Women's empowerment and domestic abuse: Experimental evidence from Vietnam

Erwin Bulte and Robert Lensink

Erwin Bulte: Development Economics Group, Wageningen University and Utrecht University; email: Erwin.bulte@wur.nl

Robert Lensink (corresponding author): Faculty of Economics and Business, University of Groningen and Development Economics Group, Wageningen University. Email: b.w.lensink@rug.nl

Acknowledgements: we would like to thank 3IE (OW3.1132: Business training services for financial support.

Abstract: Intimate partner violence is an important global health problem that policy makers seek to address by a variety of interventions, including efforts to promote "women's empowerment." We use data from a randomized control trial in Vietnam and find that this strategy may backfire: women who participated in a gender and entrepreneurship training program suffer more frequent abuse than women in the control group. We conjecture that increased female income is the mechanism linking the training program to domestic violence. We also make a methodological contribution and show that the outcomes of our impact analysis depend on how we measure intimate partner violence.

Keywords: spousal violence, gender training, domestic abuse

JEL Codes: D10, D74, I10, J12

1. Introduction

Violence against women is a widespread and pervasive problem. According to recent estimates, one-third of all women have experienced physical or sexual abuse during their lives (Garcia-Moreno et al., 2005; Devries et al., 2013). The most common form of violence against women is intimate partner violence (IPV), which has significant (public) health and societal impacts. In addition to reduced agency of women, which is a human rights problem in itself, these impacts include costs due to poor physical and mental health. Fearon and Hoeffler (2014) estimated that these costs exceed the global economic cost of civil war. Not surprisingly, curbing IPV has emerged as an important policy priority. In this paper, we will study the impact of a business and gender empowerment training program in Vietnam on the incidence of violence against women.

Broadly speaking, two sets of theories try to explain domestic violence. *Expressive* violence theories start from the premise that violence serves the purpose of relieving frustration and anxiety triggered by challenges to traditional gender roles and norms emphasizing male authority. Expressive theories assume that violence enters directly in the perpetrator's utility function (e.g. Farmer and Tiefenthaler, 1997). In contrast, *instrumental* violence theories are predicated on the assumption that violence helps male household members to assert control over scarce household resources, or to influence the behavior of other household members (e.g. Eswaran and Malhotra, 2011). In the theoretical model developed by Tauchen et al. (1991), domestic violence is explained by both expressive and instrumental motives. Depending on the underlying cause of IPV in specific settings, policy makers may consider various responses to mitigate the incidence of abuse (see next section).

The empirical evidence on the effectiveness of anti-IPV interventions is rich, but also ambiguous and heavily skewed toward high-income countries. Early studies established correlations between IPV measures and observational data on employment status, income, credit use, or education (e.g. Hornung et al., 1981; Farmer and Tiefenthaler, 1997; Bloch and Rao, 2002; Heath, 2014). However, interpretation of such correlations is hampered by simultaneity, omitted variables, and selection. To attenuate endogeneity concerns, instrumental variable estimators may be used, leveraging exogenous variation in supply of, and demand for, female labor (Aizer, 2010; Bhattacharrya, et al., 2011; Chin, 2012; Lenze and Klasen, 2017).

A recent series of experimental studies goes one step further, and randomly assigns subjects to interventions aimed at mitigating IPV. Several studies took place in developing countries, and a handful evaluates interventions that aim to promote female employment or entrepreneurship.[1] Pronyk et al. (2006) considered a combined microfinance and HIV training program for poor women in South Africa, finding strong and encouraging results – a 55% reduction in IPV during the study period of 12 months. However, a follow-up analysis demonstrated that the microfinance intervention itself had little effect on IPV, highlighting instead the importance of complementary training sessions and suggesting the existence of synergy effects across intervention modules (Kim et al., 2009). This appears consistent with evidence from a pilot randomized control trial (RCT) in rural Côte d'Ivoire that studied a combination of economic assistance (group savings) and so-called "gender dialogue groups" (Gupta et al., 2013). This study found large effects of a packaged intervention on physical IPV (but not sexual IPV). Green et al. (2015) studied a combination of a training module and

[1] Another type of intervention, transfer programs targeting women as beneficiaries, has received ample attention (for an interesting non-experimental study on transfers and IPV, refer to Bobonis et al. [2013]). Such programs have improved school attendance, health, and nutrition outcomes (Duflo 2012) and also appear to have the potential to affect domestic abuse. Angelucci (2008) studied alcohol-related abuse and found that the effect of transfers varies with the size of the grant. Hidrobo and Fernald (2013) studied the impact of a cash transfer program in Ecuador and found that the impact varies with characteristics of the receiving household, such as education levels of the partners. Hidrobo et al. (2016) used an RCT to compare the effects of cash, vouchers, and food transfers on IPV and document that all transfer modalities are equally effective in curtailing violence – reducing violence from 16% to 10%. In contrast, Peterman et al. (2017) studied the impact of the Zambian unconditional cash transfer program on IPV (measured using a list experiment) and found no evidence of impact.

financial intervention in Uganda. A one-off cash transfer is combined with a business skills training intervention intended to promote entrepreneurship by developing non-farm microenterprises. While this intervention resulted in large employment gains for participants (Blattman et al., 2016), it did not mitigate abuse. Taken together, these studies send mixed signals about the scope to leverage development interventions to tackle domestic violence.

We used an RCT design to empirically explore the impact of a gender and business training program on IPV for a large sample of poor female entrepreneurs in rural northern Vietnam. The subjects in our sample are female micro-entrepreneurs and members of a microfinance organization (and therefore not necessarily representative for women in Vietnam generally). We randomly assigned credit centers of the microfinance organization to a control or experimental arm, and used both survey data and data from a list experiment to gauge the prevalence of IPV 6 months after the training program. We also collected data on income and a proxy for women's bargaining power based on decision-making.[2]

The contribution of this paper to the literature is threefold. First, we use a randomized intervention to contribute to the experimental literature on the effects of female empowerment on domestic violence, focusing on a novel sub-sample of subjects. Second, we make a methodological contribution by reporting diverging results for different approaches to measuring IPV. Specifically, our results suggest underreporting of IPV when using standard survey-based measures of self-reported domestic violence, obscuring the signal-to-noise ratio that is evident from our preferred measure (the list experiment). Third, we demonstrate that the concept of "women empowerment" is broad and that not all of its dimensions co-vary in the same way with IPV. Specifically, promoting economic empowerment appears to have

[2] Participation in decision-making is an often-used but imperfect proxy for autonomy or bargaining power (e.g. Seymour and Peterman, 2018). Heterogeneity in the strength of the association between decision-making and autonomy depends on the domain and whether partners agree on who normally makes decisions, implying that details about measurement and context are important for understanding bargaining power. Taking decisions can also be a burden and it is not evident how decisions map onto underlying preferences (women can take the decisions they believe their husbands would favor).

different effects on IPV than promoting gender equality and more equal intra-household bargaining (even if the two concepts may be related).

If we measure IPV via a direct survey question, we find that access to the training program has a small IPV reducing effect. However, this conclusion is reversed if we use our preferred measurement approach. When using a list experiment to measure IPV, we find that access to the training program has a large and significant *accentuating* effect on domestic violence. Women in the treatment group are much more likely to suffer from abuse than women in the control group. We also find that participating in the training program increases both female income and bargaining power. Additional patterns in our data are consistent with the hypothesis that an increase in female income is the mechanism linking the training program to more frequent abuse. Potential mitigating effects, via an increase in bargaining power due to the training program, are too small to offset the income effect. However, the nature of our data does not allow a full mediation analysis of the mechanisms linking the training program to domestic violence, and alternative theories may also have explanatory power.

The paper is structured as follows. In section 2, we discuss how women empowerment and domestic abuse may hang together, and propose some testable hypotheses. In section 3, we introduce our Vietnamese case study and the intervention we will study. Section 4 presents our data and identification strategy, paying special attention to the measurement of IPV. In section 5, we present reduced-form regression results based on two types of measures of IPV and shed some light on the mechanism linking the training program to abuse. The conclusions and discussion ensue.

2. **Conceptual framework**

The intervention we study in the remainder of this paper aims to achieve two objectives: (i) to raise women's income (*economic empowerment*) and (ii) to promote gender equity within the household and beyond (*social empowerment*). The expectation of the implementing agency was that women empowerment along both dimensions would curb IPB. But the relationship between empowerment and IPV is complex, and it is not evident that promoting economic and social empowerment will necessarily have the same type of impact on the prevalence of domestic abuse.

Consider economic empowerment interventions or efforts to increase the income flow that women bring into the household. Examples include microfinance interventions or cash-, asset-, or voucher-transfer programs targeting women, but also efforts to promote female entrepreneurship. Such interventions are expected to increase female agency. Increasing the flow of resources that women bring into the household may alter their bargaining position because it increases the appreciation for (and value of) women, but it may also improve post-marriage "outside options" – shifting the threat point in cooperative bargaining (World Bank, 2011). A standard household bargaining model predicts that husbands will curb the dishing out of violence so as not to lose their wives and their incomes (e.g. Bueno and Henderson, 2017).[3] But Eswaran and Malhotra (2011) demonstrated that violence does not need to be declining in female reservation utility, assuming that IPV increases male bargaining power within the family.

Observe that the effects of empowerment may also vary with the cultural context, especially with the stigma associated with divorce (the outside option). If cultural norms are such that divorced women are stigmatized and unable to lead a fulfilling life henceforth, then threats to end a marriage are not credible (as in the case studied by, for example, Bloch and

[3] In case economic empowerment is achieved via work outside the house, another channel may be relevant: reduced exposure (e.g. Chin, 2012).

Rao, 2002).[4] If women cannot opt out of their marriage, economic empowerment may invite adverse consequences. For example, conflicts may emerge over the allocation of the additional income, possibly resulting in additional violence due to instrumental reasons – the "extraction effect" (Chin, 2012). In addition, economic empowerment may pose a threat to male roles of dominance and authority, inviting extra expressive violence. Economically empowering women in a context with strong traditional norms may then be counterproductive, resulting in outcomes referred to as "male backlash" (Luke and Munshi, 2011; Bueno and Henderson, 2017).

Next, consider efforts to promote gender equality, for example through education programs aiming to alter attitudes and norms or through engaging groups of men in training and discussion sessions (Ellsberg et al., 2014; Jewkes et al., 2015). More equal status and bargaining power likely reduces tensions and conflict in the household as this may attenuate differences in opinion about how much each partner should contribute to the "household good" and reduce income hiding (Malapit, 2012). Both instrumental and expressive theories generally predict that more gender equality reduces IPV.[5] However, since traditional norms tend to be deeply rooted within local culture, it may not be easy to change them. It is also not evident to what extent gender equality can be promoted by "training" only women. The insights and lessons they bring back to the household may be rejected by their husbands, possibly inviting additional tensions and violence.

