



USAID
FROM THE AMERICAN PEOPLE



Impact Evaluation of the Women's Leadership in Small and Medium Enterprises Activity in India

October 2017 (Revised March 6, 2019)

This publication was produced at the request of the United States Agency for International Development for the E3 Analytics and Evaluation Project. It was prepared independently by Management Systems International, A Tetra Tech Company.

ABSTRACT

This report presents findings, conclusions, and lessons learned from an impact evaluation of the Women's Leadership in Small and Medium Enterprises (WLSME) activity in India. The evaluation used a quasi-experimental design, comparing four randomly assigned treatment groups to a matched comparison group across four outcome measures: business growth, entrepreneurial leadership, networks, and business knowledge and practices. The activity targeted 210 cashew-processing microenterprises and SMEs owned and managed by women in the Panruti block of Tamil Nadu State. After a baseline survey was conducted on eligible applicants, the sample was randomly assigned to four treatment groups. To avoid contamination, the comparison group was not randomized from the same sample as the treatment arms. Instead, it was randomly selected from a list of similar SMEs located in a geographically removed part of the Panruti block. The endline survey took place in May and June 2017, six months after the WLSME activity ended. The short-term findings are encouraging, primarily the entrepreneur leadership outcomes. Whereas most outcomes related to business growth, networks, and business knowledge and practices do not show a statistically significant link to the WLSME activity, the data show significant effects of sizable magnitude with respect to entrepreneur leadership.

IMPACT EVALUATION OF THE WOMEN'S LEADERSHIP IN SMALL AND MEDIUM ENTERPRISES ACTIVITY IN INDIA

October 26, 2017 (Revised March 6, 2019)

Contracted under AID-OAA-M-13-00017

E3 Analytics and Evaluation Project

Prepared By:

Alberto Chong, Principal Investigator
Irene Velez, Technical Manager

Photo Caption: Executive committee members of the Women Cashew Entrepreneurs Network, which was established under the WLSME India activity to strengthen collective bargaining power and increase information flow and knowledge sharing.

Credit: Irene Velez, MSI

DISCLAIMER

The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

CONTENTS

Acknowledgments	iii
Acronyms	iv
Executive Summary	v
Activity Description	v
Evaluation Design.....	v
Findings.....	vii
Conclusions.....	viii
Lessons Learned.....	ix
Introduction	1
Activity Overview	1
WLSME Initiative.....	1
India WLSME Activity	2
Development Hypotheses	2
Evaluation Questions	3
Outcome Measures.....	4
Gender Aspects of the Questions.....	4
Evaluation Design	4
Data Analysis Methods.....	6
Evaluation Limitations.....	7
Endline Sample	8
Balance Among Treatment and Control Groups	9
Primary Question Findings – Overall Impact	10
Business Growth.....	10
Entrepreneurial Leadership	12
Networks.....	16
Business Knowledge and Practices	16
Secondary Questions Findings – Component Analysis	18
Business Growth.....	31
Entrepreneurial Leadership	31
Networks.....	33
Business Knowledge and Practices	34
Conclusions	34
Lessons Learned	35
Annex A: Evaluation Statement of Work	36
Annex B: Literature Review	47
Annex C: Empirical Methods	49
Annex D: Data Collection and Quality Assurance	52
Annex E: Team Composition	54
Annex F: References	56

FIGURES:

Figure 1: Theory of Change	3
Figure 2: India Evaluation Design	5

TABLES:

Table 1: Survey Sample Distribution	8
Table 2: Focus Group Sample Distribution	9
Table 3: Endline Balance Test	10
Table 4: Business Growth Outcomes.....	11
Table 5: Firm Size Outcome	12
Table 6: Entrepreneurial Leadership Outcomes	13
Table 7: Networks Outcomes.....	16
Table 8: Business Knowledge and Practices Outcomes	17
Table 9: Component A vs. Component ABC Impacts	20
Table 10: Component B vs. Component ABC Impacts	24
Table 11: Component C vs. Component ABC Impacts.....	28
Table 12: Firm Size Outcomes by Component.....	31

ACKNOWLEDGMENTS

The evaluation team would like to thank the women from Cuddalore District who shared their time to participate in this study, Market Xcel for conducting the surveys, and CARE India for its collaboration throughout the study. The team would also like to thank Angelo Cozzubo for his statistical support, Swarna Rajagopalan for facilitating the focus group discussions and conducting the qualitative analysis, and Sudaroli Ramasamy for her transcription and translation support. Additional thanks to Gary Glass and Jacob Patterson-Stein for their technical reviews, and Jennifer Crawford and Toni Melnyk for copy editing. The evaluation team benefited from the shared thoughts and experiences of the WLSME Community of Practice, including the WLSME Peru team (GRADE/Sector 3 Social Venture Group) and the WLSME Kyrgyzstan team (ACDI/VOCA). Finally, the evaluation team would like to thank Anastasia de Santos (USAID/E3/TRR), Wade Channell (USAID/E3/GenDev), and Bhavani Pathak (USAID/E3/PLC) for their continued support of this study, and FHI 360 for its work on the original evaluation design and baseline data.

ACRONYMS

E3	Bureau for Economic Growth, Education, and Environment (USAID)
EB	Entropy Balancing
FGD	Focus Group Discussion
IEC	Information, Education, and Communication
LIBA	Loyola Institute of Business Administration
OLS	Ordinary Least Squares
PSM	Propensity Score Matching
SME	Small and Medium Enterprise
TRR	Office of Trade and Regulatory Reform (USAID/E3)
USAID	United States Agency for International Development
WLSME	Women's Leadership in Small and Medium Enterprises

EXECUTIVE SUMMARY

This is the final report for the impact evaluation of the Women's Leadership in Small and Medium Enterprises (WLSME) activity in India, which was commissioned by the Office of Trade and Regulatory Reform (TRR) in the United States Agency for International Development's Bureau of Economic Growth, Education, and Environment (USAID/E3). This impact evaluation consists of a quasi-experimental design to test how the WLSME interventions affect women and their businesses across four sets of indicators: business growth, entrepreneurial leadership, business knowledge/practices, and social/business networks. This report summarizes the data collection and analysis methods used, and provides initial findings and conclusions based on post-intervention data analysis of the WLSME India activity. Although the initial evaluation design planned to collect multiple follow-up rounds of survey data, USAID decided not to conduct additional data collection after the first follow-up round. Thus, this report contains the short-term findings (hereafter referred to as the endline) observed six months after the WLSME activity in India ended in December 2016.

The short-term findings are encouraging, primarily, the entrepreneur leadership outcomes. Whereas the majority of outcomes related to business growth, networks, and business knowledge and practices do not show a statistically significant link to the WLSME activity, the data show significant effects of sizable magnitude with respect to entrepreneur leadership.

Activity Description

USAID's WLSME initiative aims to address women's under-representation in the small and medium enterprise (SME) sector by funding and testing innovative interventions to increase the entry and growth of women-owned and women-managed SMEs in the developing world. The WLSME India activity directly addressed three critical barriers: (i) agency constraints, which impede adequate accumulation of human capital and managerial capital; (ii) relationship constraints, which limit women's access to information and reduce opportunities to build and draw on social capital; and (iii) structural constraints, which place gender-specific barriers that limit the presence and success of women entrepreneurs.

CARE India, in partnership with its sub-grantee, the Loyola Institute of Business Administration (LIBA), implemented the WLSME activity in India between September 2014 and December 2016. The activity targeted 210 cashew-processing microenterprises and SMEs owned and managed by women in the Panruti block of Tamil Nadu State, and aimed to promote women's leadership and the sustainable growth of these SMEs. The activity was implemented in separate components: (A) human capital, which included a semi-structured curriculum of business and skills trainings on different topics, peer learning, one-to-one mentoring, and guest speakers; (B) information and social capital, which established a network of women entrepreneurs in cashew processing to facilitate linkages with value chain actors, strengthen collective bargaining power, and increase information flow and knowledge sharing among women; and (C) structural constraints, which used information, education, and communication materials to support discussion of gender roles within the business, household, and community. The materials were used for joint counseling with spouses and to engage service providers to strengthen support services for women.

Evaluation Design

The WLSME India activity is based on the development hypothesis that if women business owners had greater human capital, social capital, and access to market information, and more supportive external structures, they would be more likely to grow their businesses and become entrepreneurial leaders. This impact evaluation aims to test this hypothesis with a quasi-experimental design. Eligible applicants to

the WLSME activity were randomly assigned to four treatment arms. Three of the treatment arms each received access to only one of the activity components, while the fourth treatment arm received access to all three activity components. To avoid contamination and spillover into the comparison group, assignment to the comparison group was not randomized from the same sample as the treatment arms. Instead, the comparison group was randomly selected from a list of similar SMEs, identified by CARE India, and located in a geographically removed part of the Panruti block. Since randomization was done at the individual level, there is potential for contamination across the treatment arms. The women assigned to the different treatment arms came from the same area and may have known each other. Therefore, women in one treatment group could have gained indirect access to WLSME activities in other treatment groups, as beneficiaries shared information learned in their own group.

The key evaluation questions to be answered are:

1. Primary Question (combined impact of treatment vs. control): Compared to a propensity score-weighted sample of participants from the control region, **do study participants who were assigned to one of the four treatment arms have higher mean values on the following post-intervention outcomes: entrepreneurial leadership, business growth, business knowledge/practices, and social/business networks?**
2. Secondary Question (separate estimates across treatment arms): Compared to participants in each of the other treatment arms (Component A, Component B, and Component C), **do participants who were randomly assigned to receive the combination of components of the intervention (Component ABC) have higher mean values on the same set of outcomes listed under the primary question?**

Sample size at baseline, conducted between August 2013 and June 2014, consisted of 210 participants in the treatment groups and 53 women in the comparison group. The endline survey took place in May and June 2017, six months after the end of the activity, with a 95 percent response rate. The endline sample consisted of 202 participants in the treatment groups, and 48 women in the comparison group. In addition, the evaluation team increased the size of the comparison group with 100 new women who own or manage a cashew-processing SME. This was to improve the matching requirements for the empirical methods employed, and data efficiency in the subsequent follow-up rounds. This new sample was randomly selected from the same villages as the original comparison group. In addition, eight focus groups with a total of 57 WLSME participants were conducted to explore specific opinions and experiences with the activity in greater depth, as well as produce narratives that address the continuity of personal experiences over time.

Several design and implementation issues make the empirical methods for this quasi-experimental evaluation more complex than would be needed under a pure experimental design. The small sample sizes of each treatment arm, the non-randomized comparison group, and the inability to use the baseline data¹ led the evaluation team to select different statistical methods to present a more robust analysis. The team answered both the primary and secondary questions using three empirical methods: difference in means, propensity score matching with a kernel estimator, and entropy balancing.

The issues described above (small sample size, non-randomized comparison group, and inability to use the baseline) apply to both the primary and secondary questions, and may have resulted in unbalanced, biased samples. Given these issues, the empirical methods are useful techniques to account for all observable differences between the groups. However, the estimates obtained should not be interpreted causally, but merely as a sign of association between the activity components and outcome variables.

¹ The baseline data were not labeled by treatment arm, and did not contain any personally identifiable information. Hence, it was not possible to link the survey observations to specific respondents so as to merge baseline and endline data.

Findings

Regarding the primary question (overall impact), the short-term findings indicate a statistically significant link between the WLSME activity and outcomes related to entrepreneur leadership, but not to outcomes related to business growth, networks, or business knowledge and practices. More specifically, the WLSME activity shows an impact on the following outcomes for female entrepreneurs who had access to any of the WLSME activities:

Business growth:

- A small decrease (about one-third of a month less) in the number of good months for the business in the previous year.

Entrepreneurial leadership:

- Increase in the likelihood of often (or always) managing sales and client relations without consulting anyone else (10 percentage points);
- Decrease in the likelihood that (strongly) agrees with the statement that “women should do what men say” (17–19 percentage points);
- Increase in the likelihood of agreeing in being flexible when it comes to making decisions (8–15 percentage points);
- Increase in the likelihood of being willing to invest in risky projects (18–21 percentage points);
- Increase in the likelihood of ensuring a conducive working atmosphere for their workers (24–25 percentage points);
- Increase in the likelihood being able to persuade their workers to work well (19–22 percentage points);
- Increase in the likelihood that (strongly) agrees that workers are part of the business’s decision-making process (21–22 percentage points);
- Increase in the likelihood of helping their workers with their financial constraints (15–17 percentage points);
- Increase in the likelihood of being open-minded and willing to listen to new ideas (17–21 percentage points);
- Increase in the likelihood of having the ability to place workers in positions appropriate to their capabilities (24 percentage points);
- Increase in the likelihood of being very active and getting completely involved, completing whatever they do (16–18 percentage points); and
- Increase in the likelihood of always ensuring that their staff makes progress (23–25 percentage points).

Networks:

- No effects found.

Business knowledge and practices:

- Decrease in the likelihood of not performing physical validation of inventory levels (15-19 percentage points);
- Increase in the likelihood that the owner will receive a fixed salary (9-11 percentage points); and
- Decrease in the likelihood of never checking the performance of the business (13-15 percentage points).

Analysis of the secondary question aimed to determine which activity component was most effective, and whether bundling the interventions resulted in additional effects beyond what could be achieved through the individual activity components. The analysis showed that most of the changes in the entrepreneurial leadership outcomes were driven by Components A and B, although Component C showed some, but fewer, positive effects as well. The few positive significant effects found for the business knowledge and practices outcomes were driven by Component A, while Component C showed a positive effect on only one outcome, and Component B showed a negative effect on one outcome in this category. **Across all the statistically significant outcomes, none showed a statistically significant marginal difference between the bundled treatment arm and the individual treatment arms. Thus, the short-term findings do not show an additional effect from bundling the interventions.** It is difficult to make conclusive inferences, however, regarding the effectiveness of the different treatment arms, given the potential contagion across the groups.

Conclusions

Five general conclusions can be drawn from the WLSME activity in India.

First, whereas there are some modest short-term positive impacts, mostly in terms of entrepreneurial leadership, they do not translate into business growth (higher sales, profits, or employment), the key outcome measurements of this intervention. This may be due to a combination of low statistical power due to small sample sizes, comparability of the control group, and likely indirect contagion and spillovers among treated groups.

Second, cashew-processing enterprises in the evaluation study area have distinct ownership and management characteristics that may affect how the positive entrepreneurial leadership outcomes are interpreted outside of this context. These businesses tend to be family run and roles are systematically divided, with husbands or male family members in charge of sales and client relations, and women in charge of processing operations and personnel. Thus, women were already assigned the role of supervising operations and personnel, and the WLSME activity contributed significantly to improving their entrepreneurial leadership in personnel management. These effects may be hard to replicate in a context where women were not already assigned that role.

Third, access to finance remained a constraint for these businesses throughout the duration of the activity. Available loan products were either microloans, which are too small given the upfront capital needed to procure bulk raw cashew nuts; or loans that required large collaterals not accessible to these businesses. While the implementing partner attempted to work with government and industry institutions to develop appropriate loan products for these micro-, small-, and medium-sized enterprises, they were not successful. In turn, many participants stated that without liquidity and working capital they could not scale up their businesses and make a profit. Instead, they were forced to process smaller quantities of cashews to simply stay afloat.

Fourth, assuming perfect conditions for the evaluation, it is likely that the time allowed for the activity to achieve statistically significant impacts was not long enough. This is probably true for outcomes that the literature has already identified as complex, such as sales and productivity, as well as empowerment. This is consistent with the fact that several outcome variables yield non-statistically significant coefficients, but do yield the expected sign. The latter is also consistent with some comments made by WLSME participants during the focus groups, where they expressed opinions that were aligned with the program objectives.

