amer. Econom. Rev.* 100(2):619–623.

Bruhn M, Zia B (2013) Stimulating managerial capital in emerging markets: The impact of business training for young entrepreneurs. *J. Development Effectiveness* 5(2):232–266.

Bruhn M, Karlan D, Schoar A (2010) What capital is missing in developing countries? *Amer. Econom. Rev.* 100(2):629–633.

Burbidge J, Magee L, Rob AL (1988) Alternative transformations to handle extreme values of the dependent variable. *J. Amer. Statist. Assoc.* 83(401):123–127.

Calderon G, Cunha J, de Giorgi G (2013) Business literacy and development: Evidence from a randomized trial in rural Mexico. NBER Working Paper w19740, National Bureau of Economic Research, Cambridge, MA.

Charness G, Gneezy U (2012) Strong evidence for gender differences in risk taking. *J. Econom. Behav. Organ.* 83(1):50–58.

Charness G, Rustichini A (2011) Gender differences in cooperation with group membership. *Games Econom. Behav.* 72(1):77–85.

Croson R, Gneezy U (2009) Gender differences in preferences. *J. Econom. Literature* 47(2):448–474.

de Mel S, McKenzie D, Woodruff C (2008) Returns to capital in microenterprises: Evidence from a field experiment. *Quart. J. Econom.* 123(4):1329–1372.

de Mel S, McKenzie D, Woodruff C (2009) Are women more credit constrained? Experimental evidence on gender and microenterprise returns. *Amer. Econom. J.: Appl. Econom.* 1(3):1–32.

de Mel S, McKenzie D, Woodruff C (2014) Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka. *J. Development Econom.* 106(January): 199–210.

Drexler A, Fischer G, Schoar A (2014) Keeping it simple: Financial literacy training and rule of thumbs: Evidence from a field experiment. *Amer. Econom. J.: Appl. Econom.* 6(2):1–31.

Duvvury N, Carney P, Nguyen HM (2012) *Estimating the Cost of Domestic Violence Against Women in Viet Nam*, 1st ed. (UN Women Vietnam, Vietnam).

Fafchamps M, McKenzie D, Quinn S, Woodruff C (2011) Female microenterprises and the fly-paper effect: Evidence from a randomized experiment in Ghana. Case study, World Bank, Washington, DC.

Fenwick GD, Neal D (2001) Effect of gender composition on group performance. *Gender, Work Organ.* 8(2):205–225.

Gerber A, Green D (2012) *Field Experiments: Design, Analysis and Interpretation* (Norton, New York).

Giné X, Mansuri G (2014) Money or ideas? A field experiment on constraints to entrepreneurship in rural Pakistan. Policy Research Working Paper 6959, World Bank, Washington DC.

Hoogendoorn S, Oosterbeek H, van Praag M (2013) The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Sci.* 59(7):1514–1528.

International Labour Organization (2010) Report on labour force surveys Vietnam, January 9, 2009. Ministry of Planning and Investment, General Statistics Office Vietnam Ha Noi.

Karlan D, Valdivia M (2011) Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *Rev. Econom. Statist.* 93(2):510–527.

Lee D (2009) Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *J. Amer. Statist. Assoc.* 83(401): 123–127.

McKenzie D, Woodruff C (2014) What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Res. Observer* 29(1):48–82.

Rahman M, Hoque A, Makinoda S (2011) Intimate partner violence against women: Is women empowerment a reducing factor? A study from a national Bangladeshi sample. *J. Family Violence* 26(5):411–420.

Sayinzoga A, Bulte EH, Lensink R (2016) Financial literacy and financial behaviour: Experimental evidence from rural Rwanda. *Econom. J.* 126(594):1571–1599.

Valdivia M (2014) Business training plus for female entrepreneurship? Short and medium-term experimental evidence from Peru. PIERI Working Papers 2014-04, http://ssrn.com/abstract=2510944.

van Velzen R (2014) The impact of business training on business outcomes: A randomized control trial on female microfinance clients in Vietnam. MsC thesis, Faculty of Economics and Business, University of Groningen, Groningen, Netherlands.

Vu NTH, van Velzen R, Lensink R, Bulte E (2015) The impact of gender and business training for female microfinance clients in Vietnam. Grantee final report, International Initiative for Impact Evaluation (3ie), New Delhi.

World Bank (2011) *Vietnam Gender Assessment* (World Bank, Washington DC).