The main insight is that the impact of women's empowerment on levels of domestic violence is ambiguous and may vary with the dimension of empowerment and the cultural

[4] If divorce is too costly or not viable, the non-cooperative Cournot–Nash equilibrium within an existing marriage may serve as the only credible threat point (see Lundberg and Pollack, 1993).

[5] However, there are exceptions to this rule. For example, in a context where women can credibly threaten to abandon their husband (low stigma associated with divorce), husbands must respect their wife's participation constraint when making intra-household allocation and domestic abuse decisions. In equilibrium, women must be as well-off married as alone. If enhanced gender equality within the household implies that the wife's utility in marriage improves relative to the outside pay-off, then husbands can decide to "bring back" their wives' well-being to the initial (reservation) level of utility by dishing out more violence.

context. The net effect of interventions seeking to promote several dimensions of the empowerment of women simultaneously is therefore an empirical issue. We now turn to such an example.

3. Female empowerment in Vietnam, and the intervention

Rural Vietnam is a strongly patriarchal society, with Confucian traditions placing high value on female obedience to men (UN Women, 2015). According to a recent article in the Financial Times (2011), divorced women in Vietnam are likely to stay alone for the rest of their lives due to strong stigma effects. While the divorce rate (or the percentage of marriages ending in divorce) is rising, it is still only 1.7%. In our sample of women, the divorce rate is even lower. The reason is that filing for divorce is primarily a matter for well-educated young women in urban areas – not the women in our sample. The reputational damage for women in rural areas is so severe that few women consider divorce a realistic alternative. Men have freedom to set and enforce allocation rules, and IPV is not uncommon in rural Vietnam. Krause et al. (2016) estimated that 31% of all women in Vietnam have experienced physical abuse during their lifetime, and Vung et al. (2008) found that the prevalence during the previous year alone was nearly 10%. Among the Vietnamese population, especially in rural areas, attitudes about gender equity are weakly developed, and the tolerance for IPV is considerable. Tran et al. (2016) argued that almost 40% of the Vietnamese respondents indicated that it is acceptable for men to use violence against their spouse.

We evaluate the impact of a gender and business training program on domestic violence for a sample of female members of the largest microfinance organization in northern Vietnam: the TYM fund. TYM is the largest women's organization in northern Vietnam, whose members are not (necessarily) representative of the north Vietnamese population because he organization targets women who are (i) relatively poor (have incomes below the country average) and (ii) entrepreneurial. All members are engaged in economic activities to earn a living (i.e. have a "micro-enterprise"). TYM seeks to empower its female members – both by helping them to increase their incomes (promoting economic independence) and by raising awareness about gender equity. For full details of the intervention, refer to Bulte et al., (2017) who studied the impact of this training program on a range of business outcomes (including profits). Also see Huis et al. (2017), who considered the training program's impact on female empowerment and found that the training program increased personal control beliefs and relational empowerment.[6]

The training intervention took place between February and October 2012 in two districts: Hanoi and Vinh Phuc. The program is based on the Gender and Entrepreneurship Together (GET) Ahead for Women in Enterprise Training Package and Resource Kit, of the International Labor Organization, and contains nine monthly modules lasting 45 minutes each. The first three training modules are devoted to gender issues and are intended to promote gender equity and to improve the bargaining position of women in the household. The remaining six modules cover standard business subjects, including marketing, financial literacy, cash management, and cost calculations. These modules should help women to turn their business activities into (more) successful endeavors, increasing the flow of income they earn from these activities. Details of the training program content are provided in Appendix A. Training sessions were delivered during monthly meetings at TYM credit centers, with weekly refresher meetings. Participating in the training sessions was voluntary and for free.

Our sample contains 187 credit centers. We stratify randomization by lending branch, and initially assigned centers randomly to one of three experimental arms: (i) 31 credit centers to a treatment arm where only female clients were invited; (ii) 70 credit centers to a treatment

[6] Huis et al. (2017) did not analyze IPV separately, but a survey-based IPV measure featured as one element in an aggregate index containing seven relational frictions items.

arm where male partners were also invited to join the training program; and (iii) 86 credit centers were assigned to the control group (and did not receive training). We purposefully assigned many centers to the treatment arms where husbands were invited to participate because we feared extensive non-compliance among husbands. By increasing the number of centers, we attempted to attenuate power problems due to low attendance of husbands (unsuccessfully, as will become clear). We randomly selected 23 members per center for the interviews, and for those centers with less than 23 clients, we interviewed all members.[7]

While one of the underlying ideas of this study was to explore how the participation of men in the training sessions would affect the outcomes of the intervention, our data do not allow us to explore this question in a satisfactory way. The reason is extensive non-compliance among husbands, who most of the time did not show up at the training sessions. As summarized by Bulte et al. (2017), the participation rate of husbands declined from nearly 40% in the first session to (much) less than 10% in later ones.[8] Nearly half of the invited husbands did not attend any training session, and only 2% of them attended all sessions. Due to extensive male non-compliance, the differences between the two treatment arms are very small, and outcomes for the two sub-samples are statistically very similar. To increase statistical power, we pool the data from the treatment arms and consider only two experimental groups: 101 credit centers receiving training, and 86 credit centers from the control group not receiving training. We demonstrate that our results are robust with respect to distinguishing between the two initial treatment arms.

More than 80% of the invited women attended all training modules, and nearly all women attended some modules. Approximately 90% of the women in the sample belong to

[7] The maximum size of a TYM center is 30 members. On average, there was no difference in sample size between centers assigned to the treatment group and centers assigned to the control group (approximately 21.5 members, on average). A t-test reveals that the small difference is not statistically different from zero.

[8] Attendance was relatively high in the first session because we financially incentivized attendance then.

the Kinh ethnicity (the dominant ethnic group in Vietnam). The remaining 10% belong to a mix of ethnic minorities from a range of geophysical backgrounds, with (potentially) quite different culture and gender norms.[9] Since gender norms may mediate the intervention's impact, lumping different ethnicities (cultures) together may obscure important results for individual ethnicities. This is why the main empirical analysis focuses on the Kinh only. However, we will also present estimates for the full sample of respondents.

4. **Data and identification**

We conducted three waves of data collection. Baseline data were collected in October–November 2011, about three months before the training program started. Midline data were collected in March–May 2013, about six months after the last training module. Finally, endline data were collected in October–November 2013, about one year after the last training module. We collected survey data on IPV data during all three data waves, but only used the list experiment during the midline. We believe that a six-month interval since the final session (and a 15-month interval since the first one) provides the best opportunity for the intervention to have an impact on IPV – this interval is long enough for the business training program to influence business practices and (the expectation of future) business outcomes and hopefully not "too long" for the bargaining impacts to erode due to lack of reinforcement or salience. To make the analyses based on survey question and list experiment comparable, we also use the midline survey question on IPV. In most analyses, we use baseline data as controls.

Table 1 summarizes information about sample size for the three rounds of data collection. During the baseline, a total of 4041 female borrowers were sampled. During the

[9] According to a report by UN Women (2015), ethnic minority women suffer from a much larger education gap than Kinh women and have less access to a wide range of (social protection) services. Many ethnic minorities' livelihoods depend on slash and burn agriculture, characterized by a "rigid division of labor," providing limited opportunities for women to "participate in decision-making" so that "wives rarely contribute their opinion." For discussions on how traditional livelihoods agriculture affect gender roles and the balance of power within households, refer to Alesina et al. (2013).

midline, we were able to re-survey 3511 women from the baseline (87%). The attrition rate of 13% is explained by a small minority of 95 women who did not want to participate in the survey and by the fact that 435 women were not available during data collection. To compensate for attrition between baseline and midline, we surveyed a group of 351 new women (for whom we lack baseline values), increasing the full midline sample to 3862 women.[10] Since not all women answered all questions, sample size varies across the models we estimate.

<< Insert Table 1 about here >>

We did not use any selection criteria, but trained and surveyed a representative sample of TYM members. The organization specifically targets women who have a per capita income below the country's average income and who are interested in running a micro-enterprise. More than 85% of the women in our sample are married, and we will estimate models for the full sample and for the sub-sample of married women separately. The age of our subjects ranges between 19 and 72 years (at baseline), with a mean (median) age of 44 (45) years, while TYM women are of all ages over 18 years old (new members should be between 18 and 65 years old). Additional information about the sample can be obtained by looking at the constants of regression models to test "balance", provided in Tables 3a–c below, but is also summarized in Appendix Table C8.

All data were collected by trained female surveyors. We used two alternative approaches to collect data on IPV six months after completion of the intervention. Our main analysis is based on a so-called list experiment, also known as the item count technique. List experiments seek to elicit truthful answers to sensitive questions by using an indirect

[10] During the last wave of data collection, we surveyed almost everybody included in baseline and midline, resulting in a sample of 4350 women. These data are not used in the current paper. A group of 3412 women were surveyed in all survey waves.

elicitation method, in contrast to direct elicitation approaches that might invite underreporting of sensitive or embarrassing behavior. The advantage of list experiments is that subjects have no incentive to lie as they do not have to admit they have been the victim of IPV.[11] The main drawback is that list experiments enable identification of effects at the level of experimental groups and do not enable the analyst to identify which individual women are victims of abuse.

A standard list experiment works as follows. A sample of respondents is randomly assigned to one of two groups: a "control group" receiving a list of J non-sensitive (or neutral) statements, and a "treatment group" receiving the same statements plus one sensitive item – the one about domestic abuse in our case (a list of J+1 statements). Individuals from both groups are asked to report with how many of the statements they agree, *but not with which particular statements they agree*. Due to the random assignment of the "sensitive statement" to one of the sub-samples, the difference-in-means between the two groups gives an unbiased estimate of the proportion of the sample agreeing with the sensitive issue. Observe that the anonymity of individual women is guaranteed.

One example is from the domain of microfinance. Lenders are interested in how borrowing entrepreneurs use loan proceeds, but borrowers are reluctant to admit using loans for consumptive purposes as this might compromise their loan eligibility or the socially desirable image they seek to project. Karlan and Zinman (2012) indeed find that direct elicitation causes underreporting of the non-enterprise use of loans, compared with the list experiment. A meta-analysis based on 48 comparisons of list experiments and survey questions found that more than 60% of the estimates for socially undesirable behavior were significantly larger when elicited via the list experiment (Holbrook and Krosnick, 2010).