Fifth, it appears that the WLSME activity design was overly broad in terms of the number of interventions relative to the size of the target group and local context. Although early consultations with the evaluation team led to a scaled-down design, an activity with even fewer treatment arms could have

allowed for a more intense intervention dosage and less issues with contagion, increasing statistical power and the chances of detecting impact of the intervention.

Lessons Learned

Entrepreneurial leadership is a multidimensional and complex construct. Further research is needed to develop rigorous methods for measuring different dimensions of leadership.

There are varying conceptual and operational definitions of entrepreneurial leadership, and there is limited evidence on validated measures and scales to be used across different contexts. Given this evaluation's significant positive findings on entrepreneurial leadership, additional research to develop rigorous measures would strengthen the validity of findings from future evaluations.

Access to finance remains an important constraint for SMEs, so SME support programs should conduct a constraints analysis as early as possible to better understand the local supply and demand factors underpinning this barrier. While the WLSME activity aimed to make the cashew-processing businesses more attractive to formal moneylenders (through formalization, record-keeping, sales, and productivity), the nature of these businesses requires large upfront capital to buy bulk raw cashew nuts, making the existing microloans inadequate for their needs. While the implementing partner attempted to address the supply-side gap during implementation, it was done out of field feedback and was not part of the initial activity work plan. More research is needed on how to best structure the provision of finance.

Evaluators should be involved from the start of the activity implementation design. The WLSME activity did two things very well. First, the implementing partner had a local presence and was very involved in the communities, which resulted in timely and meaningful feedback to the evaluators early on. Second, close collaboration between the implementing partner and the evaluation team during start-up enabled quick reaction to the specific needs or changes required from the intervention without critically compromising the needs of a rigorous evaluation. These two factors highlight the importance of engaging the evaluators from the inception of projects.

“Early warning” mechanisms should be put in place during the design and early implementation phase to allow for timely course corrections. The design showed some structural weaknesses from the outset, in particular, identification of a feasible location to function as a credible comparison group, as well as the potential for extensive spillover or indirect contagion within the treatment groups. This was compounded by a delay in the implementation of one activity component (Component A). Putting in place “early warning” mechanisms, such as an external peer review of the design report, and linking a formal implementation fidelity system between the implementing partner and evaluation team can allow for timely modifications to the design to mitigate the potential threats to the validity of the evaluation.

Future designs should explicitly take into account the trade-offs between broadness and depth. This WLSME activity was overly broad in terms of the number of interventions relative to the size of the target group, even after the treatment arms were scaled down. Given the local context, a more modest intervention would have helped in all aspects of the evaluation, from the theory of change, to the intervention design, to statistical power, and likely impact evaluation measurement.

INTRODUCTION

This is the final report for the impact evaluation of the Women's Leadership in Small and Medium Enterprises (WLSME) activity in India. It was commissioned by the Office of Trade and Regulatory Reform (TRR) in the United States Agency for International Development's Bureau of Economic Growth, Education, and Environment (USAID/E3). The Bureau's E3 Analytics and Evaluation Project² has provided post-baseline support for the evaluation, including ongoing implementation monitoring and follow-up data collection. This impact evaluation consists of a quasi-experimental design to test how the WLSME interventions affected women and their businesses across four sets of indicators: business growth, entrepreneurial leadership, business knowledge/practices, and social/business networks.

The report summarizes the data collection and analysis methods used, and provides initial findings and conclusions based on post-intervention data analysis of the WLSME India activity. While the initial evaluation design planned to collect multiple follow-up rounds of survey data, USAID decided not to conduct additional data collection after the first follow-up round. Thus, this report contains the short-term findings (hereafter referred to as the endline) that were observed six months after the WLSME activity ended in December 2016.

ACTIVITY OVERVIEW

WLSME Initiative

USAID's WLSME initiative aims to address women's under-representation in the small and medium enterprise (SME) sector by funding and testing innovative interventions to increase the entry and growth of women-owned and women-managed SMEs in the developing world. The initiative focuses on reducing three critical barriers by implementing specific measures to produce structural change so that women may benefit from labor market participation in both the short and medium terms. The three barriers are: (i) agency constraints, which impede adequate accumulation of human capital and managerial capital, and thus limit women's knowledge and business practices; (ii) relationship constraints, which limit women's access to information and, consequently, reduce opportunities for female entrepreneurs to build and draw on social capital; and (iii) structural constraints, which place gender-specific barriers that limit the presence and success of women entrepreneurs.

In September 2012, USAID awarded three WLSME activities in India, Kyrgyzstan, and Peru through a competitive process, each with a performance period of three years and a budget of around \$1.5 to \$2 million. Impact evaluations for each of these three activities were also initiated under a cooperative agreement with FHI 360 that concluded on September 30, 2014, following the evaluation design and baseline data collection and analysis. Subsequently, implementation of two of the impact evaluations (for Kyrgyzstan and India) was transferred to the E3 Analytics and Evaluation Project, while the Peru evaluation is being funded separately and completed through the Multilateral Investment Fund of the Inter-American Development Bank. The implementation timeline is different for each of the WLSME activities; the India activity received an extension and concluded in December 2016.

² The E3 Analytics and Evaluation Project team consists of a team lead, Management Systems International, A Tetra Tech Company, and team partners, Development and Training Services, a Palladium company; and NORC at the University of Chicago.

India WLSME Activity

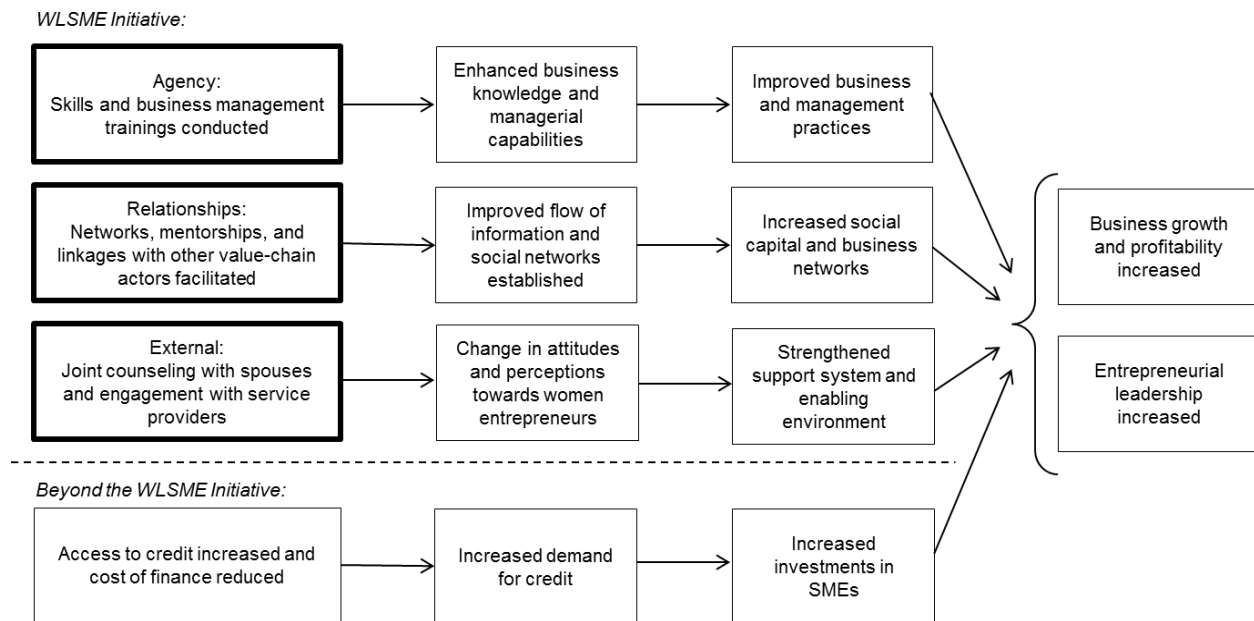
CARE India, in partnership with its sub-grantee, the Loyola Institute of Business Administration (LIBA), implemented the WLSME activity in India. The activity was designed to promote women's leadership and the sustainable growth of 210 cashew-processing microenterprises and SMEs owned and managed by women in the Panruti block of Tamil Nadu State. The components of this activity, summarized below, were implemented in separate groups, and aimed to address the three constraints outlined earlier.

- **Component A – Human Capital (Agency):** Strengthen skills, capacities, and capabilities of women to own and manage sustainable enterprises. This included business and skills training on entrepreneurship, technology adoption, accounting, marketing, human resource management, the business registration process, and financial management. The trainings focused on theme-based lectures, but also incorporated peer learning, one-to-one mentoring, and guest speakers. The training curriculum was semi-structured, there were no minimum numbers of sessions or hours that needed to be completed, and beneficiaries decided which sessions to attend.
- **Component B – Information and Social Capital (Relations):** Facilitate effective relationships between women entrepreneurs and the value chain actors. This included establishing a network of women entrepreneurs in cashew processing to facilitate linkages with value chain actors, strengthen collective bargaining power, and increase information flow and knowledge sharing among women. The network collected member dues and provided a portion of those funds to members as short-term loans. Additionally, the activity team selected members who had exhibited leadership, knowledge, confidence, and previous experience interacting with service providers and other value chain actors to be “change leaders;” change leaders were expected to mentor and provide peer support to other SMEs in this component.
- **Component C – Structural Constraints (External):** Promote an enabling environment and more positive attitude toward women entrepreneurs from family members and other stakeholders. The activity team developed information, education, and communication (IEC) materials to support discussion of gender roles within the business, household, and community. The team used the materials for joint counseling with spouses and to engage service providers to strengthen support services for women. Additionally, family members who were supportive of women entrepreneurs were selected as “role models,” to play a facilitator role with other families and SMEs in addressing the structural barriers faced by women entrepreneurs.
- **Component ABC** – Received all three of the above components, and had access to all activity activities.

DEVELOPMENT HYPOTHESES

USAID's development hypotheses for the WLSME activities are displayed graphically in Figure 1, which highlights each of the intended results of the project and the presumed causal linkages (arrows). The diagram focuses on three parallel constraints that are hypothesized to impede business growth and entrepreneurial leadership, and which the WLSME activities aim to address. It also includes the possibility of increased investments in SMEs that may occur as a result of increased access to and reduced cost of finance (shown under the dotted line). However, access to finance was not delivered directly by this WLSME activity.

FIGURE I: THEORY OF CHANGE



WLSME’s theory of change for this activity is that if agency constraints are reduced by increasing women entrepreneurs’ business knowledge and managerial capabilities, their business and management practices will improve, leading to increased business growth and profitability, and increased entrepreneurial leadership. In addition, if relationship constraints are reduced by increasing women entrepreneurs’ flow of information and links to other value-chain actors, their social capital and business networks will increase, which, in turn, will increase business growth, profitability, and entrepreneurial leadership. Further, if external constraints are reduced by changing the attitudes and perceptions of spouses and service providers toward women entrepreneurs, women’s support system and enabling environment will be strengthened, and, in turn, business growth, profitability, and entrepreneurial leadership will increase. Finally, if these three constraints are reduced simultaneously, this will lead to larger increases in business growth and profitability and entrepreneurial leadership than if each barrier is targeted separately.

EVALUATION QUESTIONS

Little empirical research exists that provides convincing evidence about which interventions have the greatest chance of success in creating female-led SMEs, and in helping women entrepreneurs grow their businesses (see Annex B). The purpose of this impact evaluation is to provide a learning, accountability, and decision-making platform by clarifying the most important constraints to women’s business growth and leadership, and thereby, the most effective means to unleash the potential of women’s entrepreneurship in the cashew value chain sector in India. This evidence is expected to be useful to USAID staff, other donors, host governments, and stakeholders to improve future programming to better address the barriers to women’s entrepreneurship at the SME level.

Evaluation questions for an impact evaluation are structured around the development hypotheses being tested. The evaluation questions included here are taken directly from the Evaluation Protocol designed by FHI 360 and from USAID’s Statement of Work for the evaluation. Therefore, the questions include

references to the evaluation design that are directly addressed in subsequent sections of this Evaluation Design Proposal. The actual evaluation questions are highlighted in bold.

1. Primary Question (combined impact of treatment vs. control): Compared to a propensity score-weighted sample of participants from the control region, **do study participants who were assigned to one of the four treatment arms have higher mean values on the following, post-intervention outcomes: entrepreneurial leadership, business growth, business knowledge/practices, and social/business networks?**
2. Secondary Question (separate estimates across treatment arms): Compared to participants in each of the other treatment arms (Component A, Component B, and Component C), **do participants who were randomly assigned to receive the combination of components of the intervention (Component ABC) have higher mean values on the same set of outcomes listed under the primary question?**

Outcome Measures

Several outcomes (dependent variables) linked to the theory of change were defined to measure whether and how much change WLSME activities caused for women entrepreneurs. These include:

- **Business Growth**: Measures include, but are not limited to: sales, profits, number of employees, number and type of paid employees, hours worked, investments, and formality.
- **Entrepreneurial Leadership**: This includes measures on decision-making in the business, entrepreneurial vocation, level of independence, and women's empowerment.
- **Business Knowledge and Practices**: This includes measures on marketing, inventory management, costing and record keeping, and financial planning.
- **Social/Business Networks**: This includes measures regarding involvement in professional networks, such as the number of other business owners with whom the women discuss business matters, as well as commercial networks.

Gender Aspects of the Questions

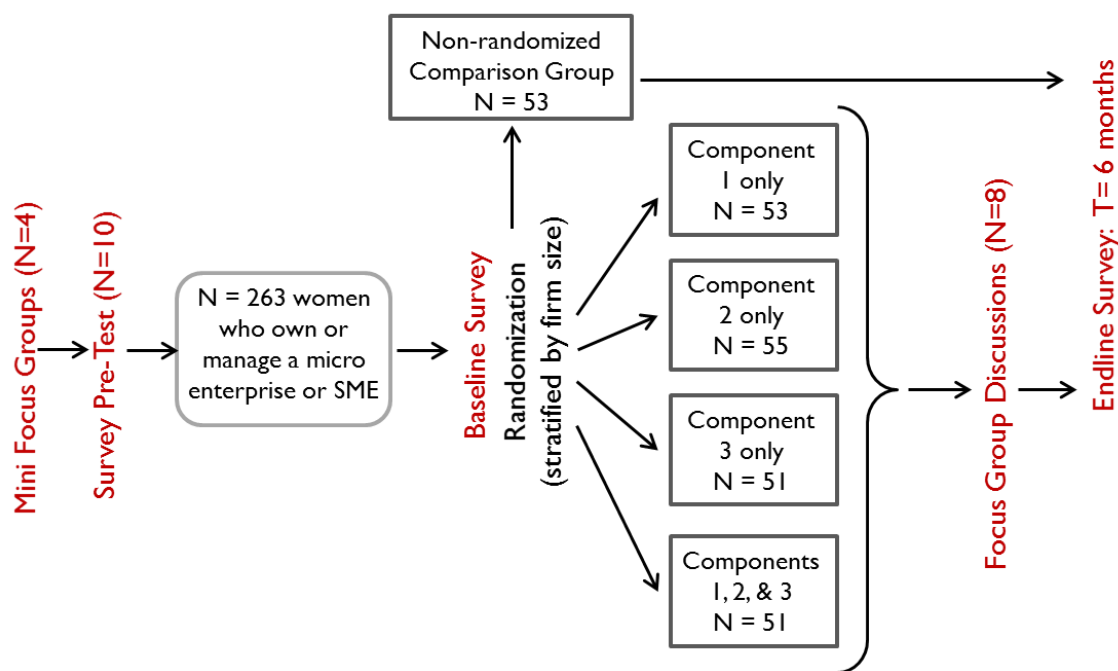
USAID evaluation guidance calls upon Agency staff and evaluation teams to examine evaluation questions from a gender perspective and to incorporate gender issues into study designs. While this WLSME activity is designed to address gender issues, it is targeted at women-owned/managed SMEs only, so the evaluation cannot disaggregate findings by gender or look at the differential gender effects of the activity components because only women are the direct beneficiaries. The WLSME India activity did include male stakeholders (either spouses or male value chain actors) in some activities, though not as direct beneficiaries so changes in the four outcomes above cannot be measured for these male stakeholders.