[11] Agüero and Frisancho (2017) discuss four reasons why abused women may be reluctant to expose themselves as a victim and may respond differently to direct elicitation than to a list experiment where their anonymity is guaranteed: (i) emotional cost due to attachment to the offender; (ii) potential loss of economic support by the offender; (iii) the fear of retaliation and escalation of violence; and (iv) stigmatization.

Work by Tsuchiya et al. (2007) and LaBrie and Earleywine (2000) confirms that especially *sensitive* questions, such as concerning shoplifting or unprotected sex, are answered more truthfully. Non-sensitive issues such as blood donations produce non-significant differences between the measurement approaches (see also Blair et al. 2018).

However, not all studies find significant differences between direct and indirect elicitation approaches (Tourangeau and Yan, 2007).[12] Agüero and Frisancho (2017) compared the measurement of IPV in Peru using direct elicitation and a list experiment. They found that the two approaches yield similar measures of IPV *on average*, but also that direct elicitation invites underreporting for the sub-group of educated women. The explanation they propose for this finding is that especially educated women wish to provide socially desirable answers and avoid shame or stigma associated with admitting victimization.

Since our main purpose is to evaluate whether the gender and business training program reduces the incidence of IPV, our design involves stratified randomization: we stratify by experimental arms in the RCT (being invited to take part in the training program, or not). We randomly assign subjects from each treatment arm to one of two sub-groups each: one group receiving questionnaire A (with five statements) and another group receiving questionnaire B (with four statements). The four non-sensitive statements that are common to questionnaire A and B are:

1. "*I have money in a saving account*"
2. "*I prefer local fruits over Chinese fruits*"
3. "*My household does not have a television*"
4. "*I usually buy pears.*"

[12] This need not be surprising. It is difficult to detect sensitive outcomes when they are low prevalence, especially when non-sensitive questions have a lot of variance (e.g. Karlan and Zinman, 2012).

Ideally, responses to non-sensitive questions should not display too much variation, as this reduces precision of the coefficient of interest. However, at the same time, there should be "some variation" across respondents (i.e. not everybody should answer the same way) to preserve confidence among the respondents that their individual response to the sensitive question remains anonymous.

The additional (sensitive) statement included only in questionnaire A reads as follows: "*I am regularly hit by my spouse*." We did not further specify the word "regularly," which implies this statement is rather vague and imprecise. The loose nature of the statement does not follow Demographic and Health Surveys (DHS) recommendations but is in keeping with the style of the non-sensitive statements in the questionnaire. We hope that using the word "regularly" implies that our subjects respond to this statement while considering violence during the post-training period, i.e. during the previous six months.[13] The question that followed both lists of statements was as follows: "*With how many of the statements in the list above do you agree?*"

We also gauge IPV using direct elicitation in the survey. Specifically, we asked each respondent the following sensitive question: "*How often did your husband push, slap, beat, or hit you during the last 6 months?*"[14] Observe that this question is intentionally not phrased exactly the same as the sensitive statement included in the list experiment. We hope this attenuates concerns about one response leading the other. The single survey question is (also) much simpler than the full IPV module used in, for example, DHS. Specifically, the DHS use nine direct questions to measure physical and sexual violence, capturing a range of behaviors including pushing and slapping (as in our question), but also pulling of hair, forced sex, and

[13] Observe that we do not expect a difference between the trained and control group if subjects also consider violence during the pre-intervention period – the analysis would suffer from attenuation bias then, with coefficients biased to zero.

[14] We provided the following categories for answering: 0 = Never; 1 = Rarely; 2 = Sometimes; 3 = Often; 4 = Very often; 99 = Don't know; and 88 = Refused to answer.

other forms of violence. However, lumping together multiple dimensions of IPV in one general question is not uncommon in multi-topic surveys.

The list experiment was part of the survey, and the order in which the list experiment and direct IPV question were presented was fixed: first the list experiment, and later in the survey the direct question. We have no reason to believe that this fixed order affects the choices of the respondents in the two experimental arms differently.

Since centers were randomly assigned to the training intervention in the RCT, and subjects were also randomly assigned to arms in the list experiment, treatment status is orthogonal to subject characteristics. Our main analysis is therefore based on a simple linear reduced-form model:

$$Y_i = \alpha + \beta_1 Training_i + \beta_2 ListA_i + \beta_3 Training_i * ListA_i + \varepsilon_i \qquad (1)$$

In (1), Y_i captures the number of statements that respondent *i* agrees with, *Training* is a dummy variable taking the value of 1 when the respondent was invited to participate in the gender and business training program (and 0 otherwise), and *ListA* is a dummy taking the value of 1 when she received the questionnaire with five statements (and 0 otherwise). Interpretation of the outcomes of the linear model is straightforward. The coefficient of interest is β_3, which takes a negative value if the training program reduces exposure to violence and which takes a positive value if the training program increases exposure to violence. As robustness tests, we will estimate various variants of (1). To increase the precision of our estimates, we control not only for branch dummies but also for the subject's age and marital status and the size of her household in some of the models. In all models, we will cluster standard errors at the training center level to tackle within-center correlation in outcomes.

We also control for boundedness of the response variable by using a non-linear least squares estimator due to Imai (2011) as an additional robustness analysis:

$$Y_i = f(X_i, \gamma) + ListA_i f(X_i, \partial) + \varepsilon_i \quad (2)$$

where $E(\varepsilon_i | X_i, ListA_i) = 0$, (γ, ∂) is a vector of unknown parameters, and X is a vector of observed covariates, including the intercept and the binary indicator *Training*. As before, *Training* = 1 again implies the respondent belongs to the (treatment) group invited to participate in the training program, and *Training* = 0 indicates subjects from the control group. Next, $f(x, \gamma)$ and $g(x, \partial)$, where $x \in X$ are regression models (logit models) for the conditional expectations of control and sensitive statements, given the covariates. The non-linear least squares estimator uses a two-step procedure in which $f(x, \gamma)$ is first fitted to the group receiving four statements and next $g(x, \partial)$ is fitted to the group receiving five statements. Heteroscedasticity-consistent robust standard errors are obtained via method of moments (as in Imai [2011]). This model implies the linear multivariate regression model with treatment-by-covariate interaction terms (as in Holbrook and Krosnick [2010]) if the two sub-models are assumed to be linear. That is:

$$Y_i = X_i^T \gamma + ListA_i X_i^T \partial + \varepsilon_i \quad (3)$$

Finally, and as outlined below, we also seek to open the “black box” and document whether additional patterns in the data are consistent with our predictions about the mechanisms linking the training intervention to domestic violence, or not. Additional information about the measurement of female income and bargaining power is provided in Appendix B.

5. **Empirical results**

5.1 *Evaluating the design*

A condition for unbiased measurement with list experiments is the absence of so-called design effects – answers to control statements should not be affected by the inclusion of the sensitive statement.[15] Blair and Imai (2012) developed a likelihood ratio test to examine whether the no-design effect assumption is violated. This test uses the assumption that the addition of a sensitive statement should not change the sum of affirmative answers to the control statements, and that (cumulative) proportions of different respondent types should be non-negative.[16] If one of the proportions is negative, the no-design effect assumption is violated. The test statistic indicates whether negative values arise by chance, in which case we do not reject the null hypothesis of no-design effects. We calculate the proportions of each respondent type, assuming no-design effects, as follows:

$$\begin{aligned} \pi_{y1} &= \Pr\left(Y_i \leq y \middle| T_i = 0\right) - \Pr(Y_i \leq y | T_i = 1), \\ \pi_{y0} &= \Pr\left(Y_i \leq y \middle| T_i = 1\right) - \Pr(Y_i \leq y - 1 | T_i = 0) \end{aligned} \tag{4}$$

This amounts to subtracting the cumulative proportions of the observed data from the same rows for π_{y1} and cumulative proportions of consecutive rows for π_{y0}. For our data, outcomes are reported in Table 2.

<< *Insert Table 2 about here* >>

For one respondent type, the proportion is negative. Following Blair and Imai (2012), we use the *list* package in R to test for design effects. The *p*-value of the test equals 0.96 for the full sample and 0.64 for the Kinh sample, so we cannot reject the null of no-design effects.

[15] Note that the no-design effect hypothesis does not require respondents to truthfully answer the control items.
[16] Note that this refers to the two groups randomly determined in terms of the list experiment; it does not refer to the groups randomly assigned to the training.

We also need to consider the possibility that respondents might lie about the sensitive statement. Lying is particularly likely to occur in case of *ceiling* or *floor* effects. A ceiling effect occurs if subjects agree with all control statements, and a floor effect occurs if subjects agree with none of the control statements. In either case, affirmative responses to the sensitive statement are easy to identify and it is difficult to protect the privacy of respondents. However, for our list experiment, the potential of *floor* or *ceiling* effects is small. In our control group of the full sample (Kinh sample) of 1922 (1699) respondents, only 1.98% (1.59%) agreed with zero statements and only 0.16% (0.18%) agreed with all control statements. The modal response was just below two.

5.2 Balancing tests and attrition

Before presenting our regression results, we test whether there is balance across the groups receiving and not receiving the sensitive statement in the list experiment. We also conduct balancing tests for respondents invited to participate in the training program and respondents from the control group in the RCT. The balance test is based on baseline data for the control variables: age of the female TYM member (*Age*); marital status (*Married*); household size (*Members*), and a regional dummy for Vinh Phuc (*Region*). In addition, we consider whether at baseline the following variables were balanced: number of loans from TYM (*Credit access*); intra-household bargaining (*Bargain*); a dummy for agricultural activities (*Agr*); land size in agriculture in acres (*Land size*); an index for business knowledge (*Knowledge*); our observational measure of IPV (*IPV survey*); and a measure of last month's profits (*Profit*). Results for the groups in the list experiment are given in Tables 3a and 3b below, based on models where we regress baseline values on the *ListA* dummy and a constant.

<< Insert Tables 3a, 3b, and 3c about here >>

These tables suggest balance across the two sub-groups, as the mean for the group with the sensitive statement does not significantly differ from the mean of the group responding to only non-sensitive statements – *ListA* is insignificant in all regressions.

In Table 3c we present balance tests for the training intervention. The table shows that baseline means of the treatment group are not significantly different from baseline means of the control group.[17] The number of subjects in these balance test deviates somewhat from the number of subjects in Tables 3a and b. The reason is attrition in the midline, when we conducted the list experiment (limiting the sample for which we have baseline data, analyzed in Tables 3a and b).[18]

<< Insert Table 4 about here >>

In Table 4, we report the outcomes of models explaining the determinants of attrition – subjects dropping from our sample after the baseline. To explain attrition, we use the same variables as in the balance tests above, and also interact these variables with the treatment dummy to consider differences in attrition across treatment arm (T1 and T2). There is no difference in attrition probability across treatment arms (column 2). As is evident, however, attrition is not purely random either. It is negatively correlated with age and marital status, as well as with the number of loans with TYM. There is also differential attrition across treatment arms. Women who followed the training program and suffered from high (survey-based) IPV at baseline are less likely to drop out of the sample. To reduce potential bias due

[17] We have also estimated a model that compares the four groups simultaneously: treatment list A, treatment list B, control list A, and control list B. We regressed the same baseline variables in Tables 3a–c on *ListA*, the training dummy, and the interaction of these variables. Again, this balance test does not indicate any reason for concern (details available on request).

[18] Consider, for instance, the number of observations for *Age* in Table 3c, which equals 3804. This number is made up as follows. We sampled 4041 women, which is the sum of the remaining women from the initial sample plus the additional women we included to compensate for attrition. Of this sample, 137 women did not answer the "Age" question, leaving us with 3904 observations. The total number of observations for *Age* in Tables 3a and 3b equals only 3374. These are the remaining women from the baseline resampled during the midline (3511), minus the number of women that did not answer the "Age" question (137): 3511 - 137 = 3374.

to non-random attrition, we control for baseline values in most models that we estimate and also provide estimates based on attrition probability weighting. Our main results are not affected by non-random attrition – see below.

5.3 Women empowerment and domestic abuse

We now report our main reduced-form regression results. Column (1) of Table 5 summarizes the results of the linear regression model explaining variation in the number of affirmative answers to the statements in the list experiment. Column (2) does the same, but here we also add a vector of baseline controls. The results are almost similar.

<< Insert Table 5 about here >>

The results of the linear model are relatively easy to interpret. Consider column (1). On average, respondents from the control group agreed with 1.89 statements from the set of control statements. Respondents from the treatment group agreed with slightly fewer control statements, namely 1.75 (or 1.89 - 0.14). This divergence in outcomes is not really surprising, as earlier analyses found that participating in the training program encourages investments in economic activities (Bulte et al., 2017) and therefore affects the probability that households have (residual) money in a saving account. Inclusion of the training dummy therefore controls for the fact that treated women agree with fewer non-sensitive items than control women.[19]

More interesting are the coefficients associated with the sensitive statement. From the control group, 11% of the members agree with the statement that they are "regularly hit" by their husband. For women from the treatment group, this share increases to 21%: this

[19] Chuang et al. (2017) tested the internal validity of list tests and pointed to the observation that different lists of non-sensitive items within the same sample lead to different estimates of the prevalence of the sensitive behavior. This might point to the problem that the number of "true" statements that are non-sensitive has an influence on compliance to the tool and its reliability. If so, this could introduce estimation bias. However, in the present study, the difference in the number of non-sensitive statements with which the respondents agrees across experimental groups is small (0.14 statements).

indicates that the training program induced additional violence. The effect is not only statistically significant; it is also meaningful: women participating in the training program are 10 percentage points more likely to report being the victim of regular abuse, amounting to an increase of 90% compared with the control group. This result contrasts with Huis et al. (2017) who find that, based on survey based data, the training program decreased relational frictions between the spouses.

It is not so easy to compare these numbers with those in the existing literature. As mentioned earlier: DHS-type estimates suggest that slightly more than 30% of the Vietnamese women have been exposed to IPV during their lives. However, we consider the issue of regular abuse (as opposed to, possibly, incidental abuse), include women from a broader age range, and focus on a non-representative sample of entrepreneurial women who have joined a microfinance organization and have their own business.

In column (3) of Table 5, we introduce the additional treatment dummy *T2* to distinguish between the two sub-samples of women whose husbands were and were not invited to participate in the training program. Importantly, *Training* still takes the value 1 for respondents from all centers assigned to treatment (with and without invitations extended to husbands), so *T2* captures the additional effect of inviting husbands. Treatment dummy (*Training*) remains significant when interacted with the *ListA* variable. Reflecting the extensive non-compliance among invited husbands, we find the dummy variable associated with one of the sub-groups is small and not significantly different from zero, and neither is the interaction of *T2* and *ListA*. Also, observe that the coefficient of the latter interaction term is very small. Estimating regression models with separate dummies for the two sub-groups produces the outcome that the associated coefficients are not significantly different from each other (according to a Wald test).

Columns (4) and (5) present additional robustness analyses. Column (4) is based on the sub-sample of women living together with a male (omitting widows, divorcees, and single women from the full sample). Some 96% of these women are married and the rest live with another person; we obtain similar results if we only include married women with a male household head. As is evident, the results are virtually unchanged for this sub-sample. In column (5), we enlarge the sample by introducing women from different ethnic groups. We again introduce controls and now also interact these controls with the ethnic dummy. Pooling all ethnic groups does not affect our main result. The coefficient of interest, *Non-Kinh*ListA*Training*, is not significantly different from zero, suggesting the results for non-*Kinh* women are comparable with those of *Kinh* women (but the sample of non-*Kinh* women is small and we may be underpowered to pick up meaningful differences).

Our regression results are rather robust with respect to alternative modifications. For example, in column (6) we present the results of a model based on attrition probability weighting for the sample of Kinh women. In the probit model that is used to study attrition we include a full vector of interaction terms of treatment status and baseline controls (as in column 3 of Table 4). In column (7), we report the results of the non-linear model. These results are qualitatively consistent with the earlier findings.

A threat to unbiased identification of the treatment effects reported in Table 5 emerges if the training program affects the accuracy of measurement – a case of non-classical measurement error. If the content of the training program raises women's knowledge about IPV, then this may affect their response to the statement about abuse: "*I am regularly hit by my spouse*."[20] This would occur if trained women interpret the word "*regularly*" differently than women from the control group, or if trained and untrained women have different

[20] This is akin to the familiar question whether business trainings affect the accuracy of reporting sales and profits. The empirical literature is divided on the relevance of this concern (e.g. De Mel et al., 2009; Drexler et al., 2014).

perceptions about what it means "*to be hit.*" Specifically, if treated women are more (less) likely to answer affirmatively, all else equal, then we will overestimate (underestimate) the true accentuating effect of the training program. However, this concern is attenuated by the fact that the training program did not contain any content focusing on domestic violence – trained women do not have additional knowledge about this issue.[21]

A related concern is social desirability bias, which would bias our estimates of impact *downward* if trained women are inclined to report fewer cases of abuse. Such bias might materialize if treated respondents are more aware of the social undesirability of IPV than women from the control group and are ashamed to admit victimization. As explained earlier, however, we believe that such bias is more relevant for our survey-based measure of IPV, discussed below, where responses can be linked to individual respondents. The same is true for another potential challenge to proper identification, namely that members of the control group may strategically overstate their exposure to IPV to qualify for future support. The rationale for such strategic responses is diminished in the list experiment. Nevertheless, we cannot eliminate all concerns regarding reporting bias.

Finally, observe that the "vagueness" of the IPV statement in the list experiment may affect the response of *all* women in the list experiment, in which case it would bias our estimate of β_2 (rather than the coefficient of interest, β_3).

In Appendix C, we report the regression outcomes of nearly all models for the full sample of female respondents – including the ethnic minority women (non-*Kinh*) with potentially different norms of gender equity. As a result, the coefficients of interest take a slightly smaller value but remain statistically significant in most models (albeit at the 10%

[21] Ideally, we would have followed up on this statement by assessing whether women from the treatment and control groups feel differently about how "normal" or "acceptable" it is for men to be violent toward their wife. Unfortunately, we lack the data to do this.

level). This finding is consistent with attenuation bias introduced by respondents with different gender norms (levels of stigma) and, hence, diverging responses to the training program.

We next analyze the data on domestic violence obtained through direct elicitation – using the single simple IPV question. Interestingly, 92% (or 96%) of the respondents from the control (or treatment) group now indicates "never" to be "pushed, slapped, beat, or hit" by their husband during the previous 6 months. No respondents were unwilling or unable to answer the question, and nobody answered that they "did not know" the answer. Admittedly, it is difficult to compare the answers to our IPV survey question with those in the literature. We use a single direct question, rather than a validated module consisting of multiple questions, and we focus on experiences during the previous 6 months, as opposed to the previous year's or lifetime exposure to IPV.

Vung et al. (2008) found that the prevalence of IPV during the previous year was almost 10%, which seems not inconsistent with our finding that 8% of the control group (4% of the treatment group) has "rarely" or "sometimes" experienced IPV during the previous 6 months. However, the finding that *not a single client* from the treatment or control group indicated themselves to be the victim of abuse "often" or "very often" during the previous six months appears rather incredible, especially in light of the finding that 11% (or 21%) of the women from the control (or treatment) group admits to "regular" victimization in the list experiment. Recall: these are the same individuals, responding to the survey question and participating in the list experiment. Consistent with earlier findings reported by Karlan and Zinman (2012) and others (e.g. Holbrook and Krosnick, 2010), this suggests underreporting of socially undesirable behavior when using direct elicitation, for example due to experimenter demand effects. We believe that our data support the conjecture that respondents are prone to misreport about sensitive issues in surveys and are more likely to reveal outcomes truthfully when their anonymity is guaranteed.[22]

<< Insert Table 6 about here >>

Not surprisingly, patterns in the survey-based data differ from the ones found in the experimental data. This is evident from Table 6, which summarizes regression models suggesting that trained women are abused *less* frequently by their husbands. This is true for the various sub-samples considered above (*Kinh* women, married *Kinh* women, all ethnicities), and this finding essentially reverses the results of the list experiment. As discussed above, this may be explained by experimenter demand effects, which are relevant for the (individually traceable) survey question to a much greater extent than the anonymous list experiment. Of course, the survey-based analysis may also suffer from a lack of statistical power, because so few surveyed women admit to being abused. Observe that the samples used for the survey-based IPV analyses diverge slightly from the samples used for the list-based IPV analyses. This is because some women did not participate in the list experiment and a slightly larger number of women refused to answer the IPV question in the survey.[23]

5.4 "Mediation analysis": The mechanisms linking the training program to IPV

We next probe the mechanisms linking participation in the training program and the prevalence of IPV. Full-fledged mediation analysis, as outlined by Barron and Kenny (1986),

[22] To further probe the validity of this idea, we tried to follow up on work by Aqüero and Frisancho (2017) who compare IPV measurements based on a list experiment and direct elicitation (DHS style). Like us, they consider the IPV prevalence result emanating from the list experiment as the "benchmark measure" because anonymity minimizes the cost of being exposed as a victim. Surprisingly, on average, they do not find significant differences in prevalence rates for their sample of Peruvian women. Additional analysis revealed a gap between measurement approaches for the category of "educated women," who may suffer most from shame about admitting victimization. To explore this issue for our sample, we divided our sample in two sub-samples, based on the median education level. We indeed find higher prevalence of IPV among "educated" women when measured with a list experiment but not with the survey question. However, the difference is not statistically significant.