EVALUATION DESIGN

This impact evaluation employs a quasi-experimental design, comparing four randomly assigned treatment groups to a matched comparison group. Between August 2013 and June 2014, women-owned/managed cashew processing firms (n=210) from the eastern part of the Panruti block in Cuddalore District were interviewed for the baseline, then randomly assigned to different treatment

arms.³ Randomization was stratified by firm size (SME and micro), where firms with less than five employees are microenterprises and firms with 5 to 250 employees are considered a SME. Participants assigned to each treatment arm did not receive access to the activities in the other treatment arms, except for the fourth treatment group (ABC), which received access to all activities (see Figure 2). To avoid contamination and spillovers into the comparison group, assignment to the comparison group was not randomized from the same universe as the treatment arms, as is usually the case in pure experimental designs. Instead, the comparison group was randomly constructed from a list of similar microenterprises and SMEs (n=53) identified by CARE India and located in the western part of the Panruti block. The baseline survey for the comparison group was administered at the same time as the treatment group. Implementation of the WLSME activity started after the baseline ended and continued until December 31, 2016. Since randomization was done at the individual level, there is potential for contamination across the treatment arms. The women assigned to the different treatment arms came from the same area and may have known each other. Therefore, women in one treatment group could have gained indirect access to activities in other treatment groups, as beneficiaries shared information learned in their own group. While it was not possible to eliminate completely this potential for contamination, CARE India put in place safeguards as part of implementation to minimize this risk. The first follow-up survey took place between May and June 2017, six months after the end of the WLSME activity. While the initial evaluation design planned to collect multiple follow-up rounds of survey data, USAID decided not to conduct additional data collection after the first follow-up round (hereafter referred to as the endline).⁴ The revised evaluation design, timeline, and sample size are shown in Figure 2.

FIGURE 2: INDIA EVALUATION DESIGN



³ Between baseline and the start of the WLSME activity, 20 women selected to be in the project were unable to take part. Reasons included moving locations, switching businesses, or lack of interest. The baseline data from these 20 participants were discarded and replaced with 20 new participants who were surveyed at baseline and randomly assigned to the treatment arms.

⁴ USAID determined that the benefit from collecting additional data did not outweigh the cost of additional survey rounds due to (1) the inability to use the baseline data, (2) diluted intervention dosage because of delayed implementation and potential contagion across treatment groups, and (3) slow change in outcomes as seen in the short-term findings.

Data Analysis Methods

There are several design and implementation issues that make the empirical methods for this quasi-experimental evaluation more complex than would be under a pure experimental design. The small sample sizes of each treatment arm, the non-randomized comparison group, and the inability to use the baseline data⁵ led the evaluation team to select different statistical methods to present a more robust analysis. Both the primary and secondary questions are answered using three empirical approaches outlined below.⁶ Additional details on these empirical models can be found in Annex C.

- **Difference in Means:** This simple method was used instead of the more robust difference-in-differences method because only one round of survey data was usable. To answer the hypothesis, this method estimates the treatment effect by comparing the outcomes between two groups after an intervention. Given the non-randomized selection of the comparison group in this activity, this method is very limiting in its ability to answer the primary question, as there may have been observable differences between the treatment and comparison groups before the intervention started that cannot be controlled for. This method is also weak in its ability to answer the secondary question. Even though the treatment arms were randomized, the small sample size may have compromised the randomness of the corresponding treatment samples. Therefore, for this evaluation, the difference in means method is not appropriate to objectively answer the evaluation questions from an attribution perspective. However, the results of this method are reported as a referential measure in order to make comparisons with the results of the other two methods considered below. The evaluation team computed the treatment effect through an ordinary least squares (OLS) regression of the form:

$$Y_i = \beta_0 + \beta_1 D_i + \delta X_i + \varepsilon_i$$

Here, Y_i is the outcome variable, D_i is the treatment status dummy, X_i is a matrix of relevant covariates, and ε_i is the error term. Specifically, X_i contains the following variables: age, marital status, literacy, business ownership, and participation in previous trainings. Cashew processing activity fixed effects were also included. In this model, β_1 is the treatment effect.

- **Propensity Score Matching (PSM):** This method is used to calculate the treatment effect by comparing two groups while accounting for the covariates that predict receiving the treatment. PSM attempts to reduce the bias due to confounding variables by comparing outcomes between treated units versus a weighted average of non-treated units with similar observable characteristics. This technique is particularly useful in cases of non-randomized selection, which is the case for the comparison group in this study. However, this matching technique is limited when sample sizes are small. For this study, the evaluation team increased the size of the comparison group by 100 new SMEs to address this issue. The team estimated the treatment effect using PSM by first computing a predictive model of the probability of being assigned to a particular treatment given a set of observable covariates (propensity score). The variables used included age, marital status, literacy, business ownership, and participation in previous trainings. The team then used this propensity score to match treated and untreated units, and estimated the treatment effect by comparing treated units to a weighted average of the matched untreated units. The team used the kernel estimator to obtain the treatment effect. This avoids the loss of observations and improves the

⁵ The baseline data were not labeled by treatment arm, and the unique identification numbers did not match the records kept by CARE India. Given that the baseline did not contain any personally identifiable information (nor was this information kept separately), it was not possible to link the survey observations to specific respondents to merge the baseline with follow-up data.

⁶ These empirical methods are different from those proposed in the Evaluation Design Report because of the inability to use both rounds of data.

efficiency of the matching process by including all individuals in the comparison group as counterfactuals, weighted by the distance between the propensity score values. The OLS regression formula noted above was then employed.

- **Entropy Balancing (EB):** In this methodology, proposed by Hainmueller (2012), the data are preprocessed with the aim of achieving balance in the value of the moments (mean, variance, and skewness) of a vector of independent variables of interest. To this end, the method generates a vector of weights to conduct a weighted least squares regression. Some advantages of EB in comparison to PSM are (i) the former does not assume a model for the probability of being treated, (ii) it does not rely on the results of the common support, and (iii) the weights obtained can be used in any regression method. EB involves a reweighting scheme that directly incorporates covariate balance into the weight function that is applied to the sample units. Using that weighted vector, the evaluation team computed the treatment effect using the same OLS regression formula.

It should be noted that the issues described above (small sample size, non-randomized comparison group, and inability to use the baseline) apply to both the primary and secondary questions and may have resulted in unbalanced, biased samples. Given these issues, the PSM and EB approaches are useful techniques to account for all observable differences between the groups. However, the estimates obtained should not necessarily be interpreted causally, but merely as a sign of association between the activity components and the outcome variables.

In addition, the evaluation also includes a qualitative component designed to help interpret and better understand the quantitative analyses. The qualitative component included focus group discussions (FGDs) with women across the different treatment arms three months after the WLSME activity ended. The FGDs explored specific opinions and experiences of the activity in greater depth, and produced narratives that address the continuity of personal experiences over time. These discussions provide insights into the social and cultural dynamics by which the effects happen, and help explain *why* and *how* the WLSME activity worked. Focus group transcripts, translated into English, were analyzed with a focus on participant opinions and feedback related to WSLME activity effectiveness, utility, and outcomes, as well as perceptions on networks and personal empowerment. Responses were sorted categorically and assigned descriptive “codes” to facilitate frequency and cross-comparison. Common trends and themes were tracked across all eight FGDs, as were divergences and outliers from those trends and themes. Qualitative analysis was done parallel to the quantitative analysis and could not be combined.⁷ The key findings from this analysis are embedded along with the survey results.

Evaluation Limitations

This impact evaluation is a rigorous approach to addressing the evaluation questions and contributing to global knowledge on women’s entrepreneurship in the SME space. Despite efforts to minimize potential threats to validity, some limitations remain:

- **Small sample size limits the evaluation’s ability to detect changes in the outcomes.** The FHI 360 Evaluation Protocol powered the study for a MDES of ~0.5 to answer the Primary Objective (Evaluation Question 1), which is considered a medium-sized effect. The evaluation team increased the comparison group in the endline, which improves the MDES to 0.36 standard deviations, but this could still not be enough. Answering the Secondary Objective (Evaluation

⁷ The evaluation team chose not to video record FGDs to avoid inhibiting participants in voicing their opinions. Since only audio recording was used, it was not possible to tag each voice to a specific participant, then link the qualitative data to the survey data. Therefore, analysis for the two data sources was done in parallel and could not be combined.

Question 2) is even less plausible, since the sample sizes for those groups are smaller and the marginal effect size across activity components is even smaller.

- **Treatment and comparison groups are not balanced on characteristics and outcomes measures at baseline.** The initial baseline data analysis prepared by FHI 360 shows significant variance in characteristics and outcomes measures between treatment and comparison groups before the activity started. While this imbalance can be partially accounted for with PSM, there might be other unobservable characteristics that remain unbalanced between the groups, and this will introduce bias into the impact estimates.
- **Inability to use the baseline during the endline data analysis.** The baseline data were not labeled by treatment arm and did not contain any personally identifiable information, so it was not possible to link the survey observations to specific respondents to merge the baseline with endline data.
- **Spillover of the interventions among treatment arms represents a potential reduction in observable treatment effect.** The women randomized into treatment groups come from the same area and may know each other; therefore, beneficiaries easily shared the information learned in the activity with women in other treatment arms. While CARE India staff made an outstanding effort to limit direct contamination through the delivery of activities, women still knew each other and passed on information by word of mouth. This contamination would potentially result in a reduction of the observable treatment effect.

ENDLINE SAMPLE

The endline survey took place between May and June 2017, six months after the end of the activity, and achieved a 95 percent response rate (see Annex D for details on the data collection and quality assurance process). Table I provides the sample distribution by treatment and comparison groups, including the number of new observations in the comparison group at endline.

TABLE I: SURVEY SAMPLE DISTRIBUTION

	Baseline	Endline		% of Sample
		Original Obs.	New Obs. at Endline	
Treatment group	210	202	---	96.2%
<i>Component A</i>	53	50	—	94.3%
<i>Component B</i>	55	53	—	96.4%
<i>Component C</i>	51	50	—	98.0%
<i>Component ABC</i>	51	49	—	96.1%
Comparison group	53	48	100	90.6%
Total	263	250	100	95.1%

Note: Component A= human capital (agency); Component B= information and social capital (relations); Component C= structural constraints (external).

Prior to endline survey data collection, eight FGDs were held with WLSME participants roughly three months after the activity ended. This time frame was chosen to minimize respondent fatigue. Doing so allowed enough time to pass after a busy closing period from the WLSME activity, but not too much time that the FGDs were too close to the survey. The FGDs were structured by activity component, two FGDs for each activity component, given that distinctly different activities were provided to each

group. The FGDs had between 6 and 9 participants, for a total of 57 WLSME participants (see Table 2). All FGDs discussed a consistent but open-ended series of questions related to the results of the WLSME activity. These lines of questioning generally corresponded to the four sets of outcome variables. However, the format and venue of the FGDs was intended to provide more opportunities for additional questioning and open-ended participant response.

TABLE 2: FOCUS GROUP SAMPLE DISTRIBUTION

Component	Participants - Morning Session	Participants - Afternoon Session	Total
Component A	9	6	15
Component B	7	6	13
Component C	8	8	16
Component ABC	6	7	13

BALANCE AMONG TREATMENT AND CONTROL GROUPS

Balance across key demographic and outcome variables is necessary to show that the treatment and comparison groups are similar prior to the start of the intervention. However, the baseline report for the India activity showed significant differences between the overall treatment group and comparison group for several variables—age, education level, household income, having a business bank account, number of workers, and number of buyers. These differences, in conjunction with the other issues outlined in the previous section, limit the causal inferences that can be made from the results of this evaluation.

Given changes to the sample due to the expansion of the comparison group, the balance tests were conducted again, this time using the endline data. The evaluation team applied the student t-test (also known as a t-test) for two independent samples with unequal variances. This test provides previewing evidence about the differences between the treatment and comparison groups across variables that are time-invariant and not related to the treatment assignment. Eight characteristics were chosen to establish whether there were significant differences between treatment and comparison groups at endline that were not related to the treatment assignment. If the p-value associated with the t-test is small (p-value <0.05), there is evidence to suggest that the average value is different for both groups. Namely, the mean difference is significantly different from zero. On the contrary, when the p-value associated with the test is not small, it can be concluded that the means of both groups are not different.

Table 3 shows the tests of balance between the treatment and comparison groups at endline. The sample proves to be unbalanced on a few characteristics, namely literacy, ownership of business, and type of dwelling. While it is possible that, at the endline, some variables may have become unbalanced as a result of the intervention,⁸ there is indication that balance was already a problem at baseline, as reported in the baseline report.

⁸ The number of time-invariant characteristics available are limited so a more thorough analysis on balance cannot be conducted.

TABLE 3: ENDLINE BALANCE TEST

Characteristics	Control	Treatment	Difference	p-value
Age	41.91	40.50	-1.41*	0.086
Married (=1)	0.98	0.96	-0.02	0.284
Literate (=1)	0.21	0.33	0.12***	0.010
Owner (=1)	0.69	0.44	-0.25***	0.000
Full-time employees	12.05	12.96	0.91	0.151
Children under 18	0.78	1.00	0.22*	0.066
Household size	4.77	4.56	-0.21	0.134
Dwelling type (1=pucca, 2=semi-pucca, 3=kuccha)	2.20	1.95	-0.25***	0.000
<i>Number of observations</i>	<i>148</i>	<i>202</i>	—	—

Note: Statistical significance is denoted by the following system: *** p<0.01, ** p<0.05, * p<0.1

PRIMARY QUESTION FINDINGS – OVERALL IMPACT

This section reports the findings related to the WLSME activity’s overall impact based on application of the difference in means, PSM, and EB methods. More specifically, this section estimates whether the presence of any intervention component, including the combination of intervention components, has an impact on a broad set of outcome variables. The estimations below can be interpreted as “intent to treat,” which represents the average effect of having access to the WLSME interventions; that is, all the women assigned to the treatment group remain part of the treatment group regardless of actual participation.

To make consistent interpretations of the empirical findings, this evaluation introduces the following simple reliability criteria: when both the PSM and EB methods show analogous results in terms of significance and sign of coefficient, the evaluation team considers them reliable findings that are statistically significant and merit interpretation. When PSM and EB results diverge in either statistical significance or sign of coefficient, the findings are not considered reliable and, therefore, no credibility is assigned to those results. In addition, the findings using the difference in means method serve only as referential findings, and are not included in any part of the analysis.⁹

Business Growth

Business growth includes variables related to sales, profit, business cycle, time spent working in the business, number and type of paid employees, investments, and loans.¹⁰ Using the reliability criteria described above, most of the outcome variables are statistically insignificant, as shown in Table 4. While some outcome variables yield the expected coefficient sign for both PSM and EB, they are not statistically significant. This is the case for average sales in a good month, average profits in a good month, likelihood that a firm has a business bank account, number of paid non-household workers, and several other variables. The lack of statistical significance for sales and profit outcomes based on

⁹ It is important to note, however, that when coefficient estimates using PSM and EB coincide in terms of significance and sign, the Difference in Means method mostly does too. Also, the criterion of “equal signs of coefficients” between the PSM and EB methods is, in a way, redundant as this reliability criterion is always complied with in all cases.

¹⁰ Given the presence of outliers in some of these variables, the calculations were performed after trimming the sample by five percent on each tail. This is the case for “average sales in a good month” and “average profits in a good month,” among others.

different methods could be due to a combination of low statistical power and difficulty in measuring the sales variable, with the latter a factor frequently reported in related literature (McKenzie 2012).