[23] A common sample of 3322 women both participated in the list experiment and answered the IPV survey question. Re-estimating the IPV models for this sub-sample with controls gives us nearly identical results as the ones reported in columns (2) of Tables 5 and 6 (not separately reported).

is rare in economics, and Deaton (2010) observes that treatments in most economic analyses are best interpreted as a "black box." One reason why mediation analysis is controversial in economics is that it is difficult to assess the plausibility of the so-called "sequential ignorability assumption" – the assumption that there is no confounder affecting both the outcome (IPV) and mediator variables (income and bargaining power). Unobserved heterogeneity implies that we should be careful with causal claims with respect to the mechanism.

The mediation variables of interest are female income and bargaining power. We may of course treat these mediating variables as dependent variables, and explain variation in (post-training) income and bargaining power by the training program. We therefore first estimate the following two well-identified models:

$$Profit_i = \alpha + \beta Training_i + \gamma X_i + \varepsilon_i, \text{ and} \tag{5a}$$

$$Bargaining\ power_i = \alpha + \beta Training_i + \gamma X_i + \varepsilon_i. \tag{5b}$$

We used the same survey instrument (and the same sample) to collect information on the two mediating variables: *profits* and *bargaining power*. Our measure of profits (or income derived from the business activity undertaken by the wife) is based on a direct estimate of last month's total profits. To deal with outliers and zero-responses, we use the inverted hyperbolic sine transformation of this variable. As a measure of bargaining power, we construct an index capturing how often the wife is involved in decisions regarding certain expenditures (such as consumer durables, financial assistance to family members, and so on). Please refer to Appendix B for additional details about these variables. As before, *Training* is a dummy indicating whether the respondent was invited to participate in the training intervention, *X* is a vector of controls, and we cluster standard errors at the center level.

Regression results for the two mediating variables are reported in Table 7.

<< Insert Tables 7 and 8 about here >>

The main thing to observe is that – as intended – the training program increases both female income and bargaining power. The magnitude of these causal effects is meaningful, consistent with the theory, and in line with earlier results reported in Bulte et al. (2017) and Huis et al (2017). Specifically, we find that participating in the training program causes a 0.33 standard deviation increase in profits and a 0.17 standard deviation increase in our index of bargaining power. Importantly, the standardized effect size on profits is twice as large as the effect size on bargaining power. This reflects that, while gender norms may be malleable, they are more resistant to change than economic variables such as income.

Next, we move to more treacherous academic terrain and consider the second link in the narrative: increases in income and female bargaining power may invite, respectively, higher and lower levels of IPV. We do not have random variation in the mediating variables, so this part of the analysis is not experimental but based on observational data.[24] We simply document the association between our mediating variables and the prevalence of IPV, and examine whether these associations are consistent with expectations. Using data from the list experiment (column 1) and the survey-based IPV measure (column 2), we estimate linear models with controls. Non-linear models or omitting the controls gives similar results (not reported). Regression results are reported in Table 8. We should compare the coefficients of the interaction terms in column 1 (*profits*×*ListA* and *bargaining*×*ListA*) with the level terms in column 2 (*profits* and *bargaining*).

[24] We also cannot assume that the training is a valid instrument for either mediating variable, as the exclusion restriction is likely violated.

Both models suggest an increase in female income is associated with more domestic abuse: the interaction between *Profits* and *ListA* is positive in the list experiment and *Profits* is positive (but very small) for our direct IPV question. Both models also yield that greater female *bargaining powe*r is associated with less abuse (again; see the relevant interaction term in column 1 and the level term in column 2). Interestingly, the significance level of the channels varies across the models. The model based on the list experiment data suggests the income channel is significant, and the effect via bargaining power is not. The model based on survey data suggests the opposite. This is consistent with the opposing reduced form impacts, summarized in Tables 5 and 6.

6. **Discussion and conclusion**

We analyze the complex relationship between female empowerment and domestic violence. IPV is a widespread and persistent problem adversely affecting the lives of hundreds of millions of women and posing a first-order challenge for policy makers worldwide. Since violence against women is often regarded as a symptom of lack of gender equity, it seems natural to assume that efforts to promote female empowerment will help to reduce IPV. Indeed, many governments and NGOs pursue such an agenda by both promoting economic independence of women and increasing their bargaining power.

In this paper, we demonstrate that such strategies may be counterproductive in a context with strong stigma associated with divorce. We speculate that two common dimensions of female empowerment – economic independence and bargaining power – may have complex and opposite effects on the prevalence of domestic violence. Specifically, while increasing the bargaining power of women within the household, or their say in how to spend household resources, tends to attenuate IPV, the reverse may be true when the income that women bring into the household increases. The latter effect may seem counterintuitive but can be explained with both instrumental and expressive theories of domestic abuse. It is also in line with a review by Vyas and Watts (2009). In reality, the situation is even more complex, as income and bargaining power are likely to be inter-linked in household models.

We analyze the impact of a gender and business training program for a sample of female entrepreneurs in rural Vietnam. We developed an RCT in which a random sub-sample of women benefitted from a training intervention that aimed to promote gender equity and helped them to scale up and professionalize their micro-enterprises. Most importantly, for analyses based on IPV data collected with a list experiment, we obtain the robust result that participating in the training program invited additional domestic abuse. Our survey-based IPV data suggest the reverse effect. We believe the discrepancy in results based on survey-based data and list experiment data can be explained by experimenter demand effects. Anonymity and credible deniability guaranteed by the list experiment implies that we believe these data are more credible. Consistent with the stated objectives and theory of change of the implementing organization, we also find that participating in the training program indeed resulted in higher incomes and a greater say in major spending decisions of the household. Moreover, the income effect is much stronger and seems to dominate the bargaining power effect – the standardized effect size is twice as large.

Identifying the causal mechanism linking the training program to IPV is not straightforward. The training program may have affected several other factors associated with IPV, such as expressive motives for violence and the time allocated to housework (during the training program, and afterwards). Distinguishing between competing channels requires additional data and perhaps additional experiments (e.g. experiments with a placebo treatment for women in the control group, reducing their time for housework during the intervention). A key limitation of the current study is that we do not have a design where respondents were (exogenously) exposed to either the business module (economic empowerment), the gender quality module (social empowerment), or a combination of the two. Therefore, we cannot identify the effects of these dimensions separately, and the reduced form treatment effects that we identify therefore are the net effect of potentially opposing underlying effects.

Importantly, we do not wish to convey the message that increasing female income will always invite more intense levels of abuse. This depends on the context, including the viability of the "exit option" for married women. In particular, this depends on stigma associated with divorce. In conservative societies, such as in rural Vietnam, divorce hardly occurs and women typically cannot credibly threaten to abandon their husband. This is the context where promoting women's empowerment through economic independence appears especially risky. It is interesting to observe that Hidrobo et al. (2016) obtain results, based on data from Ecuador, that appear opposite to ours. In Ecuador, divorce rates are also very low (perhaps because of the influence of the Catholic Church), yet Hidrobo et al. (2016) find that, on average, transfers to women reduce IPV. However, it is important to observe that the divorce to marriage ratio in Ecuador is some 50% higher than in Vietnam. Moreover, the *separation rate* is also higher. From the perspective of our model, the distinction between divorce and separation is moot; hence, it is possible that predictions with respect to the effect of empowerment interventions on the prevalence of IPV are opposite.

Acknowledgments

We thank seminar participants in Wageningen, Essen, Geneva, Oxford, Amsterdam, Heidelberg, and Gottingen for feedback and suggestions. We also thank two anonymous referees for helpful comments and suggestions. 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. Remaining errors are our own.

References

Aizer, A. 2010. The gender wage gap and domestic violence. Am. Econ. Rev.100, 1847–1859.

Alesina, A., Giuliano, P., Nunn, N. 2013. On the origins of gender roles and the plough. Q. J. Econ. 128, 469–529.

Angelucci, N. 2008. Love on the rocks. B E J. Econ. Anal. Policy 8, DOI: https://doi.org/10.2202/1935-1682.1766.

Agüero, J., Frisancho, V. 2017. Misreporting in sensitive health behaviors and its impact on treatment effects: an application to intimate partner violence. IDB Working Paper Series RG-K1462.

Barron, R., Kenny, D. 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations. J. Pers. Soc. Psychol. 51, 1173–1182.

Bhattacharya, M., Bedi, A., Chhachhi, A. 2011. Marital violence and women's employment and property status: evidence from north Indian villages. World Dev. 39, 1676–1689.

Blair, G., Imai, K. 2012. Statistical analysis of list experiments. Polit. Anal. 20, 47–77.

Blair, G., Chou W., Imai, K. 2018. List experiments with measurement error. Polit. Anal., forthcoming. Available at: https://imai.princeton.edu/research/files/listerror.pdf.

Blattman, C., Green, E., Jamison, J., Lehman, M., Annan, J. 2016. The returns to cash and microenterprise support among the ultra-poor: a field experiment in post-war Uganda. Am. Econ. J.: Appl. 8, 35–64.

Bloch, F., Rao, V. 2002. Terror as a bargaining instrument: a case study of dowry violence in rural India. Am. Econ. Rev. 92, 10291042.

Bobonis, G., Gonzalez-Brenes, M., Castro, R. 2013. Public transfers and domestic violence: the roles of private information and spousal control. Am. Econ. J.: Econ. Policy 5, 179–205.

Bueno, C.C., Henderson, E.A. 2017. Bargaining or backlash? Evidence on intimate partner violence from the Dominican Republic. Femin. Econ. 23, 90–116.

Bulte, E., Lensink, R., Vu, N. 2017. Do gender and business trainings affect business outcomes? Experimental evidence from Vietnam. Manage. Sci., 63: 2885-2902

Chin, Y.M. 2012. Male backlash, bargaining or exposure reduction? Women's working status and physical spousal violence in India. J. Popul. Econ. 25, 175–200.

Chuang, E., Dupas, P., Huillery, E., Seban, J. 2017. Sex, lies and measurement. Working paper. Available from: https://web.stanford.edu/~pdupas/CDHS_measurement.pdf.

Deaton, A. 2010. Instruments, randomization, and learning about development. J. Econ. Lit. 48, 424–455.

De Mel, S., McKenzie, D., Woodruff, C. 2009. Measuring microenterprise profits: must we ask how the sausage is made? J. Dev. Econ. 88, 19–31.

Devries, K., Mak, J., Garcia-Moreno, C., et al. 2013. The global prevalence of intimate partner violence against women. Science 340, 1527–1528.

Drexler, A., Fischer, G., Schoar, A. 2014. Keeping it simple: financial literacy training and rule of thumbs: evidence from a field Experiment. Am. Econ. J.: Appl. Econ. 6, 1–31.

Duflo, E. 2012. Women empowerment and economic development. J. Econ. Lit. 50, 1051–179.

Ellsberg, M., Arango, D.J., Morton, M., et al. 2014. Prevention of violence against women and girls: what does the evidence say? Lancet 385,1555–1566.

Eswaran, M., Malhotra, N. 2011. Domestic violence and women's autonomy in developing countries: theory and evidence. Can. J. Econ. 44, 1222–1263.

Farmer, A., Tiefenthaler, J. (1997). An economic analysis of domestic violence. Rev. Soc. Econ. 55, 335–358.

Fearon, J., Hoeffler, A. 2014. Peaceful, stable, and resilient societies. Technical Report prepared for Copenhagen Consensus Center.

Financial Times. 2011. Vietnam sees more women filing for divorce. Available from: https://www.ft.com/content/60466878-d39a-11e0-bc6b-00144feab49a.

Garcia-Moreno, C., Jansen, H.A.F.M., Ellsberg, M., Heise, L., Watts, C. 2005. WHO Multi-country Study on Women's Health and Domestic Violence Against Women. Geneva, Switzerland: World Health Organization.

Green, E., Blattman, C., Jamison, J., Annan, J. 2015. Women's entrepreneurship and intimate partner violence: a cluster randomized trial of microenterprise assistance and partner participation in post-conflict Uganda. Soc. Sci. Med. 133, 177–188.

Gupta, J., Falb, K., Lehmann, J., et al. 2013. Gender norms and economic empowerment intervention to reduce intimate partner violence against women in rural Côte d'Ivoire: a randomized controlled pilot study. BMC Int. Health Hum. Rights 13, 46.

Heath, R. 2014. Women's access to labor market opportunities, control of household resources, and domestic violence: evidence from Bangladesh. World Dev. 57, 32–46.

Hidrobo, M., Fernald, L. 2013. Cash transfers and domestic violence. J. Health Econ. 32, 304–319.

Hidrobo, M., Peterman, A., Heise, L. 2016. The effect of cash, vouchers, and food transfers on intimate partner violence: evidence from a randomized experiment in northern Ecuador. Am. Econ. J.: Appl. Econ. 8, 284–303.

Holbrook, I.L., Krosnick, J.A. 2010. Social desirability bias in voter turnout reports: tests using the item count technique. Public Opin. Q. 74, 37–67.

Hornung, C., McCullough, B., Sugimoto, T. 1981. Status relationships in marriage: risk factors in spouse abuse. J. Marriage Fam. 43, 675–692.

Huis, M., Lensink, R., Vu, N., Hansen, N. 2017. Teaching gender equality: impacts of a gender and business training on empowerment among female microfinance borrowers in Northern Vietnam. Groningen, Netherlands: University of Groningen.

Imai, K. 2011. Multivariate regression analysis for the item count techniques. J. Am. Stat. Assoc. 106, 407–416.

Jewkes, R., Flood, N., Lang, J. 2015. From work with men and boys to changes of social norms and reduction of inequities in gender relations: a conceptual shift in prevention of violence against women and girls. Lancet 385, 1580–1589.

Karlan, D.S., Zinman, J. 2012. List randomization for sensitive behavior: an application for measuring use of loan proceeds. J. Dev. Econ. 98, 71–75.

Kim, J., Ferrari, G., Abramsky, T., et al. 2009. Assessing the incremental effects of combining economic and health interventions: the image study in South Africa. Bull. World Health Organ. 87, 824–832.

Krause, K., Gordon-Roberts, R., van der Ende, K., Schuler, S., Yount, K.M. 2016. Why do women justify violence against wives more often than do men in Vietnam? J. Interpers. Violence 31, 3150–3173.

LaBrie, J., Earleywine, M. 2000. Sexual risk behaviors and alcohol: higher base rates revealed using the unmatched count technique. J. Sex Res. 37, 321–326.

Lenze, J., Klasen, S. 2017. Does women's labor force participation reduce domestic violence? Evidence from Jordan. Fem. Econ. 23, 1–29.

Luke, N., Munshi, K. 2011. Women as agents of change: female income and mobility in India. J. Development Econ. 94, 1–17.

Lundberg, S., Pollak, R.A. 1993. Separate spheres bargaining and the marriage market. J. Pol. Econ. 101, 988–1010.

Malapit, H. 2012. Why do spouses hide income? J. Socio-Econ. 41, 584–593.

Peterman, A., Palermo, T., Handa, S., Seidenfeld, D.; Zambia Child Grant Program Evaluation Team 2017. List randomization for soliciting experience of intimate partner violence: application to the evaluation of Zambia's unconditional child grant program. Health Econ. Lett. 27, 622–628. DOI: 10.1002/hec.3588.

Pronyk, P., Hargeaves, J.R., Kim, J.C., et al. (2006). Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: a cluster randomised trial. Lancet 368, 1973–1983.

Seymour, G., Peterman, A. 2018. Context and measurement: an analysis of the relationship between intrahousehold decision making and autonomy. World Dev. 111, 97–112.

Tauchen, H.V., Witte, A.D., Long, S.K. 1991. Domestic violence: a non-random affair. Int. Econ. Rev. 32, 491–511.

Tourangeau, R., Yan, T. 2007. Sensitive questions in surveys. Psychol. Bull. 133, 859–883.

Tran, T.D., Nguyen, H., Fisher, J. 2016. Attitudes towards intimate partner violence against women among women and men in 39 low- and middle-income countries. PLoS ONE November 28, 1–14.

Tsuchiya, T., Hirai, Y., Ono, S. 2007. A study of the properties of the item count technique. Public Opin. Q. 71, 253–272.

UN Women. 2015. Social protection for women and girls in Viet Nam. Hanoi: United Nations Entity for Gender Equality and the Empowerment of Women. Viet Nam Country Office.

Vung, N.D., Ostergren, P.-O., Krantz, G. 2008. Intimate partner violence against women in rural Vietnam – different socio-demographic factors are associated with different forms of violence: need for new intervention guidelines? BMC Public Health 8, 55.

Vyas, S., Watts, C. 2009 How does economic empowerment affect women’s risk of intimate partner violence in low and middle income countries? A systematic review of published evidence. J. Int. Dev. 21, 577–602.

World Bank. 2011.World Development Report 2012: Gender equality and development. Washington DC: World Bank.

Table 1. The sample of women.

	Kinh sample	Total sample
Baseline	3805	4041
Midline	3374	3862
Endline	3799	4350

Table 2.

Proportions of respondent types, Kinh sample.

	Observed data				Estimated proportion of respondent types			
Y-value	Control counts	Control proportion (cumul)	Treatment counts	Treatment proportion (cumul)	π_{yo}	SE (π_{yo})	π_{y1}	SE (π_{y1})
0	27	1.59 (1.59)	30	1.79 (1.79)	1.79	0.32	-0.20	0.44
1	479	28.19 (29.78)	364	21.77 (23.56)	21.97	1.08	6.22	1.52
2	980	57.68 (87.46)	976	58.37 (81.94)	52.16	1.45	5.52	1.24
3	210	12.36 (99,82)	213	12.74 (94.68)	7.22	0.97	5.14	0.56
4	3	0.18 (100)	88	5.26 (99.94)	0.12	0.12	0.06	0.06
5			1	0.06 (100)				
Total	1699		1672		83.26		16.74	

Notes: Y-value: indicates amount of answers respondent agrees with; *Control counts:* how many people in control group agree with *Y-value* answers, where control group refers to group of female confronted with four statements; *Control proportion (cumul):* the proportion (cumulative proportion) of female in control group that agrees with *Y-value* answers; *Treatment counts:* how many people in treatment group agree with *Y-value* answers, where *treatment* group refers to group of female confronted with five statements; *Treatment proportion (cumul):* the proportion (cumulative proportion) of female in treatment group that agrees with *Y-value* answers.

$$\pi_{y1} = \Pr\left(Y_i \le y \middle| T_i = 0\right) - \Pr(Y_i \le y | T_i = 1),$$

$$\pi_{y0} = \Pr\left(Y_i \le y \middle| T_i = 1\right) - \Pr(Y_i \le y - 1 | T_i = 0)$$

where *T = 1* refers to treatment (five statements) and *T = 0* refers to control (four statements). Thus π_{yo} equals *Treatment proportion (cumul) Y-value - Control proportion (cumul) Y-value*$_{-1}$ and π_{y1} equals *Control proportion (cumul) Y-value - Treatment proportion (cumul) Y-value.* SE refers to standard error.

Table 3a. Balance tests treatment group (invited to the training program), Kinh sample.

Variable	*Age*	*Married*	*Members*	*Region*	*Credit access*	*Bargain*
List	-0.01	-0.02	0.01	-0.02	0.01	0.08
_cons	43.79***	0.87***	4.76***	0.28***	1.14***	4.13***
Observations	1817	1817	1817	1817	1817	1817
R-squared	0.00	0.00	0.00	0.00	0.00	0.00

Notes: List refers to a dummy indicating groups with (1) and without (0) sensitive questions. *Age:* Age of the female TYM member; *Married:* marital status; *Members:* household size; *Region:* a regional dummy for Vinh Phuc; *Credit access:* number of loans from TYM; *Bargain:* intra-household bargaining. p-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 3a (cont.) Balance tests treatment group (invited to the training program), Kinh sample.

Variable	*Agr*	*Land size*	*Knowledge*	*Obs IPV*	*Profit*
List	0.01	10.97	-0.03	0.00	-0.10
_cons	0.81***	1506.78***	8.98***	0.06***	2.78***
Observations	1817	1817	1817	1813	1817
R-squared	0.00	0.00	0.00	0.00	0.00

Notes: List refers to a dummy indicating groups with and without sensitive question. *Agr:* a dummy for agricultural activities; *Land size:* land size in agriculture in acres; *Knowledge:* an index for business knowledge; *Obs IPV:* our observational measure of intimate partner violence (IPV); *Profit:* inverted hyperbolic profits in previous month. p-values based on cluster robust standard errors. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 3b. Balance tests control group (not invited to the training program), Kinh sample.