TABLE 4: BUSINESS GROWTH OUTCOMES

Row	Outcome Variable	Difference in Means	PSM (Kernel)	Entropy Balancing
1	Average sales in a good month (rupees) [†]	-1,697.95	-1,034.50	-1,052.73
2	Average sales in a bad month (rupees) [†]	-2,482.23*	14.99	-2,522.02**
3	Average sales in an average month (rupees) [†]	-2,547.20	-459.44	-5,892.24*
4	Average profit in a good month (rupees) [†]	-1,908.93	-2,959.18	-2,004.85
5	Average profit in a bad month (rupees) [†]	-1,886.04	-1,565.62	-2,278.97
6	Average profit in an average month (rupees) [†]	-3,944.82	-1,503.20	-1,581.15
7	Number of good months in the last year	-0.38***	-0.27***	-0.37***
8	Number of bad months in the last year	-0.26	-0.23	-0.16
9	Number of average months in the last year	-0.18	-.29**	-0.26
10	Sales in the last 12 months (rupees)	20,918.94	30,272.39	41,915.98*
11	Likelihood of sales doing better in the last 12 months	-0.12*	-0.05	-0.07
12	Likelihood of having a bank account for business	0.10	0.04	0.07
13	Number of people from household who have worked in the business in the last 12 months	-0.17	-0.04	-0.12
14	Number of people from outside household who have worked in the business in the last 12 months	1.67	0.05	0.50
15	Number of people from household who worked in the business and were remunerated with cash	-0.21	-0.16	-0.24
16	Number of non-household people who worked in the business and were remunerated with cash	1.81	0.07	0.51
17	Number of months per year spent working in the business owned or managed	0.14	-0.56	-0.27
18	Number of days per week spent working in the business owned or managed	0.16**	0.06	0.07
19	Number of hours per day spent working in the business owned or managed	0.14	0.07	0.06
20	Likelihood of purchasing raw materials, goods, or equipment for business with a loan in the last 12 months	0.01	-0.01	0.00
21	Likelihood of applying for a loan from a financial institution in the last 12 months	-0.04	-0.04	-0.02
22	Likelihood of loan approval	-0.03	-0.03	-0.03
23	Likelihood of registering business with government	-0.03	0.02	0.02

Note: Coefficients were obtained by difference in means, PSM with a regular kernel estimator, and EB regressions. The regressions included cashew processing activity fixed effects. Age, marital status, literacy, business ownership, and participation in previous trainings were included as control variables. Standard errors were corrected by clusters at the village level. Statistical significance is denoted by the following system: *** p<0.01, ** p<0.05, * p<0.1.

Outcome variables stated as “likelihood” can be interpreted as a percentage point change by multiplying the coefficient (treatment effect) by 100.

[†] For these variables, the sample was trimmed 5 percent at the tails for outliers.

Using the reliability criteria, the evaluation team found one statistically significant result corresponding to the number of good months in the last year, which shows a small negative effect. This means that firms in the treatment group reported a small decrease (about one-third of a month less) in the number of good months in the previous year in relation to firms in the comparison group. While this self-reported assessment may be due to an exogenous shock in the treatment group—as the treatment and comparison groups came from different geographic areas—it may also reflect pessimism, which may or may not be linked to the intervention.

This pessimism was reflected in the FGDs, where participants voiced their concerns about the low harvesting season due to scarce rain. Since the supply of local cashews has been low, their only option is to procure raw cashews from importers, which tend to be more expensive and require larger purchases and upfront payments. Most participants said they did not have the capital nor the access to finance to make these purchases, so they could not buy the amount of raw cashews they would have been able to process. Most participants agree that their biggest challenge is lack of capital and lack of access to credit and loans. Without liquidity and working capital they cannot scale up their businesses and make a profit; instead, they are forced to process smaller quantities of cashews to simply stay afloat.

The team also examined whether there was an association between the WLSME activity and crossing the firm size threshold from a microenterprise into a SME, where SMEs have 5 to 250 employees. Using our standard reliability criteria, Table 5 shows no statistically significant association between the overall WLSME activity and size of firm.¹¹

TABLE 5: FIRM SIZE OUTCOME

Outcome Variable	Difference in Means	PSM (Kernel)	Entropy Balancing
Likelihood of crossing SME threshold (5 to 250 <u>total</u> workers)	-0.04	-0.06	-0.07*

Note: Coefficients were obtained by difference in means, PSM with a regular kernel, and EB regressions. The regressions included cashew processing activity fixed effects. Age, marital status, literacy, business ownership, and participation in previous trainings were included as control variables. Standard errors were corrected by clusters at the village level. Statistical significance is denoted by the following system: *** p<0.01, ** p<0.05, * p<0.1.

Entrepreneurial Leadership

The second key category, entrepreneurial leadership, includes variables related to decision-making in the business, entrepreneurial vocation, level of independence, and women’s empowerment, among others. The main results are shown in Table 6. The evaluation team found that variables related to having a role in decision-making (rows 2 to 9) are not statistically significant at conventional levels when using the reliability criteria. The outcomes related to decisions taken without consulting anyone else (rows 10 to 17) also show no effect, mostly, except for the likelihood of often (or always) managing sales and client relations without consulting anyone else. This shows a positive effect of 10 percentage points. While WLSME beneficiaries do not report being in charge of managing sales and client relations, they do report often/always doing this without consulting others. This is an interesting finding given that roles within the cashew processing businesses tend to be systematically divided, with husbands or male family members in charge of sales and client relations, and women in charge of processing operations and personnel. It could be that while women still consider their husbands to be in charge of that aspect of the business, they are now taking part in it. The FGDs reflect this duality in decision-making and business roles in

¹¹ We also tested three additional proxies. The results using these additional definitions are mostly statistical zeros although in some cases yield the wrong sign. We also tested a sample that excludes the new observations from the comparison group and found analogous results.

more in detail. Given the division of labor, men tend to be away either procuring the raw cashews or selling the final product. While they are away, the participants interact with the service providers and make home sales on their own. The participants from the FGDs expressed bigger changes in decision-making power, but these changes are not reflected in the quantitative data.

TABLE 6: ENTREPRENEURIAL LEADERSHIP OUTCOMES

Row	Outcome Variable	Difference in Means	PSM (Kernel)	Entropy Balancing
1	Likelihood that prefers to work as an employee in a business instead of managing/owning one	0.01	-0.03	-0.01
2	Likelihood that (with partner/spouse or another household member) is in charge of general business planning decisions	0.05	0.04	0.04
3	Likelihood that (with partner/spouse or another household member) decides what inputs to buy for production	0.01	0.09	0.10
4	Likelihood that (with partner/spouse or another household member) is in charge of sales and client relations	0.06	-0.02	-0.01
5	Likelihood that (with partner/spouse or another household member) decides if she should apply for a loan	0.06	0.04	0.06
6	Likelihood that (with partner/spouse or another household member) decides her own singular wage	0.09	0.10	0.10
7	Likelihood that (with partner/spouse or another household member) decides what type of work she will do	0.10	0.13**	0.12
8	Likelihood that (with partner/spouse or another household member) is in charge of marketing and advertising decisions	0.10	0.05	0.09
9	Likelihood that (with partner/spouse or another household member) is in charge of staffing of business decisions	0.12	0.12*	0.11
10	Likelihood that often (or always) makes general business planning decisions without consulting anyone else	-0.05	-0.02	0.00
11	Likelihood that often (or always) decides what inputs to buy for production without consulting anyone else	0.07	-0.05	-0.03
12	Likelihood that often (or always) manages sales and client relations without consulting anyone else	0.04	0.10*	0.10**
13	Likelihood that often (or always) decides whether to apply for a loan without consulting anyone else	-0.06	-0.03	-0.04
14	Likelihood that often (or always) decides her own singular wage without consulting anyone else	-0.02	-0.04	-0.06
15	Likelihood that often (or always) decides what type of work she will do without consulting anyone else	0.04	0.00	0.03
16	Likelihood that often (or always) makes marketing and advertising decisions without consulting anyone else	-0.06	-0.03	-0.06
17	Likelihood that often (or always) makes staffing of business decisions without consulting anyone else	-0.01	0.04	0.04
18	Likelihood that (strongly) agrees that “women should do what men say”	-0.08	-0.17**	-0.19**
19	Likelihood that (strongly) agrees that “women must share their income with their husbands”	0.05	-0.08	-0.07

Row	Outcome Variable	Difference in Means	PSM (Kernel)	Entropy Balancing
20	Likelihood that (strongly) agrees that “it is OK if men abandon women if they wish to”	0.04	-0.07	-0.10
21	Likelihood that (strongly) agrees that “it is OK if men chide women because they went out without any permission”	0.05	0.07	0.11
22	Likelihood that (strongly) agrees that “it is OK if men chide women if they do not take care of children”	0.04	-0.03	-0.04
23	Likelihood that (strongly) agrees that “the role of women is to earn money and take care of her family”	0.02	0.02	0.01
24	Likelihood that (strongly) agrees that “a mother who works can establish a relationship as warm and solid with her children as a mother who does not work”	0.08*	0.09	0.11**
25	Likelihood that (strongly) agrees that “father’s and mother’s dedication is equally important for the learning and effective development of children”	0.11*	0.12	0.17**
26	Likelihood that (strongly) agrees that there are no gender inequality problems in her community	0.04	0.05	0.02
27	Likelihood that (strongly) agrees in being flexible when it comes to taking decisions	0.09**	0.15***	0.08*
28	Likelihood that (strongly) agrees that she is willing to invest in risky projects	0.16**	0.21***	0.18**
29	Likelihood that (strongly) agrees that she ensures a conducive working atmosphere for workers	0.25***	0.25***	0.24***
30	Likelihood that (strongly) agrees that she is able to persuade workers to work well	0.15**	0.22***	0.19**
31	Likelihood that (strongly) agrees that she always is empathetic toward workers	0.05	0.11	0.11
32	Likelihood that (strongly) agrees that she avoids unnecessary conflict between herself and workers	0.12	0.13*	0.12
33	Likelihood that (strongly) agrees that workers are part of process of decision-making for business	0.11	0.22***	0.21**
34	Likelihood that (strongly) agrees that she takes feedback from workers	0.08	0.15**	0.17
35	Likelihood that (strongly) agrees that she helps workers with their financial constraints	0.16***	0.15**	0.17***
36	Likelihood that (strongly) agrees that she keeps her ears open for suggestions from workers	0.12**	0.18**	0.18**
37	Likelihood that (strongly) agrees that she appreciates any new things that are introduced by workers in the work process	0.06	0.06	0.02
38	Likelihood that (strongly) agrees that she is open-minded and willing to listen to new ideas	0.11	0.21***	0.17**
39	Likelihood that (strongly) agrees that she has the ability to place people in positions appropriate to their capacities	0.17**	0.24***	0.24***
40	Likelihood that (strongly) agrees that she is very active and gets involved completely in whatever she does	0.13*	0.18**	0.16**

Row	Outcome Variable	Difference in Means	PSM (Kernel)	Entropy Balancing
41	Likelihood that (strongly) agrees that she always ensures that her staff makes progress	0.18***	0.25***	0.23***
42	Likelihood that (strongly) agrees that it is hard to keep workers motivated	0.13*	0.14*	0.15

Note: Coefficients were obtained by difference in means, PSM with a regular kernel estimator, and EB regressions. The regressions included cashew processing activity fixed effects. Age, marital status, literacy, business ownership, and participation in previous trainings were included as control variables. Standard errors were corrected by clusters at the village level.

Statistical significance is denoted by the following system: *** p<0.01, ** p<0.05, * p<0.1.

Outcome variables stated as “likelihood” can be interpreted as a percentage point change by multiplying the coefficient (treatment effect) by 100.

The empowerment outcomes (rows 18 to 26) also mostly show no effect, except for a 17-19 percentage point decrease, depending on the method, in (strongly) agreeing with the statement that “women should do what men say.” The baseline report showed that most women in both the treatment and comparison groups strongly agreed with this statement. Given the cultural context, this is a rather large effect. The FGDs show positive results. Mainly, participants feel the WLSME activity has given them more confidence in themselves and more mobility. They explained how, previously, they would not usually leave their homes. Attending activities allowed them to get out and ask their husbands to accompany them to the bank or market. While they have gained confidence, they explained that women are still limited in where they can travel and how they are treated in the market or with sellers.

The entrepreneurial leadership outcomes (rows 27 to 42) show remarkable findings. Ten out of 16 variables show positive effects that meet the reliability criteria. In particular, WLSME beneficiaries are:

- 8-15 percentage points more likely, depending on the method reported, to agree to being flexible when it comes to making decisions;
- 18-21 percentage points more likely to be willing to invest in risky projects;
- 24-25 percentage points more likely to ensure a conducive working atmosphere for their workers;
- 19-22 percentage points more likely to be able to persuade their workers to work well;
- 21-22 percentage points more likely to agree that workers are part of the business’ decision-making process;
- 15-17 percentage points more likely to help their workers with their financial constraints;
- 17-21 percentage points more likely to be open-minded and willing to listen to new ideas;
- 24 percentage points more likely to have the ability to place people in positions appropriate to their capabilities;
- 16-18 percentage points more likely to be very active and get completely involved in whatever they do; and
- 23-25 percentage points more likely to always ensure their staff makes progress.

These findings are also reflected in the FGDs, where women claim to be more confident and less shy, and have learned how to better supervise their employees and increase their productivity through better personnel management.

While the evaluation cannot claim causality of the WLSME activity, it is reasonable to conclude that the WLSME interventions played a significant role in these results for the simple reason that it is rather unrealistic to believe that a non-observed endogenous shock may have produced such big changes in the short-term. Moreover, the baseline report shows that members of the comparison group were slightly more in agreement with these statements before the WLSME activity started. Overall, there is strong

indication that there is a statistical association between the WLSME activity and improved entrepreneurial leadership outcomes.

Networks

For the outcome variables related to business and social networks, shown in Table 7, the evaluation team did not find any statistically significant findings consistent with the reliability criteria. The FGDs provide more insights into this dimension. While participants agree that the WLSME activity has increased the number of people they know, and they share information with each other, they are not a support network and would not consider doing business with each other. It seems like trust is not there, but also they do not want to risk falling out with each other if they incur a loss. Some participants stated they knew more cashew processing businesses inside and outside of their communities, mainly because of the exposure visits conducted by the WLSME activity. However, most FGD participants did not place much weight on these connections.

TABLE 7: NETWORKS OUTCOMES

Row	Outcome Variable	Difference in Means	PSM (Kernel)	Entropy Balancing
1	Likelihood that there are social groups in her community	0.05	0.02	0.02
2	Likelihood of being an active member of any social group	-0.01	-0.06**	-0.07
3	Number of people she can ask for business advice	-0.40	0.37	0.07
4	Likelihood that has negotiated lower prices with suppliers in the last three months	0.03	0.04	0.03
5	Number of people she could contact for hiring	0.23	0.28	-0.31
6	Number of business leaders from other companies who regularly meet with her	-0.49	0.17	-0.16
7	Number of community members who are not employees and whom participant can count on to help with the business	-0.84	-1.65**	-1.32
8	Number of suppliers currently working with	-0.54	-1.26	-0.92

Note: Coefficients were obtained by difference in means, PSM with a regular kernel estimator, and EB regressions. The regressions included cashew processing activity fixed effects. Age, marital status, literacy, business ownership, and participation in previous trainings were included as control variables. Standard errors were corrected by clusters at the village level.

Statistical significance is denoted by the following system: *** p<0.01, ** p<0.05, * p<0.1.

Outcome variables stated as “likelihood” can be interpreted as a percentage point change by multiplying the coefficient (treatment effect) by 100.

Business Knowledge and Practices

Table 8 shows the findings with respect to business knowledge and practices, which include implementation of marketing, operations, and accounting practices; negotiation skills; recording of budgets; and future expectations, among other practices targeted by the WLSME activity.

Of 14 outcome variables considered, only 3 are statistically significant and consistent with the reliability criteria. The evaluation team found that there is a 15-19 percentage point decrease, depending on the method, in the likelihood of not performing physical validation of inventory levels. In addition, the team found a 9-11 percentage point increase in the likelihood that the owner will receive a fixed salary.

Finally, the team found a 13-15 percentage point decrease in the likelihood that participants never check the performance of the business.

In general, the estimates on the other outcome variables are not statistically significant, not consistent with the reliability criteria, and very small in magnitude. It is hard to explain the lack of statistical significance in these variables and the FGD findings are also mixed. For example, some participants expressed that the WLSME activity did not teach them anything new about running their business. They have been doing this for years and they know what they are doing. However, other participants expressed that the activity helped them make their operations more hygienic and safe, and taught them how to keep balanced records, keep track of inventory, and separate business from household expenses. Two business practices that received greater positive consensus in the FGDs were branding their business and defining business goals over the next five years. The mixed findings from the FGDs, along with the positive effects for the three variables in the quantitative data, are a promising sign that may be linked to the WLSME activity.

TABLE 8: BUSINESS KNOWLEDGE AND PRACTICES OUTCOMES

Row	Outcome Variable	Difference in Means	PSM (Kernel)	Entropy Balancing
1	Likelihood that no marketing activities were implemented during the last three years	-0.01	-0.10	-0.02
2	Likelihood that participant does not perform a physical validation of inventory levels	-0.11	-0.19**	-0.15*
3	Likelihood that business runs out of inventory at least one time a month	-0.00	-0.09	-0.08
4	Likelihood that participant tried to negotiate a lower price with suppliers during the last three months	0.03	0.04	0.03
5	Likelihood that participant compared price and quality of inputs with other suppliers' products during the last three months	0.06	0.10	0.08
6	Likelihood that owner has a fixed salary	0.08	0.09*	0.11**
7	Likelihood that participant records salary of the owner in a notebook, registry, or computer	0.18**	0.09	0.08
8	Likelihood that participant does not keep track of business purchases and sales	-0.10*	-0.11	-0.07
9	Likelihood that participant has a written expense budget	0.07	0.05	0.05
10	Likelihood that participant has no written goals for next 12 months	0.00	-0.11	-0.06
11	Likelihood that business has no accountancy documents prepared annually	0.01	-0.01	0.04
12	Likelihood that participant never checks performance of business	-0.10	-0.13*	-0.15**
13	Likelihood that participant never compares actual performance and goals	-0.07	-0.06	-0.10
14	Likelihood that business does not have any insurance	0.00	0.00	0.00

Note: Coefficients were obtained by difference in means, PSM with a regular kernel estimator, and EB regressions. The regressions included cashew processing activity fixed effects. Age, marital status, literacy, business ownership, and participation in previous trainings were included as control variables. Standard errors were corrected by clusters at the village level.

Statistical significance is denoted by the following system: *** p<0.01, ** p<0.05, * p<0.1.

Outcome variables stated as "likelihood" can be interpreted as a percentage point change by multiplying the coefficient (treatment effect) by 100.

SECONDARY QUESTIONS FINDINGS – COMPONENT ANALYSIS

The previous section covered the overall impact from participation in the WLSME activity. It measured the access to any treatment component or combination of components with respect to the outcomes' measures. This section goes a step further to measure not only access to any treatment component, but whether the bundling of the three treatment components shows a higher association with the outcome variables than that of any individual component. In other words, we measure the association of each particular treatment arm (A, B, and C) with respect to receiving the bundled treatment arm (ABC). The objective of this analysis is to compare these two estimates in order to identify the treatment arm(s) with the greatest effects on the study outcomes. As is the case in the primary question section, this analysis cannot clearly claim causal attribution despite the fact that treatment arms were randomized. This because of the previously discussed limitations of the data, in particular, the use of a non-randomized comparison group and the inability to use the baseline data.

There are two potential ways to answer this secondary question, both of which are methodologically equivalent. The first approach is to compare (i) the individual treatment arm relative to the comparison group and (ii) the bundled treatment arm relative to the comparison group, and then (iii) estimate the difference between (ii) and (i). By definition, the estimate of this difference will yield the marginal difference of the bundled treatment arm and the individual treatment arm. In other words:

$$\text{Marginal Difference} = (\text{Bundled Component} - \text{Comparison Group}) - (\text{Individual Component} - \text{Comparison Group})$$

This is equivalent to:

$$\text{Marginal Difference} = (\text{Bundled Component}) - (\text{Individual Component})$$

Doing this enables the team to understand whether participants who were randomly assigned to receive the bundled treatment arm (ABC) have higher mean values on any given outcome variable when compared to participants in each of the individual treatment arms (A or B or C).

The second approach, seemingly simpler, is to directly compare the impact of the individual treatment arm to that of the bundled treatment arm and bypass the comparison group. That is:

$$\text{Marginal Difference} = (\text{Bundled Component} - \text{Individual Component})$$

This second approach may be superior to the first since there would be no need to use the non-randomized comparison group. The second approach, however, suffers from greater shortcomings than the first, which uses the non-randomized comparison group. The main problem with the second approach is that the sample size drops significantly, to about 50 observations for each group. Another problem, present in both approaches, is that contagion issues are further exacerbated since most treated individuals know each other and are likely to exchange information on the different activities received.

The findings, using the first (preferred) approach, are shown below. Tables 9-11 each show the two parts used to calculate the marginal differences. First, the tables present the effects from the treatment arms relative to the comparison group using the three empirical methods used to answer the primary question above. To simplify the tables, only the outcome variables that were statistically significant with at least one method (regardless of the method) are shown. Second, the marginal differences between the bundled and individual treatment arms are shown in the last column. The evaluation team applied the same reliability criteria described in the primary question to determine “reliable” findings that are consistently statistically significant and, thus, merit interpretation. The marginal differences are calculated

only for outcome variables that meet the reliability criteria for both parts, the individual component and bundled component. The reliability criteria employed in this section are particularly strict. To warrant interpretation, the following must occur for both PSM and EB: (i) the coefficient of the individual component is statistically significant, (ii) the coefficient of the bundled component is statistically significant; (iii) the signs of the individual and the bundled components are the same. If the sign of the coefficient of the individual component is positive, the sign of the coefficient of the bundle component must also be positive. In the context above, the difference in means of the bundled and individual components (the “marginal difference”) must be statistically significant to warrant statistical association.

There were some outcome variables in which (i) and (ii) occurred, but (iii) did not. That is, the sign of the coefficient of the individual component was different than the sign of the coefficient of the bundled component, or vice-versa. However, in the cases of some outcome variables, this difference in sign occurred with one or both coefficients being very close to zero.¹² To be consistent with the criteria, the team did not deem these results valid.

If the marginal difference is positive, then the bundled treatment arm has a marginal effect above the effect from the individual treatment arm. If the marginal difference is zero (i.e., the effect from the individual treatment arm is equal to the effect of the bundled treatment arm), then the observed effect is due to the individual activity component and not to the bundling of activity interventions. Statistical significance for the marginal difference was computed as a t-test between the bundled treatment arm coefficient and the point value of the estimator for the individual treatment arm.

Overall, based on the reliability criteria, the evaluation team found results similar to those presented for the primary question above. The team found evidence of an association between several entrepreneurial leadership outcomes in all three individual treatment arms. However, most of the entrepreneurial leadership outcomes were driven by Components A and B, although Component C shows some, but fewer, positive effects as well. Moreover, the short-term findings do not show an additional effect from bundling the interventions. It is hard to make conclusive inferences regarding the effectiveness of the different treatment arms given the potential contagion across the groups. Detailed tables and additional comments follow.

¹² As a hypothetical example, one coefficient (say, the coefficient of the bundled treatment, ABC) might be -0.0001, and the other coefficient (say, the coefficient of component A) might be 0.1. According to the team’s criteria, the corresponding outcome variable would be discarded, as the signs of these two coefficients are different. However, it is also true that the coefficient of the bundled component is economically a zero. Since the team is taking a strict, statistical approach to the results, it has discarded this case. However, one should be aware that this is a borderline case; this type of result may mean that not enough time has passed to be able to observe consistent signs between both methods.

TABLE 9: COMPONENT A VS. COMPONENT ABC IMPACTS

Row	Outcome Variable	Component A vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – A)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
Panel I: Business Growth Outcomes										
2	Average sales in a bad month (rupees) [†]	-2,965.85**	1,100.73	-3,194.57**	-2,030.60	-2,407.92	-1,539.73			
7	Number of good months in last year	-0.38***	-0.35**	-0.41***	-0.41***	-0.47***	-0.43***	-0.04	-0.12	-0.02
8	Number of bad months in last year	-0.29	-0.26	-0.26	-0.47*	-0.32	-0.44*			
9	Number of average months in last year	-0.34	-0.41	-0.44*	-0.16	0.03	-0.16			
10	Sales in the last 12 months (rupees)	-13,373.86	7,167.12	22,055.92	80,047.83	77,970.07	82,068.70*			
17	Number of days per week she spent working in the business she owns or manages	0.06	-0.00	0.00	0.23***	0.22*	0.19***			
20	Likelihood of applying for a loan from a financial institution in the last 12 months	-0.10**	-0.10*	-0.09	-0.00	-0.00	0.01			
21	Likelihood of loan approval	-0.06*	-0.06	-0.06	-0.00	-0.01	0.00			
22	Likelihood that business is registered with government	0.03	0.12	0.12	-0.06**	-0.06	-0.05*			
Panel II: Entrepreneurial Leadership										
7	Likelihood that (with partner/spouse or another household member) decides what type of work she will do	0.15	0.13	0.14	0.17*	0.16*	0.15			
9	Likelihood that (with partner/spouse or another household member) is in charge of staffing of business decisions	0.17*	0.18	0.12	0.20*	0.15	0.20*			
12	Likelihood that often (or always) manages sales and client relations without consulting anyone else	0.07	0.14*	0.15**	0.08	0.04	0.01			

Row	Outcome Variable	Component A vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – A)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
18	Likelihood that (strongly) agrees that “women should do what men say”	-0.22**	-0.26**	-0.28**	-0.18**	-0.19*	-0.25***	0.04	0.08	0.04
24	Likelihood that (strongly) agrees that “a mother who works can establish a relationship as warm and solid with her children as a mother who does not work”	0.17**	0.19*	0.20***	-0.06	0.06	-0.03			
25	Likelihood that (strongly) agrees that “father’s and mother’s dedication is equally important for the learning and effective development of children”	0.13	0.19*	0.20**	0.13	0.17*	0.15			
27	Likelihood that (strongly) agrees in being flexible when it comes to taking decisions	0.10**	0.11*	0.08	0.07	0.10	0.06			
28	Likelihood that (strongly) agrees that she is willing to invest in risky projects	0.17*	0.15	0.15	0.15	0.10	0.19*			
29	Likelihood that (strongly) agrees that she ensures a conducive working atmosphere for her workers	0.24**	0.26**	0.25***	0.26**	0.23**	0.28***	0.02	-0.03	0.03
30	Likelihood that (strongly) agrees that she is able to persuade her workers to work well	0.18*	0.21**	0.24**	0.19	0.21**	0.19*		0.00	-0.05
31	Likelihood that (strongly) agrees that she always is empathetic toward her workers	0.19*	0.22**	0.23**	0.13	0.05	0.17*			-0.05
32	Likelihood that (strongly) agrees that she avoids unnecessary conflict between herself and workers	0.09	0.08	0.09	0.22**	0.26***	0.22***			
33	Likelihood that (strongly) agrees that workers are part of process of decision-making for business	0.13	0.24**	0.26*	0.24**	0.20**	0.25**		-0.04	-0.01

Row	Outcome Variable	Component A vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – A)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
34	Likelihood that (strongly) agrees that she takes feedback from her workers	0.09	0.17	0.19*	0.17	0.13	0.15			
35	Likelihood that (strongly) agrees that she helps her workers with their financial constraints	0.16	0.13	0.14	0.20**	0.18*	0.19**			
36	Likelihood that (strongly) agrees that she keeps her ears open for suggestions from workers	0.19**	0.23**	0.23**	0.17**	0.18*	0.22***	-0.02	-0.05	-0.01
38	Likelihood that (strongly) agrees that she is open-minded and willing to listen to new ideas	0.20*	0.23**	0.23**	0.09	0.10	0.11			
39	Likelihood that (strongly) agrees that she has the ability to place people in positions appropriate to their capacities	0.13	0.19*	0.19	0.27***	0.23**	0.29***		0.04	
40	Likelihood that (strongly) agrees that she is very active and gets involved completely in whatever she does	0.20**	0.23**	0.21**	0.07	0.10	0.11			
41	Likelihood that (strongly) agrees that she always ensures that her workers make progress	0.20**	0.28***	0.30***	0.21*	0.22**	0.21**	0.01	-0.06	-0.09
42	Likelihood that (strongly) agrees that it is hard to keep workers motivated	0.17	0.19*	0.21*	0.17	0.08	0.19*			

Panel III: Network Outcomes

8	Number of community members who are not employees whom she can count on to help with the business	-0.97	-0.93	-1.07	-2.03*	-1.88*	-2.22**			
---	---	-------	-------	-------	--------	--------	---------	--	--	--

Panel IV: Business Knowledge and Practice Outcomes

2	Likelihood that does not use internet for marketing purposes or to sell products/services	0.00	0.00	0.00**	0.00	0.00	-0.00			
---	---	------	------	--------	------	------	-------	--	--	--

Row	Outcome Variable	Component A vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – A)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
3	Likelihood that does not perform a physical validation of inventory levels	-0.18*	-0.26**	-0.26**	-0.14	-0.14	-0.14			
9	Likelihood that does not keep track of business purchases and sales	-0.22**	-0.17*	-0.18**	-0.08	-0.06	-0.05			
13	Likelihood that never checks performance of business	-0.18	-0.21*	-0.22**	-0.12	-0.12	-0.08			
14	Likelihood that never compares actual performance and goals	-0.16	-0.20*	-0.23**	-0.05	0.00	0.01			

Note: Coefficients were obtained by difference in means, PSM with a regular kernel, and EB regressions. The regressions included cashew processing activity fixed effects. Age, marital status, literacy, business ownership, and participation in previous trainings were included as control variables. Standard errors were corrected by clusters at the village level. Statistical significance is denoted by the following system: *** p<0.01, ** p<0.05, * p<0.1. Statistical significance for the difference component was computed as a t-test between Component ABC coefficient and the point value of the estimator for Component A. Outcome variables stated as “likelihood” can be interpreted as percentage point change by multiplying the coefficient (treatment effect) by 100. † For these variables, the sample was trimmed 5 percent at the tails for outliers.