Variable	*Age*	*Married*	*Members*	*Region*	*Credit access*	*Bargain*
List	-0.86	-0.02	0.02	-0.02	-0.02	-0.02
_cons	44.55***	0.87***	4.76***	0.29***	1.18***	4.22***
Observations	1557	1557	1557	1557	1556	1557
R-squared	0.00	0.00	0.00	0.00	0.00	0.00

Notes: List refers to a dummy indicating groups with (1) and without (0) sensitive questions. *Age:* Age of the female TYM member; *Married:* marital status; *Members:* household size; *Region:* a regional dummy for Vinh Phuc; *Credit access:* number of loans from TYM; *Bargain:* intra-household bargaining. p-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 3b (cont.)
Balance tests control group (not invited to the training program), Kinh sample.

Variable	*Agr*	*Land size*	*Knowledge*	*Obs IPV*	*Profit*
List	0.001	-64.59	-0.03	0.00	0.55
_cons	0.82***	1539.60***	9.11***	0.08***	2.26***
Observations	1556	1557	1557	1548	1557
R-squared	0.00	0.00	0.00	0.00	0.00

Notes: List refers to a dummy indicating groups with (1) and without (0) sensitive questions. *Agr:* a dummy for agricultural activities; *Land size:* land size in agriculture in acres; *Knowledge*: an index for business knowledge; *Obs IPV:* our observational measure of IPV; *Profit:* inverted hyperbolic profits in the previous month. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0$.

Table 3c
Balance tests training program, Kinh sample.

Variable	*Age*	*Married*	*Members*	*Region*	*Credit access*	*Bargain*
Treatment	-0.38	0.00	-0.05	-0.01	-0.02	-0.04
_cons	44.04***	0.86***	4.77***	0.27***	1.15***	4.20***
Observations	3804	3805	3804	3805	3804	3805
R-squared	0.00	0.00	0.00	0.00	0.00	0.00

Notes: Treatment refers to a dummy indicating groups with (1) and without (0) access to the training program. *Age*: Age of the female TYM member; *Married:* marital status; *Members:* household size; *Region:* a regional dummy for Vinh Phuc; *Credit access:* number of loans from TYM; *Bargain:* intra-household bargaining. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 3c (cont.)
Balance tests training program, Kinh sample.

Variable	*Agr*	*Land size*	*Knowledge*	*Obs IPV*	*Profit*
Treatment	-0.01	5.75	-0.10	-0.03	0.07
_cons	0.82***	1498.83***	9.08***	0.08***	2.64***
Observations	3804	3805	3805	3791	3805
R-squared	0.00	0.00	0.00	0.00	0.00

Notes: Treatment refers to a dummy indicating groups with (1) and without (0) access to the training program. *Agr:* a dummy for agricultural activities; *Land size:* land size in agriculture in acres; *Knowledge:* an index for business knowledge; *Obs IPV:* our observational measure of IPV; *Profit:* inverted hyperbolic profits in the previous month. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 4 Attrition tests, Kinh sample, with and without interaction effects.

Variable	Model 1	Model 2	Model 3
Treatment	-0.01		0.04
Age	-0.01*	-0.01*	0.00
Married	-0.16*	-0.17*	-0.20
Members	-0.03	-0.03	0.03
Region	-0.19	-0.20	-0.20
Credit access	-0.18***	-0.18***	-0.19**
Bargain	-0.02	-0.01	-0.01
Agr	-0.04	-0.04	0.02
Land size	0.00	0.00	0.00
Knowledge	0.01	0.01	-0.02
Obs IPV	0.12	0.13*	0.24**
Profit	0.00	0.00	0.01
T1		-0.09	
T2		0.15	
TMarried			0.12
TAge			0.00
TMembers			-0.11**
TRegion			0.04
TCredit access			0.02
TBargain			-0.01
TAgr			-0.11
TLand size			0.00
TKnowledge			0.05
TObsIPV			-0.29*
TProfit			-0.01
_cons	-0.44*	-0.44*	-0.51
Observations	3787	3787	3787

Notes: Treatment refers to a dummy indicating groups with (1) and without (0) access to the training program. *Age:* Age of the female TYM member; *Married:* marital status; *Members:* household size; *Region:* a regional dummy for Vinh Phuc; *Credit access:* number of loans from TYM; *Bargain:* intra-household bargaining; *Agr:* a dummy for agricultural activities; *Land size:* land size in agriculture in acres; *Knowledge:* an index for business knowledge; *Obs IPV:* our observational measure of IPV; *Profit:* inverted hyperbolic profits in the previous month. T1: A zero-one dummy indicating the group of females who have been invited to follow the training program with their husbands; *T2:* a zero-one dummy indicating the group of females who have been invited to follow the training program without husbands A *T* before a variable means that this variable is interacted with *Treatment.* * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 5
Gender and business training programs and IPV (list experiment).

Variable	Base linear model	Linear with controls	With additional treatment	Married women only	All ethnicities	Attrition probability weighting	Non-linear model
ListA	0.11**	0.11**	0.11**	0.10**	0.11**	0.09**	-2.07***
Training	-0.14*	-0.16**	-0.12	-0.16**	-0.16**	-0.15***	-0.15***
*Training*ListA*	0.10*	0.11*	0.13*	0.11*	0.11*	0.12**	0.78*
T2			-0.12				
*T2*ListA*			-0.05				
NK					-0.41		
*NK*ListA*					0.42**		
*NK*Training*					0.11		
*NK*ListA*Training*					0.38		
Constant	1.89***	2.34***	2.34***	2.24***	2.34***	2.29***	-0.11***
Controls	No	Yes	Yes	Yes	Yes	Yes	No
Observations	3371	3315	3315	2861	3448	3300	3371
R-squared	0.02	0.04	0.05	0.04	0.04	0.04	

Notes: The dependent variable indicates the number of statements in the list experiment that the participant agrees with. *Training:* a zero-one dummy indicating who has been invited to follow the training program; *T2:* a zero-one dummy indicating the group of females who have been invited to follow the training program without husbands. *ListA:* a dummy indicating which women are confronted with the sensitive statement. *NK:* a zero-one dummy for ethnicity other than Kinh. Controls are baseline values for marital status of the respondents, knowledge, land size, amount of schooling years, and a score on cognitive ability. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 6
Gender and business training programs and IPV (direct questioning), Kinh sample.

Variable	Base linear model	Linear with Controls	With additional treatment	Married women only	All ethnicities	Attrition probability weighting
Training	-0.04*	-0.04*	-0.04*	-0.05**	-0.04**	-0.03***
T2			-0.01			
Constant	0.08***	0.12**	0.12**	0.13**	0.13**	0.11***
Controls	No	Yes	Yes	Yes	Yes	Yes
Observations	3325	3270	3270	2860	3402	3255
R-squared	0.01	0.02	0.02	0.02	0.02	0.02

Notes: The dependent variable indicates exposure to IPV elicited by a single direct question. *Training:* a zero-one dummy indicating who has been invited to follow the training program; *T2:* a zero-one dummy indicating the group of females who have been invited to follow the training program without husbands. Controls are baseline values for marital status of the respondents, knowledge, land size, amount of schooling years, and a score on cognitive ability. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 7.

Impact of the training program on profits and bargaining power, Kinh sample.

Variable	PTProfits	Bargaining power
Training	2.16***	0.25**
Constant	2.62**	5.22***
Controls	Yes	Yes
Observations	3312	3285
R-squared	0.04	0.18

Notes: Dependent variable *PTProfits* refers to the (post-training) inverse hyperbolic sine transformation of last month total profits, and *Bargaining power* is an (post-training) index capturing women's bargaining power (see Appendix B). *Training* refers to a dummy indicating groups with (1) and without (0) being invited to the training program. Controls are baseline values for marital status of the respondents, knowledge, land size, amount of schooling years, and a score on cognitive ability. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 8

Profits, bargaining power, and IPV, Kinh sample.

Variable	List experiment	Survey-based IPV
ListA	0.16**	
Profits	0.01**	0.00
Bargaining power	0.02	-0.01**
Bargaining × ListA	-0.01	
Profits × ListA	0.01**	
Constant	2.13***	0.15**
Controls	Yes	Yes
Observations	3282	3237
R-squared	0.05	0.02

Notes: The dependent variable in column 2 (List) indicates the number of statements in the list experiment that the participant agrees with. The dependent variable in column 3 (Direct) indicates exposure to IPV elicited by a single direct question. *Profits* refers to the inverse hyperbolic sine transformation of last month total profits; *Bargaining power:* intra-household bargaining; *ListA:* a dummy indicating which women are confronted with the sensitive statement. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Appendix A

The training program consists of nine training modules: see Table A1 for an overview. Three modules focus on gender issues: (1) gender beliefs, roles, prejudice, and gender equality; (2) female entrepreneurs' business skills, confidence, and identifying successful business goals; and (3) difficulties and challenges for women in doing business and setting up a cooperation. The six remaining modules focused on business-related topics: (4) identifying and selecting business ideas and opportunities; (5) the importance of product, price, promotion, and place in marketing and business sales; (6) calculating interest rates and the possibility of saving; (7) opening and managing cash books; (8) managing account receivable and account payable books; and (9) calculating purchasing and production costs and costs of sold goods.

Table A1

Modules of the GET Ahead Training Package.

Module 1: Gender and gender equality
Module 2: The business woman and her self-confidence
Module 3: The business woman and her environment
Module 4: The business project: Business ideas
Module 5: The business project: Marketing and how to sell with success
Module 6: The business project-finance: Calculations and how to calculate interest rate
Module 7: The business project-finance: Managing cash
Module 8: The business project-finance: How to record accounts receivable and accounts payable
Module 9: The business project: How to calculate costs of production and cost of goods sold

GET stands for Gender and Entrepreneurship Together.

Appendix B

Intermediate variables used in the analysis:

(1) Bargaining power:

The bargaining power index is taken from Huis et al. (2017). They constructed a bargaining power index for decision making on large and small expenditures. We took the index for large expenditures, which is constructed by summing scores on seven questions (answered by women):

Who makes most decisions about asking for a loan?

Who makes most decisions about consumer durable items? (TV, fridge, tape recorder, etc.)

Who makes most decisions about what health expenditures to make?

Who makes most decisions about saving for business and for household?

Who makes most decisions about expenses for home purchase, improvement or repair?

Who makes decisions about where to invest surplus money?

Who makes decisions about how to assist family members?

Huis et al. (2017) assigned 0 points for each decision made by the husband, 0.5 points for each decision made by the couple together, and 1 point for each decision made by the wife, then summed these points for all relevant items.

(2) Profits

The indicator for profits is taken from Bulte et al. (2017), who distinguish between various indicators for profits. We used the inverse hyperbolic sine transformation of last month total profits, which attenuates the importance of outliers and accommodates zero-responses (for details, see Bulte et al. [2017]).

Appendix C: Tables for full sample (pooling all ethnicities)

Table C1

Proportions of respondents types, entire sample.