TABLE 10: COMPONENT B VS. COMPONENT ABC IMPACTS

Row	Outcome Variable	Component B vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – B)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
Panel I: Business Growth Outcomes										
3	Average sales in an average month (rupees) [†]	-5,810.64*	1,092.73	-7,443.45*	-3,245.43	-3,227.30	-3,610.90			
7	Number of good months in last year	-0.34***	-0.32**	-0.28***	-0.41***	-0.47***	-0.43***	-0.08	-0.15	-0.14
8	Number of bad months in last year	0.05	0.12	0.09	-0.47*	-0.32	-0.44*			
10	Sales in the last 12 months (rupees)	15,174.38	4,896.61	30,287.34	80,047.83	77,970.07	82,068.70*			
17	Number of days per week she spent working in the business she owns or manages	0.18**	0.09	0.08	0.23***	0.22*	0.19***			
19	Likelihood of purchasing raw materials, goods, or equipment for business with a loan in the last 12 months	-0.02	-0.02*	-0.02	0.00	0.01	0.01			
22	Likelihood that business is registered with government	-0.03	-0.01	0.01	-0.06**	-0.06	-0.05*			
23	Likelihood of sales doing better in the last 12 months	-0.06	-0.01	-0.01	-0.11*	-0.11	-0.09			
Panel II: Entrepreneurial Leadership										
7	Likelihood that (with partner/spouse or another household member) decides what type of work she will do	0.07	0.10	0.11	0.17*	0.16*	0.15			
9	Likelihood that (with partner/spouse or another household member) is in charge of staffing of business decisions	0.01	0.07	0.08	0.19*	0.15	0.19*			
17	Likelihood that often (or always) makes staffing of business decisions without consulting anyone else	0.11*	0.09	0.11**	0.04	0.03	0.05			

Row	Outcome Variable	Component B vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – B)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
18	Likelihood that (strongly) agrees that “women should do what men say”	-0.12	-0.11	-0.17	-0.18**	-0.19*	-0.25***			
21	Likelihood that (strongly) agrees that “it is OK if men chide women because they went out without any permission”	0.19**	0.20**	0.17*	0.05	0.07	0.09			
24	Likelihood that (strongly) agrees that “a mother who works can establish a relationship as warm and solid with her children as a mother who does not work”	0.17**	0.19*	0.14*	-0.06	0.06	-0.03			
25	Likelihood that (strongly) agrees that “father’s and mother’s dedication is equally important for the learning and effective development of children”	0.18**	0.20**	0.20**	0.13	0.17*	0.15			
27	Likelihood that (strongly) agrees in being flexible when it comes to taking decisions	0.12**	0.21***	0.15**	0.07	0.10	0.06			
28	Likelihood that (strongly) agrees that she is willing to invest in risky projects	0.22**	0.23**	0.27**	0.15	0.10	0.19*			
29	Likelihood that (strongly) agrees that she ensures a conducive working atmosphere for her workers	0.25***	0.25**	0.23***	0.26**	0.23**	0.28***	0.01	-0.02	0.04
30	Likelihood that (strongly) agrees that she is able to persuade her workers to work well	0.18*	0.20*	0.20*	0.19	0.21**	0.19*		0.02	-0.01
31	Likelihood that (strongly) agrees that she always is empathetic toward her workers	0.04	0.11	0.17*	0.13	0.05	0.17*			0.00
32	Likelihood that (strongly) agrees that she avoids unnecessary conflict between herself and workers	0.09	0.14	0.11	0.22**	0.26***	0.22***			

Row	Outcome Variable	Component B vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – B)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
33	Likelihood that (strongly) agrees that workers are part of process of decision-making for business	0.13	0.18*	0.20	0.24**	0.20**	0.25**		0.02	
34	Likelihood that (strongly) agrees that she takes feedback from her workers	0.16	0.20*	0.27**	0.17	0.13	0.15			
35	Likelihood that participant (strongly) agrees that she helps her workers with their financial constraints	0.20**	0.23**	0.21***	0.20**	0.18*	0.19**	-0.01	-0.06	-0.01
36	Likelihood that (strongly) agrees that she keeps her ears open for suggestions from workers	0.04	0.13	0.16	0.17**	0.18*	0.22***			
38	Likelihood that (strongly) agrees that she is open-minded and willing to listen to new ideas	0.17*	0.24**	0.23**	0.09	0.10	0.11			
39	Likelihood that (strongly) agrees that she has the ability to place people in positions appropriate to their capacities	0.16	0.20**	0.25**	0.27***	0.23**	0.29***	0.11	0.03	0.04
40	Likelihood that (strongly) agrees that she is very active and gets involved completely in whatever she does	0.19**	0.22**	0.19**	0.07	0.10	0.11			
41	Likelihood that (strongly) agrees that she always ensures that her workers make progress	0.21**	0.23**	0.26***	0.21*	0.22**	0.21**	0.00	-0.02	-0.05
42	Likelihood that (strongly) agrees that it is hard to keep workers motivated	0.12	0.11	0.18*	0.17	0.08	0.19*			
Panel III: Network Outcomes										
2	Likelihood of being an active member of any social group	-0.05	-0.09*	-0.09	-0.05	-0.02	-0.06			
8	Number of community members who are not employees whom she can count on to help with the business	-1.11	-1.30	-1.22	-2.03*	-1.88*	-2.22**			

Row	Outcome Variable	Component B vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – B)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
Panel IV: Business Knowledge and Practice Outcomes										
7	Likelihood that owner has a fixed salary	0.02	0.05	0.09*	0.10	0.11	0.11			
12	Likelihood that has no accountancy documents prepared annually	0.17**	0.21**	0.19*	0.03	0.09	0.07			
13	Likelihood that never checks performance of business	-0.10	-0.10	-0.18**	-0.12	-0.12	-0.08			

Note: Coefficients were obtained by difference in means, PSM with a regular kernel, and EB regressions. The regressions included cashew processing activity fixed effects. Age, marital status, literacy, business ownership, and participation in previous trainings were included as control variables. Standard errors were corrected by clusters at the village level. Statistical significance is denoted by the following system: *** p<0.01, ** p<0.05, * p<0.1. Statistical significance for the difference component was computed as a t-test between Component ABC coefficient and the point value of the estimator for Component B. Outcome variables stated as “likelihood” can be interpreted as percentage point change by multiplying the coefficient (treatment effect) by 100. † For these variables, the sample was trimmed 5 percent at the tails for outliers.

TABLE II: COMPONENT C VS. COMPONENT ABC IMPACTS

Row	Outcome Variable	Component C vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – C)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
Panel I: Business Growth Outcomes										
2	Average sales in a bad month (rupees)†	-3,809.59**	-892.83	-4,049.89***	-2,030.60	-2,407.92	-1,539.73			
5	Average profit in a bad month (rupees)†	-3,630.97**	-2,177.36	-3,784.87**	-1,510.61	-1,682.79	-2,002.96			
7	Number of good months in last year	-0.38***	-0.17	-0.37***	-0.41***	-0.47***	-0.43***			
8	Number of bad months in last year	-0.20	-0.11	-0.09	-0.47*	-0.32	-0.44*			
10	Sales in the last 12 months (rupees)	11,569.31	35,956.80	51,520.94	80,047.83	77,970.10	82,068.70*			
15	Number of non-household people who worked in the business and were remunerated with cash	1.78*	1.39	1.73*	0.91	0.71	0.80			
17	Number of days per week she spent working in the business she owns or manages	0.05	-0.05	0.00	0.23***	0.22*	0.19***			
22	Likelihood that business is registered with government	0.00	-0.00	0.02	-0.06**	-0.06	-0.05*			
23	Likelihood of sales doing better in the last 12 months	-0.17**	-0.07	-0.14**	-0.11*	-0.11	-0.09			
Panel II: Entrepreneurial Leadership										
7	Likelihood that (with partner/spouse or another household member) decides what type of work she will do	0.08	0.04	0.09	0.17*	0.16*	0.15			
9	Likelihood that (with partner/spouse or another household member) is in charge of staffing of business decisions	0.10	0.06	0.04	0.19*	0.15	0.19*			
12	Likelihood that often (or always) manages sales and client relations without consulting anyone else	0.09	0.14	0.12*	0.08	0.04	0.01			

Row	Outcome Variable	Component C vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – C)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
18	Likelihood that (strongly) agrees that “women should do what men say”	0.01	-0.06	-0.08	-0.18**	-0.19*	-0.25***			
20	Likelihood that (strongly) agrees that “it is OK if men abandon women if they wish to”	-0.12	-0.17*	-0.23*	0.02	-0.07	-0.03			
25	Likelihood that (strongly) agrees that “father’s and mother’s dedication is equally important for the learning and effective development of children”	0.07	0.09	0.13	0.13	0.17*	0.15			
27	Likelihood that (strongly) agrees in being flexible when it comes to taking decisions	0.07	0.13*	0.04	0.07	0.10	0.06			
28	Likelihood that (strongly) agrees that she is willing to invest in risky projects	0.15	0.18*	0.13*	0.15	0.10	0.19*			
29	Likelihood that (strongly) agrees that she ensures a conducive working atmosphere for her workers	0.22**	0.23**	0.20**	0.26**	0.23**	0.28***	0.04	-0.00	0.08
30	Likelihood that (strongly) agrees that she is able to persuade her workers to work well	0.12	0.18*	0.16*	0.19	0.21**	0.19*	0.04		0.03
31	Likelihood that (strongly) agrees that she always is empathetic toward her workers	-0.07	-0.02	-0.09	0.13	0.05	0.17*			
32	Likelihood that (strongly) agrees that she avoids unnecessary conflict between herself and workers	0.06	0.08	0.06	0.22**	0.26***	0.22***			
33	Likelihood that (strongly) agrees that workers are part of process of decision-making for business	0.09	0.17*	0.16	0.24**	0.20**	0.25**			
35	Likelihood that (strongly) agrees that she helps her workers with their financial constraints	0.12	0.15*	0.14	0.20**	0.18*	0.19**	0.02		

Row	Outcome Variable	Component C vs. Comparison Group			Component ABC vs. Comparison Group			Marginal Difference (ABC – C)		
		Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB	Diff. in Means	PSM (Kernel)	EB
36	Likelihood that (strongly) agrees that she keeps her ears open for suggestions from workers	0.13	0.17*	0.14	0.17**	0.18*	0.22***	0.01		
38	Likelihood that (strongly) agrees that she is open-minded and willing to listen to new ideas	0.11	0.20**	0.14	0.09	0.10	0.11			
39	Likelihood that (strongly) agrees that she has the ability to place people in positions appropriate to their capacities	0.22**	0.28***	0.25**	0.27***	0.23**	0.29***	0.06	-0.05	0.04
41	Likelihood that (strongly) agrees that she always ensures that her workers make progress	0.17	0.24**	0.20*	0.21*	0.22**	0.21**		-0.02	0.02
42	Likelihood that (strongly) agrees that it is hard to keep workers motivated	0.07	0.07	0.03	0.17	0.08	0.19*			
Panel III: Network Outcomes										
2	Likelihood of being an active member of any social group	-0.03	-0.071*	-0.03	-0.05	-0.02	-0.06			
7	Number of business leaders from other companies who regularly meet with her	-0.90*	-0.20	-0.88**	-0.17	-0.55	-0.23			
8	Number of community members who are not employees whom she can count on to help with the business	-0.76	-0.94	-0.75	-2.03*	-1.88*	-2.22**			
Panel IV: Business Knowledge and Practice Outcomes										
3	Likelihood that does not perform a physical validation of inventory levels	-0.13	-0.17*	-0.14	-0.14	-0.14	-0.14			
7	Likelihood that owner has a fixed salary	0.11*	0.14*	0.14**	0.10	0.11	0.11			

Note: Coefficients were obtained by difference in means, PSM with a regular kernel, and EB regressions. The regressions included cashew processing activity fixed effects. Age, marital status, literacy, business ownership, and participation in previous trainings were included as control variables. Standard errors were corrected by clusters at the village level. Statistical significance is denoted by the following system: *** p<0.01, ** p<0.05, * p<0.1. Statistical significance for the difference component was computed as a t-test between Component ABC coefficient and the point value of the estimator for Component C. Outcome variables stated as “likelihood” can be interpreted as percentage point change by multiplying the coefficient (treatment effect) by 100. † For these variables, the sample was trimmed 5 percent at the tails for outliers.

Business Growth

As was the case in the analysis of the primary question, most of the business growth outcome variables are statistically insignificant, as shown in Panel I of Tables 9-11. Looking across activity components, the evaluation team found that the small negative effect on the number of good months for the business in the previous year that was detected in the overall impact analysis was driven by Components A and B. The bundling of interventions does not have an additional marginal effect on this outcome. Given the small magnitude of this effect (between one-third to two-fifths of a month), the team does not consider this to be an effect with practical significance. The team also found a small increase (about one-fifth of a day) in the number of days per week spent working in the business for the women in Component ABC relative to the comparison group, but no effect on this outcome for women in the individual treatment arms. The lack of statistical significance in the business growth outcomes could stem from low statistical power due to the difficulty in measuring sales and profit variables and to the small sample sizes, or from the fact that it is still too early to detect impacts on these outcomes.

Moreover, Table 12 shows a negative association between Component A and firm size, that is, a 13-14 percentage point decrease in the likelihood of reaching SME status. However, this effect goes away when the new observations from the comparison group are excluded.

TABLE 12: FIRM SIZE OUTCOMES BY COMPONENT

Outcome Variable	Difference in Means	PSM (Kernel)	Entropy Balancing
Component A			
Likelihood of crossing SME threshold (5 to 250 <u>total</u> workers)	-0.09	-0.13*	-0.14*
Component B			
Likelihood of crossing SME threshold (5 to 250 <u>total</u> workers)	-0.03	-0.05	-0.06
Component C			
Likelihood of crossing SME threshold (5 to 250 <u>total</u> workers)	0.02	0.03	-0.00
Component ABC			
Likelihood of crossing SME threshold (5 to 250 <u>total</u> workers)	-0.07	-0.03	-0.08*

Note: Coefficients were obtained by difference in means, PSM with a regular kernel, and EB regressions. The regressions included cashew processing activity fixed effects. Age, marital status, literacy, business ownership, and participation in previous trainings were included as control variables. Standard errors were corrected by clusters at the village level. Statistical significance is denoted by the following system: *** p<0.01, ** p<0.05, * p<0.1.

Entrepreneurial Leadership

Analysis of the primary question shows that most of the significant links from the WLSME activity concern entrepreneurial leadership outcomes. Panel II of Tables 9-11 provides more insights into how the different activity components are related to these treatment effects.

In analyzing the treatment effects by activity component, the evaluation team found that variables related to having a role in decision-making (rows 2 to 9) were not statistically significant at conventional levels when using the reliability criteria. For outcomes related to decisions taken without consulting anyone else (rows 10 to 17), the team found that the positive effect on the likelihood of often (or always) managing sales and client relations without consulting anyone else is driven by Component A. Women in Component A, who received trainings and activities aimed at human capital skills, reported being 14 to 15 percentage points more likely often (or always) to manage sales and client relations without consulting anyone else.

While most of the empowerment outcomes (rows 18 to 26) showed mostly no effect at the overall activity level (primary question), the evaluation team found statistically significant effects on some outcomes by treatment arm. For example, the decrease in (strongly) agreeing that “women should do what men say” is driven by Component A. This outcome is also statistically significant for Component ABC, but the marginal difference between the two components is not. Thus, bundling of interventions does not have an additional marginal effect on this outcome. Four additional empowerment outcomes showed significant effects at the individual treatment arm level, but not at the overall activity level. Women in Components A and B reported roughly a 20 percentage point increase in (strongly) agreeing with each of the following two statements: “a mother who works can establish a relationship as warm and solid with her children as a mother who does not work,” and “father’s and mother’s dedication is equally important for the learning and effective development of children.” Women in Component C reported a 17-23 percentage point decrease in (strongly) agreeing that “it is OK if men abandon women if they wish to.” The fourth outcome, however, showed a negative effect. Women in Component B reported a 17-20 percentage point increase in (strongly) agreeing that “it is OK if men chide women because they went out without any permission.” Gender norms are slow to change, especially in this context, so while it is encouraging to see positive effects, it is still not possible to conclude which activity component was most effective. It is surprising, however, not to see larger effects on these outcomes for the women in Component C since they received targeted IEC materials and joint counseling to address gender roles.

The entrepreneurial leadership outcomes (rows 27 to 42) show remarkable findings for the overall activity (primary question). In addition to the 10 outcomes showing significant positive effects for the overall activity, 5 other outcomes show statistically significant positive effects for different treatment arms. While the positive effects are scattered across the treatment arms, Components B and A show more significant effects than Component C. Moreover, whenever there is a statistically significant effect for Component ABC, the marginal difference is not statistically significant, showing that the bundled treatment arm does not have an additional effect on these outcomes. One outcome is statistically significant for the bundled treatment arm only. That is, there is a 22-26 percentage point increase in women (strongly) agreeing that they avoid unnecessary conflicts between themselves and workers. The positive effects found by treatment arm are listed below:

Component A:

There were increases in women (strongly) agreeing that they:

- ensure a conducive working atmosphere for their workers (25-26 percentage points).
- are able to persuade their workers to work well (21-24 percentage points).
- are always empathetic toward their workers (22-23 percentage points).
- include workers in the business’ decision-making process (24-26 percentage points).
- keep ears open for suggestions from workers (23 percentage points).
- are open-minded and willing to listen to new ideas (23 percentage points).
- are very active and get involved completely in whatever they do (21-23 percentage points)
- always ensure their workers make progress (28-30 percentage points).

- find it hard to keep workers motivated (19-21 percentage points).

Component B:

There were increases in women (strongly) agreeing that they:

- are flexible when it comes to taking decisions (15-21 percentage points).
- are willing to invest in risky projects (23-27 percentage points).
- ensure a conducive working atmosphere for their workers (23-25 percentage points).
- are able to persuade their workers to work well (20 percentage points).
- take feedback from their workers (20-27 percentage points).
- help workers with their financial constraints (21-23 percentage points).
- are open-minded and willing to listen to new ideas (23-24 percentage points).
- have the ability to place people in positions appropriate to their capacities (20-25 percentage points).
- are very active and get involved completely in whatever they do (19-22 percentage points).
- always ensure their workers make progress (23-26 percentage points).

Component C:

There were increases in women (strongly) agreeing that they:

- are willing to invest in risky projects (13-18 percentage points).
- ensure a conducive working atmosphere for their workers (20-23 percentage points).
- are able to persuade their workers to work well (16-18 percentage points).
- have the ability to place people in positions appropriate to their capacities (25-28 percentage points).
- always ensure that their workers make progress (20-24 percentage points).

Four of these outcomes were statistically significant for all four treatment arms—the increased likelihood of a woman agreeing that (1) she ensures a conducive working atmosphere for her workers, (2) she is able to persuade her workers to work well, (3) she has the ability to place people in positions appropriate to their capacities, and (4) she always ensure that her workers make progress. As noted above, though, the marginal difference between the bundled and individual treatment arms is not statistically significant, showing that bundling the interventions does not have an additional effect on these outcomes.

Networks

The analysis of the primary question did not show any statistically significant effects on networks for the overall activity. Panel III of Tables 9-11, however, shows a negative effect for women in the bundled treatment arm who report knowing about two fewer community members (not employees) whom they can count on to help with the business. This finding could reflect a more pronounced perception from what was identified in the FGDs, where participants report knowing more people in the cashew processing business, but they are not a support network and would not consider doing business with each other. It could be that participants in the bundled treatment arm place even less weight on these connections because they were more exposed to different cashew businesses and their challenges.

Business Knowledge and Practices

With respect to business knowledge and practices, the analysis for the primary question shows three outcomes with statistically significant effects. Panel IV of Tables 9-11 shows that the effects on these three outcomes are driven by different individual treatment arms. For example, only women in Component A reported a decrease in the likelihood of not performing physical validation of inventory levels (26 percent), and never checking the performance of the business (21-22 percent). Women in Component C reported an increase in the likelihood that the owner receives a fixed salary (14 percent).

For the different activity components, three additional outcome variables show statistically significant effects that were not found for the overall impact of the activity. Significant effects for two of these outcomes were reported by women in Component A, including a 17 to 18 percent decrease in the likelihood of not keeping track of business purchases and sales, and a 23 percent decrease in the likelihood of never comparing actual performance and goals. The third statistically significant outcome, however, is a negative effect. Women in Component B reported a 19 to 21 percent increase in the likelihood of not preparing accountancy documents annually.

CONCLUSIONS

Five general conclusions can be drawn from the WLSME activity in India.

First, whereas there are some modest short-term positive impacts, mostly in terms of entrepreneurial leadership, they do not translate into business growth (higher sales, profits, or employment), the key outcome measures of this intervention. This may be due to a combination of low statistical power due to small sample sizes, comparability of the control group, and likely indirect contagion and spillovers among treated groups.

Second, cashew-processing enterprises in the evaluation study area have distinct ownership and management characteristics that may affect how the positive entrepreneurial leadership outcomes are interpreted outside of this context. These businesses tend to be family run and roles are systematically divided, with husbands or male family members in charge of sales and client relations, and women in charge of processing operations and personnel. Thus, women were already assigned the role of supervising operations and personnel, and the WLSME activity contributed significantly to improving their entrepreneurial leadership in personnel management. These effects may be hard to replicate in a context where women were not already assigned that role.

Third, access to finance remained a constraint for these businesses throughout the duration of the activity. Available loan products were either microloans, which are too small given the upfront capital needed to procure bulk raw cashew nuts; or loans that required large collaterals not accessible to these businesses. While the implementing partner attempted to work with government and industry institutions to develop appropriate loan products for these micro-, small-, and medium-sized enterprises, they were not successful. In turn, many participants stated that without liquidity and working capital they could not scale up their businesses and make a profit. Instead, they were forced to process smaller quantities of cashews to simply stay afloat.

Fourth, assuming perfect conditions for the evaluation, it is likely that the time allowed for the activity to achieve statistically significant impacts was not sufficient. This is probably true for outcomes that the literature has already identified as rather complex, such as sales and productivity as well as empowerment. This is consistent with the fact that, while several outcome variables have non-statistically significant coefficients, they do yield the expected sign. The latter is also consistent with

some comments made by WLSME participants during the focus groups, where they expressed opinions that were aligned with the program objectives.

Fifth, it appears that the WLSME activity design was overly broad in terms of the number of interventions relative to the size of the target group and local context. Although early consultations with the evaluation team led to a scaled-down design, an activity with even fewer treatment arms could have allowed for a more intense intervention dosage and less issues with contagion, increasing statistical power and the chances of detecting impact of the intervention.

LESSONS LEARNED

Entrepreneurial leadership is a multidimensional and complex construct. Further research is needed to develop rigorous methods for measuring different dimensions of leadership.

There are varying conceptual and operational definitions of entrepreneurial leadership, and there is limited evidence on validated measures and scales to be used across different contexts. Given this evaluation's significant positive findings on entrepreneurial leadership, additional research to develop rigorous measures would strengthen the validity of findings from future evaluations.

Access to finance remains an important constraint for SMEs, so SME support programs should conduct a constraints analysis as early as possible to better understand the local supply and demand factors underpinning this barrier. While the WLSME activity aimed to make the cashew-processing businesses more attractive to formal moneylenders (through formalization, record-keeping, sales, and productivity), the nature of these businesses requires large upfront capital to buy bulk raw cashew nuts, making the existing microloans inadequate for their needs. While the implementing partner attempted to address the supply-side gap during implementation, it was done out of field feedback and was not part of the initial activity work plan. More research is needed on how to best structure the provision of finance.

Evaluators should be involved from the start of the activity implementation design. The WLSME activity did two things very well. First, the implementing partner had a local presence and was very involved in the communities, which resulted in timely and meaningful feedback to the evaluators early on. Second, close collaboration between the implementing partner and the evaluation team during start-up enabled quick reaction to the specific needs or changes required from the intervention without critically compromising the needs of a rigorous evaluation. These two factors highlight the importance of engaging the evaluators from the inception of projects.

“Early warning” mechanisms should be put in place during the design and early implementation phase to allow for timely course corrections. The design showed some structural weaknesses from the outset, in particular, identification of a feasible location to function as a credible comparison group, as well as the potential for extensive spillover or indirect contagion within the treatment groups. This was compounded by a delay in the implementation of one activity component (Component A). Putting in place “early warning” mechanisms, such as an external peer review of the design report, and linking a formal implementation fidelity system between the implementing partner and evaluation team can allow for timely modifications to the design to mitigate the potential threats to the validity of the evaluation.

Future designs should explicitly take into account the trade-offs between broadness and depth. This WLSME activity was overly broad in terms of the number of interventions relative to the size of the target group, even after the treatment arms were scaled down. Given the local context, a more modest intervention would have helped in all aspects of the evaluation, from the theory of change, to the intervention design, to statistical power, and likely impact evaluation measurement.

ANNEX A: EVALUATION STATEMENT OF WORK

Statement of Work: Impact Evaluation of Women's Leadership in Small and Medium Enterprises (WLSME) Projects

I. Activity Description

USAID's WLSME initiative, commissioned by its Office of Gender Equality and Women's Empowerment in the E3 Bureau, aims to address women's relative absence in the SME sector in order to promote broad-based economic growth and poverty reduction through higher productivity and efficiency gains. This initiative focuses on reducing three critical barriers by implementing specific measures to produce structural change, so that women may benefit from labor market participation both in the short- and medium-term: (1) agency constraints, which impede adequate accumulation of human capital and managerial capital, and thus limit women's knowledge and business practices; (2) external constraints, which place gender-specific barriers that limit the presence and success of women entrepreneurs; and (3) relationship constraints, which limit women's access to information and, as a consequence, reduce the opportunities for women entrepreneurs to build and draw on social capital.

In September 2012, USAID awarded three WLSME projects through a competitive process in the Kyrgyz Republic, India, and Peru. The recipients of the cooperative agreements will be referred to as "grantees" throughout this document. Each cooperative agreement has a performance period of three years, starting late September 2012, and has a total USAID-funded budget of around \$1.5 to \$2 million. The project interventions are scheduled to end around late summer of 2015. FHI 360, through the FIELD-Support Leader with Associates award, completed the research design for all three evaluations, and also completed baseline data collection (including data cleaning and summary analysis) as well as the leadership scale validation as a component of these evaluations for the Kyrgyz Republic and India. FHI 360 also completed part of the baseline collection for Peru, and the Multilateral Investment Fund of the Inter-American Development Bank will fund the completion of that impact evaluation separately.

This SOW describes the remaining activities for the impact evaluations of the WLSME projects in the Kyrgyz Republic and India that will be carried out by the E3 Analytics and Evaluation Project to address specific constraints to the development of women's leaderships in SMEs. This includes revisions to the evaluation design as appropriate and agreed with USAID, conducting endline data collection and analysis, and preparing the final evaluation report, as well as potentially assistance in disseminating evaluation findings. WLSME project descriptions provided by the grantees are summarized below.

WLSME India: Cashew Value Chain

CARE USA in India, in partnership with its sub-grantee the Loyola Institute of Business Administration (LIBA), is implementing the WLSME project to promote women's leadership and the sustainable growth of 210 cashew-processing micro- and SMEs owned and managed by women in the Panruti block of Tamil Nadu. The project has the following three objectives aimed at addressing the critical barriers related to: (1) human capital gap (Agency), (2) information and social capital gap (Relations), and (3) external constraints (Structures).

Objective 1: Strengthen skills, capacities and capabilities of women to own and manage sustainable enterprises. Project activities include skills training for women entrepreneurs, building awareness and

knowledge of women entrepreneurs on various aspects related to their enterprises, support new SMEs and strengthen existing SMEs as sustainable enterprises, and facilitate access to financial services and government schemes.

Objective 2 Facilitate effective relationships among women entrepreneurs and with value chain actors. Project activities include establishing a network of women entrepreneurs in SME involved in cashew processing in the Panruti cluster, facilitating exchange of information and support among women entrepreneurs, and facilitating linkages with key actors in value chain.

Objective 3 Promote an enabling environment and more positive attitude and behavior toward women entrepreneurs from family members and other stakeholders. Project activities include facilitating positive attitude and support for women entrepreneurs from other household members, and engaging with service providers in the value chain to strengthen support services to women entrepreneurs.

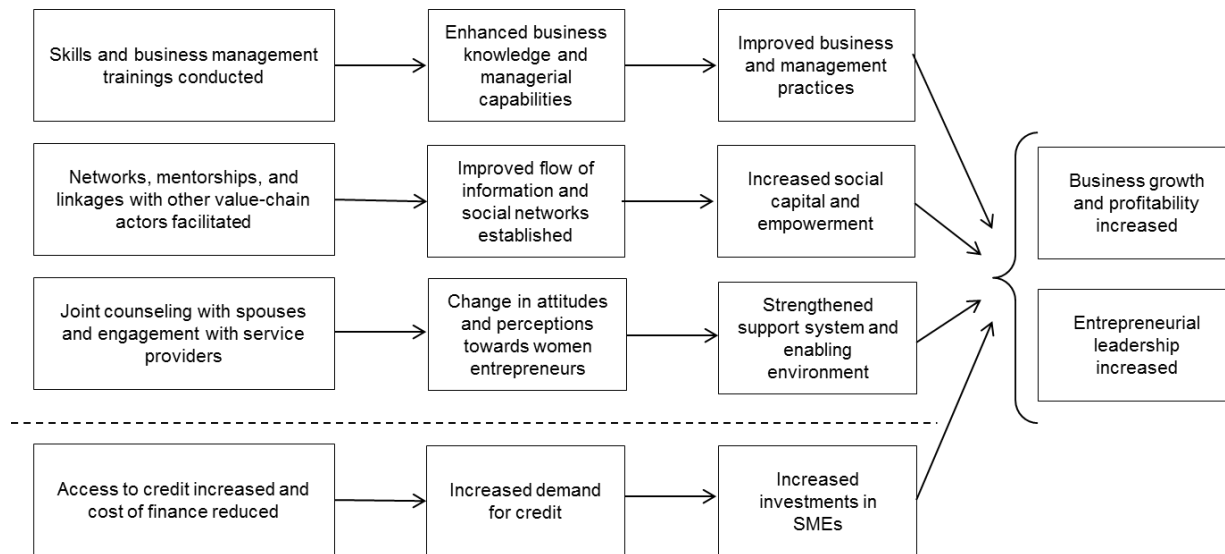
WLSME Kyrgyzstan

ACDI/VOCA, in collaboration with its partner organization Bai-Tushum Innovations Fund (BT Fund), is implementing the WLSME project in Kyrgyzstan. The project is operating nationwide and targets women who operate enterprises in priority sectors (garment, tourism, and agro-processing), meet minimum employee and loan size requirements, and are identified as potential high-growth entrepreneurs. Project activities are sequenced, with the main Business Management Training (BMT) reaching about 960 clients. As activities become more tailored and specialized, they progressively focus on fewer women. The second nested component, Market Linkages, consists of stakeholder meetings, trade fairs, workshops on value chains and sub-sectors, semi-annual value chain stakeholder meetings to address the information and social capital gap. The most intensive mentoring and skills development activities will be reserved for the 100 most promising women entrepreneurs.

II. Development Hypothesis

USAID's development hypothesis for the WLSME projects is displayed graphically in Figure I, highlighting each of the intended results of the project and the presumed causal linkages (arrows). The diagram focuses on three parallel constraints that are hypothesized to impede business growth and entrepreneurial leadership, which the WLSME projects aim to address. It also includes the possibility of increased investments in SMEs that may occur as a result of increased access and reduced cost of finance from the BT Fund partnership in Kyrgyzstan, shown under the dotted line. However, this final path of interest is not being delivered exclusively to impact evaluation treatment group members.