	Observed data				Estimated proportion of respondent types			
Y-value	Control counts	Control proportion (cumul)	Treatment counts	Treatment proportion (cumul)	π_{yo}	SE (π_{yo})	π_{y1}	SE (π_{y1})
0	38	1.98 (1.98)	38	2.00 (2.00)	2.00	0.32	-0.02	0.45
1	539	28.04 (30.02)	410	21.58 (23.58)	21.60	1.02	6,44	1.43
2	1123	58.43 (88.45)	1116	58.74 (82.32)	52,29	1.36	6.13	1.14
3	219	11.39 (99,84)	241	12.68 (95.00)	6.55	0.88	4,84	0.51
4	3	0.16 (100)	94	4.95 (99.95)	0.10	0.10	0.05	0.05
5			1	0.05 (100)				
Total	1922		1900					

Notes: Y-value: indicates amount of answers respondent agrees with; *Control counts:* how many people in control group agree with *Y-value* answers, where control group refers to group of female confronted with four statements; *Control proportion (cumul):* the proportion (cumulative proportion) of female in control group that agrees with *Y-value* answers; *Treatment counts:* how many people in treatment group agree with *Y-value* answers, where *treatment* group refers to group of female confronted with five statements; *Treatment proportion (cumul):* the proportion (cumulative proportion) of female in treatment group that agrees with *Y-value* answers.

$$\pi_{y1} = \Pr\left(Y_i \le y \middle| T_i = 0\right) - \Pr(Y_i \le y | T_i = 1),$$

$$\pi_{y0} = \Pr\left(Y_i \le y \middle| T_i = 1\right) - \Pr(Y_i \le y - 1 | T_i = 0)$$

where T = 1 refers to treatment (five statements) and $T=0$ refers to control (four statements). Thus π_{yo} equals *Treatment proportion (cumul) Y-value - Control proportion (cumul) Y-value$_{-1}$* and π_{y1} equals *Control proportion (cumul) Y-value - Treatment proportion (cumul) Y-value.* SE refers to standard error. p = value test: 0.96. Hence, we cannot reject the null of no design effects.

Table C2a
Balance tests treatment group (invited to the training program), entire sample.

Variable	*Age*	*Married*	*Members*	*Region*	*Credit access*	*Bargain*
List	0.04	-0.02	-0.01	-0.02	0.01	0.07
_cons	43.76***	0.85***	4.77***	0.28***	1.13***	4.06***
Observations	1893	1893	1888	1893	1893	1893
R-squared	0.00	0.00	0.00	0.00	0.00	0.00

Notes: List refers to a dummy indicating groups with (1) and without (0) sensitive questions. *Age:* Age of the female TYM member; *Married:* marital status; *Members:* household size; *Region:* a regional dummy for Vinh Phuc; *Credit access:* number of loans from TYM; *Bargain:* intra-household bargaining. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table C2a (cont.)
Balance tests treatment group (invited to the training program), entire sample.

Variable	*Agr*	*Land size*	*Knowledge*	*Obs IPV*	*Profit*
List	0.01	15.49	-0.04	-0.01	-0.11
_cons	0.79***	1475.75***	8.93***	0.06***	2.74***
Observations	1893	1893	1893	1889	1893
R-squared	0.00	0.00	0.00	0.00	0.00

Notes: List refers to a dummy indicating groups with (1) and without (0) sensitive questions. *Agr:* a dummy for agricultural activities; *Land size:* land size in agriculture in acres; *Knowledge*: an index for business knowledge; *Obs IPV:* baseline IPV measure based on direct survey question; *Profit:* inverted hyperbolic profits in the previous month. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table C2b
Balance tests control group (not invited to the training program), entire sample.

Variable	*Age*	*Married*	*Members*	*Region*	*Credit access*	*Bargain*
List	-0.71	-0.02	0.02	-0.02	-0.02	-0.03
_cons	44.47***	0.85***	4.74***	0.28***	1.16***	4.13***
Observations	1617	1618	1616	1618	1617	1617
R squared	0.00	0.00	0.00	0.00	0.00	0.00

Notes: See Table C2a

Table C2b (cont.)
Balance tests control group (not invited to the training program), entire sample.

Variable	*Agr*	*Land size*	*Knowledge*	*Obs IPV*	*Profit*
List	0.01	-58.90	-0.03	0.00	0.55
_cons	0.80***	1501.01***	9.06***	0.07***	2.18***
Observations	1616	1618	1618	1608	1618
R-squared	0.00	0.00	0.00	0.00	0.00

Notes: See Table C2a (cont.)

Table C2c
Balance tests training program, entire sample.

Variable	*Age*	*Married*	*Members*	*Region*	*Credit access*	*Bargain*
Treatment	-0.38	-0.01	-0.04	0.00	-0.02	-0.05
_cons	43.98***	0.82***	4.76***	0.26***	1.11***	4.05***
Observations	4035	4041	3943	4041	4037	4037
R-squared	0.00	0.00	0.00	0.00	0.00	0.00

Notes: See Table C2a

Table C2c (cont.)
Balance tests training program, entire sample.

Variable	*Agr*	*Land size*	*Knowledge*	*Obs IPV*	*Profit*
Treatment	-0.01	5.81	-0.10	-0.02	0.09
_cons	0.79***	1436.30***	8.99***	0.08***	2.53***
Observations	4036	4041	4041	4022	4041
R-squared	0.00	0.00	0.00	0.00	0.00

Notes: See Table C2a (cont.)

Table C3

Attrition tests, entire sample.

Variable	Attrition1	Attrition 2
Treatment	-0.02	
Age	-0.01*	-0.01*
Married	-0.05	-0.06
Members	-0.03	-0.03
Region	-0.20*	-0.21*
Credit access	-0.16***	-0.16***
Bargain	0.02	0.02
Agr	0.01	0.01
Land size	0.00	0.00
Knowledge	0.01	0.01
Obs IPV	0.11	0.13*
Profit	0.00	0.00
T1		-0.10
T2		0.14
Cons	-0.82***	-0.82***
Observations	3924	3924

Notes: Treatment refers to a dummy indicating groups with (1) and without (0) access to the training program. *Age*: Age of the female TYM member; *Married:* marital status; *Members:* household size; *Region:* a regional dummy for Vinh Phuc; *Credit access:* number of loans from TYM; *Bargain:* intra-household bargaining; *Agr:* a dummy for agricultural activities; *Land size:* land size in agriculture in acres; *Knowledge:* an index for business knowledge; *Obs IPV:* baseline IPV measure based on direct survey question; *Profit:* inverted hyperbolic profits in the previous month. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table C4

Gender and business training programs and IPV (list experiment), entire sample.

Variable	Base linear model	Linear with controls	With additional treatment	Attrition probability weighting	Non-linear model
ListA	0.13**	0.12**	0.12**	0.10***	-1.92***
Training	-0.15**	-0.15**	-0.12	-0.15***	-0.15***
*Training*ListA*	0.08[a]	0.10*	0.12*	0.11**	0.59[a]
T2			-0.11		
*T2*ListA*			-0.06		
Constant	1.88***	2.25***	2.26***	2.26***	-0.12***
Observations	3822	3448	3448	3425	3822
R-squared	0.02	0.04	0.05	0.04	

Notes: The dependent variable indicates the number of statements in the list experiment that the participant agrees with. *Training:* a zero-one dummy indicating who has been invited to follow the training program; *T2:* a zero-one dummy indicating the group of females who have been invited to follow the training program without husbands. *ListA:* a dummy indicating which women are confronted with the sensitive statement. Controls are baseline values for marital status of the respondents, knowledge, land size, amount of schooling years, and a score on cognitive ability. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$; [a] $p < 0.15$.

Table C5

Gender and business training programs and IPV (direct questioning), entire sample.

Variable	Base linear model	Linear with Controls	With additional Treatment	Attrition probability weighting
Training	-0.04*	-0.04**	-0.04*	-0.04***
T2			-0.01	
Constant	0.08***	0.13**	0.12**	0.13***
Controls	No	Yes	Yes	Yes
Observations	3772	3402	3402	3379
R-squared	0.01	0.02	0.02	0.02

Notes: The dependent variable is the answer to the direct survey question about IPV. *Training:* a zero-one dummy indicating who has been invited to follow the training program; *T2:* a zero-one dummy indicating the group of females who have been invited to follow the training program without husbands. Controls are baseline values for marital status of the respondents, knowledge, land size, amount of schooling years, and a score on cognitive ability. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table C6
Impact of the training program on profits and bargaining power, entire sample.

Variable	Profits	Bargaining power
Training	2.19***	0.25**
Constant	1.94*	4.75***
Controls	Yes	Yes
Observations	3445	3415
R-squared	0.05	0.17

Notes: Dependent variable *Profits* refers to the inverse hyperbolic sine transformation of last month total profits; *Bargaining power:* an index capturing women's bargaining power (see Appendix B). *Training:* a dummy indicating groups with (1) and without (0) being invited to the training program. Controls are baseline values for marital status of the respondents, knowledge, land size, amount of schooling years, and a score on cognitive ability. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; ***$p < 0.01$.

Table C7
Profits, Bargaining power, and IPV, entire sample.

Variable	*List*	*Survey*
ListA	0.18**	
Profits	0.01**	$0.01*10^{-1}$
Bargaining power	0.01	-0.01**
*Bargaining*ListA*	-0.01	
*Profits*ListA*	0.01**	
Constant	1.71***	0.10***
Controls	No	No
Observations	3782	3732
R-squared	0.03	0.01

Notes: The dependent variable in column 2 (*List*) indicates the number of statements in the list experiment that the participant agrees with. The dependent variable in column 3 (*Survey*) indicates exposure to IPV elicited by a single direct question. *Profits* refers to the inverse hyperbolic sine transformation of last month total profits; *Bargaining power:* intra-household bargaining; *ListA:* a dummy indicating which women are confronted with the sensitive statement. *p*-values based on cluster robust standard errors; * $p < 0.10$; ** $p < 0.05$; *** $p < 0$.

Table C8
Descriptive statistics.

Variable	N	Mean	SD	Min	Max
Age	4035	43.77447	10.32927	19	72
Married	4041	0.818609	0.385389	0	1
Members	3943	4.73751	1.557565	1	15
Region	4041	0.260579	0.439005	0	1
Credit access	4037	1.100322	0.701896	0	2
Bargain	4037	4.026133	1.549361	0	7
Agr	4036	0.780228	0.414143	0	1
Land size	4041	1439.437	1116.407	0	7200
Knowledge	4041	8.936402	1.719188	0	14
OBS IPV	4022	0.067877	0.306795	0	3
Profit	4041	2.571529	7.200472	-14.0215	13.67446