Figure I: WLSME Project Theory of Change



III. Existing Performance Information Sources

Given the ongoing status of the impact evaluations, USAID has provided the evaluation team with the following relevant project and evaluation documentation:

1. WLSME Kyrgyzstan project documents from ACDI/VOCA
 - Annual Work Plans
 - Annual and Quarterly Reports
 - Market Assessment Report
 - Gender Assessment Report
2. WLSME India project documents from CARE
 - Annual Work Plans
 - Annual and Quarterly Reports
 - Concept Notes on IEC materials, Joint Counseling Sessions, Change Leader identification process, Role Model selection, and Implementation Steps to avoid contamination for the RCT
3. Evaluation documents from FHI 360
 - Evaluation Design Protocols
 - Baseline Survey Questionnaires and Informed Consent forms
 - Baseline Reports
 - Baseline Survey datasets (STATA format)

The above list, which is non-exhaustive, highlights the more important sources of information that have been shared with the evaluation team. The following additional documents have not yet been provided to the evaluation team but will be shared as the evaluation progresses:

- All future quarterly project reports provided by ACDI/VOCA and CARE
- Copies or detailed descriptions of project activities
- Attendance spreadsheets pertaining to participation in each project activity

IV. Evaluation Purpose, Audience, and Intended Use

Purpose and Uses

There is little existing research to provide convincing empirical evidence about which interventions have the greatest chance of success in terms of creating female-led small and medium enterprises and helping them grow their businesses. Therefore, the purpose of these impact evaluations is to provide a learning, accountability, and decision-making platform by clarifying the most important constraints to women's business growth and leadership, and thus the most effective means to unleash the potential of women's entrepreneurship in SMEs. This evidence is expected to be useful to USAID staff to improve future programming in order to better address the barriers to women's entrepreneurship and to enhance its strategy on how to effectively support the business success of women entrepreneurs at the SME level.

This evidence will also be disseminated among practitioners and other governments and donors to contribute to the improvement of women's economic empowerment in developing countries. USAID is also supporting a core learning network, or closed Community of Practice (CoP), among the three implementing grantees and the evaluation teams. The CoP provides a space to share experiences on project implementation and the impact evaluations, and will serve as a nexus for disseminating results from these evaluations.

Audience

The primary audience for these evaluations is USAID staff in the E3 Bureau, particularly its Offices of Trade and Regulatory Reform and Gender Equality and Women's Empowerment. Findings and lessons learned from these evaluations may also be of interest to the business community, governments, donor agencies, and relevant practitioners in the field of women's economic empowerment in developing countries.

V. Evaluation Questions

The evaluation questions below were identified by USAID as reflecting Agency learning priorities for addressing women's relative absence in the SME sector, which is important to the promotion of broad-based economic growth and poverty reduction in terms of higher productivity and efficiency gains.

WLSME Kyrgyzstan

1. Primary Objective (combined impact T vs. C): Compared to participants in the control group, do participants who are randomly assigned to receive the program have higher mean values on the following, post-intervention outcomes: entrepreneurial leadership, business growth, business knowledge/practices, and social/business networks?
2. Secondary Objective (separate estimates across T arms):
 - Compared to participants who only receive Business Management Trainings, do participants also exposed to Market Linkages have higher mean values on the same set of outcomes listed under the Primary Objective?
 - Compared to participants who only receive Business Management Trainings, do participants also exposed to Technical Skills/Access to Finance have higher mean values on the same set of outcomes listed under the Primary Objective?

WLSME India

1. Primary Objective (combined impact T vs. C): Compared to a propensity score weighted sample of participants from the control region, do study participants who were assigned to one of the

four treatment components have higher mean values on the following, post-intervention outcomes: entrepreneurial leadership, business growth, business knowledge/practices, and social/business networks?

2. Secondary Objective (separate estimates across T arms): Compared to participants in each of other treatment arms (control, component 1 only, component 2 only, and component 3 only), do participants who were randomly assigned to receive the combination of components 1, 2 & 3 of the intervention have higher mean values on the same set of outcomes listed under the Primary Objective?

VI. Gender Considerations

USAID's Gender Policy (Automated Directives System 203.3.1.5) calls upon Agency staff and evaluation teams to examine evaluation questions from a gender perspective and to incorporate gender issues into study designs. As the WLSME projects are focused on women, they only intend to include male perspectives through the External Constraints components in the India project. As such, the evaluation team is not expected to collect data from male stakeholders (either spouses or male value chain actors). Thus, data collected in these evaluations will not be disaggregated by gender and will not look at the differential gender effects of the project components. Nonetheless, the main objective of the WLSME projects is to close the multiple existing gaps between women and men in SMEs.

VII. Evaluation Design and Methods

Impact Evaluation Design

USAID had previously commissioned impact evaluations of the WLSME projects under its FIELD-Support award with FHI360, under which an evaluation design was developed and baseline data were collected and reported. The FIELD-Support Cooperative Agreement concluded in September 2014, and USAID is now requesting that remaining activities for the impact evaluation be transferred to the E3 Analytics and Evaluation Project.

The Evaluation Protocols developed by FHI 360 describe the original research designs. The evaluation team should adhere to these protocols as much as possible, and deviations should only occur in consultation with the grantees and with USAID approval. The team should also note that each project has fairly comprehensive performance monitoring systems in place as well, but these were designed to minimize overlap with the evaluation indicators.

The two impact evaluation designs consist of a rigorous mixed-methods approach, with a randomized assignment to treatment complemented by qualitative data (including focus group discussions and in-depth interviews). The experimental designs involve collecting data from treatment groups and a control (counterfactual) group at multiple points in time, in order to make causal inferences with adequate statistical power and to document the size of the intervention's effects.

Data Collection Methods

Considering that baseline data collection and reporting has already occurred, endline survey research under the E3 Analytics and Evaluation Project for this evaluation will need to be conducted following the end of the respective project, in three rounds over two years. In addition, qualitative research such as focus group discussions will be conducted as soon as the projects end to access general intervention processes and content. To gain a deeper understanding of if and how the interventions improved participants' lives and businesses, in-depth interviews will be conducted after the 12-month follow-up survey is implemented and the study data analyzed. In responding to this SOW, the evaluation team

should identify where sampling and other types of selection procedures will be used to identify the individuals from whom qualitative data will be collected, and to explain those methods and why they are appropriate.

VIII. Data Analysis Methods

In its response to this SOW, the evaluation team should indicate and justify its choices for sequencing the collection of quantitative and qualitative data. For example, focus group discussions may precede further quantitative research to inform survey questions or may follow quantitative research to help explain survey findings; alternatively, these lines of data may be collected and analyzed in parallel and only synthesized once data from all sources are available. The evaluation design should also explain what statistical tests will be conducted on data collected to address all evaluation questions, how qualitative data will be analyzed, and whether that analysis will allow the evaluation team to transform some data obtained from qualitative into quantitative form.

IX. Strengths and Limitations

The evaluation designs of the two WLSME projects reflect a rigorous approach to address the proposed evaluation questions and to contribute to the global knowledge on women's entrepreneurship in the SME space. One key contribution is that both projects were specifically designed to test different pathways through which barriers affect women's economic outcomes and business success, which is a great contribution to the evidence base on women's entrepreneurship. Another strength of the envisioned evaluation designs is the use of multiple follow-up rounds and measuring outcomes over two years after the end of the project. This ensures that the evaluation will not miss potential impacts from the projects.

However, the evaluation will need to consider and address several limitations related to statistical analysis and attribution. Anticipated challenges, along with how they may be addressed by the evaluation team, are described briefly below.

- **Small sample sizes** reduce the ability to detect statistically significant differences between the groups. The original evaluation designs attempted to compensate for this by including multiple time points in the data collection and analysis plans.
- **Indirect contamination** across treatment arms and control groups may be present since eligible beneficiaries assigned to different groups reside in the same geographic areas. While WLSME staff attempt to limit direct contamination through the careful delivery of project activities, women may still pass on the information through word of mouth.
- **Attrition** presents another potential challenge, particularly since the intervention and data collection will take place over a three-year period. To help address this, the grantees have asked participants to notify the WLSME contact person if they move to other place, change contact information, or decide to leave the project. WLSME staff will call all women in the project (controls and treatments) on a regular basis, as time and resources permit, and update the database in case their contact information has changed, they have moved, or they decided to leave the project.
- **Selection bias**, given that not all project components were randomized, also poses a limitation in answering the evaluation questions, specifically when comparing across treatment groups in Kyrgyzstan and when comparing treatment versus control groups in India. While the addition of new statistical analysis methods can provide support for the findings, the bias still remains.

X. Evaluation Deliverables

It is anticipated that the evaluation team will be responsible for the deliverables listed in Table I for each evaluation. A final list of proposed deliverables and due dates will be included in the Evaluation Design Proposal to be prepared for each evaluation for USAID’s approval.

Table I: Preliminary Deliverables and Schedule for WLSME Impact Evaluations

Deliverable	Estimated Due Date
1. Evaluation Concept Paper (both evaluations), including methodological options to improve the evaluation design, and associated methods to the extent that options exist at this level.	o/a 30 days from client approval of SOW
2. Evaluation Design Proposals for each evaluation, including description of the evaluation methodology, drafts of data collection instruments and a sampling plan, as relevant	o/a 30 days from client approval to move forward with preparing Evaluation Design Proposal
3. Intermediate Reports covering follow-up data collection rounds, FGDs, and IDIs	o/a 60 days from completion of field research
4. Draft Impact Evaluation Report for each evaluation including key findings, conclusions and recommendations for USAID and its grantees	o/a 60 days from completion of field research
5. Oral Presentation on preliminary findings, conclusions and recommendations from quantitative and qualitative data collection	o/a 60 days from completion of field research
6. Final Impact Evaluation Report for each evaluation including evaluation data sets, codebooks, etc.	o/a 21 days following receipt of USAID feedback on Draft Evaluation Report
7. Debrief for grantees and partners (tentative)	As agreed following USAID approval of Final Evaluation Report

All documents and reports will be provided electronically to USAID no later than the dates indicated in the approved Evaluation Design Proposal. The format of the evaluation report should follow USAID guidelines set forth in the USAID Evaluation Report Template (<http://usaidlearninglab.org/library/evaluation-report-template>) and the How-To Note on Preparing Evaluation Reports (<http://usaidlearninglab.org/library/how-note-preparing-evaluation-reports>).

XI. Team Composition

Each evaluation will be delivered by a core evaluation team supported by technical and administrative U.S.-based evaluation and project management specialists. It is anticipated that the core evaluation team will be composed of a Principal Investigator who is an Evaluation Specialist, a Local Qualitative Researcher, and a Research Assistant. A survey research firm may also be contracted to support endline data collection. A final team composition, including proposed evaluation team members and their CVs, will be included in the Evaluation Design Proposal. Each team member will be required to provide a signed statement attesting that they have no conflict of interest, or describe any potential existing conflict of interest, and will be made available at USAID’s request.

Principal Investigator/Evaluation Specialist

The Principal Investigator must have a doctoral degree in a relevant social science and at least three years' experience conducting rigorous, experimental research in developing countries. The specialist should be able to demonstrate successful collaboration and leadership in cross-cultural contexts. Professional experience in research on gender equality and women's economic empowerment in developing countries and in the countries/region being studied is preferred. Proficiency in any of the relevant languages (Tamil, Kyrgyz, Russian) is a plus. The Principal Investigator will also possess skills in management, supervision, leadership and networking, and ability to work creatively towards solutions.

Local Qualitative Researcher

The evaluation team will also include a Local Qualitative Researcher who will complement the Principal Investigator in qualitative research and will lead the focus group discussions and in-depth interviews. The Researcher must have an advanced degree in a relevant social science, such as economics, statistics or sociology, and at least three years' experience conducting rigorous research in economic development or gender in developing countries. The specialist should be able to demonstrate successful collaboration and leadership in cross-cultural contexts. The specialist must also be proficient in any of the relevant languages (Tamil, Kyrgyz, and Russian).

Research Assistant

The Research Assistant should ideally be pursuing or have completed a graduate degree in a relevant social science such as economics or statistics, and ideally be trained in quantitative and qualitative random sampling and data collection methods. Proficiency in any of the relevant languages (Tamil, Kyrgyz, Russian) is a plus.

Survey Research Firm

Competent and experienced research firms will be selected to conduct endline data collection for the household survey, including data entry, for each evaluation. The firms will also support the piloting and translation (into Kyrgyz and Russian) of the survey instrument, as well as, transcription of focus group discussions and in-depth interviews.

XII. USAID Participation

While regular communication between the evaluation team and the designated USAID Activity Manager for this evaluation will be essential, USAID does not anticipate that any of its staff will serve as a full time team member on these evaluations, nor is it currently expected that USAID staff will join field data collection visits to project sites.

XIII. Scheduling and Logistics

The following tables provide the anticipated timeframe for evaluation activities and deliverables.

Table 2: Estimated Timeline for WLSME Kyrgyzstan

Tasks	FY15			FY16				FY17				FY18			
	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Finalize instruments and protocols for FGDs & survey	■														
<i>WLSME Kyrgyzstan Implementation</i>	■	■													
Conduct FGDs (n=6)			■												
Follow-up Round 1 (at project end)			■												
FGD Summary Report of methodology and findings				■											
Brief summary analysis and data for follow-up 1				■											
Follow-up Round 2 (at 12 months)							■								
Brief summary analysis and data for follow-up 2							■								
Instruments and protocols for IDIs							■								
Conduct IDIs (n=40)								■							
IDI summary report of methodology and findings									■						
Follow-up Round 3 (at 24 months)										■					
Brief summary analysis and data for follow-up 3											■				
Draft Endline Report												■			
Draft Evaluation Report												■			
Final Evaluation Report													■		

Table 3: Estimated Timeline for WLSME India

Tasks	FY15			FY16				FY17				FY18			
	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
<i>WLSME India Implementation (with extension)</i>															
Instruments and protocols for FGDs and survey															
Conduct FGDs (n=6)															
FGD summary report of methodology and findings															
Follow-up Round 1 (at 3 months)															
Brief summary analysis and data for follow-up 1															
Follow-up Round 2 (at 6 months)															
Brief summary analysis and data for follow-up 2															
Follow-up Round 3 (at 12 months)															
Brief summary analysis and data for follow-up 3															
Instruments and protocols for IDIs															
Conduct IDIs (n=40)															
Follow-up Round 4 (at 18 months)															
IDI summary report of methodology and findings															
Brief summary analysis and data for follow-up 4															
Follow-up Round 5 (at 24 months)															
Brief summary analysis and data for follow-up 5															
Draft Endline Report															
Draft Evaluation Report															
Final Evaluation Report															

The evaluation team will be responsible for procuring all logistical needs such as work space, transportation, printing, translation, and any other forms of communication. USAID will offer some assistance in providing introductions to partners and key stakeholders as needed, and will ensure the provision of data and supporting documents as possible.

XIV. Reporting Requirements

All members of the evaluation team will be provided with USAID’s mandatory statement of the evaluation standards they are expected to meet, shown in the following text box below, along with USAID’s conflict of interest statement that they sign and return to the E3 Analytics and Evaluation Project Home Office where necessary before field work starts.

USAID EVALUATION POLICY (APPENDIX I)

CRITERIA TO ENSURE THE QUALITY OF THE EVALUATION REPORT

- The evaluation report should represent a thoughtful, well-researched and well-organized effort to objectively evaluate what worked in the project, what did not and why.
- Evaluation reports shall address all evaluation questions included in the scope of work.
- The evaluation report should include the scope of work as an annex. All modifications to the scope of work, whether in technical requirements, evaluation questions, evaluation team composition, methodology or timeline need to be agreed upon in writing by the technical officer.
- Evaluation methodology shall be explained in detail and all tools used in conducting the evaluation such as questionnaires, checklists, and discussion guides will be included in an Annex in the final report.
- Evaluation findings will assess outcomes and impact on males and females.
- Limitations to the evaluation shall be disclosed in the report, with particular attention to the limitations associated with the evaluation methodology (selection bias, recall bias, unobservable differences between comparator groups, etc.).
- Evaluation findings should be presented as analyzed facts, evidence and data and not based on anecdotes, hearsay or the compilation of people’s opinions. Findings should be specific, concise and supported by strong quantitative or qualitative evidence.
- Sources of information need to be properly identified and listed in an annex.
- Recommendations need to be supported by a specific set of findings.
- Recommendations should be action-oriented, practical, and specific, with defined responsibility for the action.

Data Management Plan

The storage and transfer of data will adhere to the requirements laid out in ADS 579.¹³ The E3 Analytics and Evaluation Project should also follow Institutional Review Board (IRB) guidance on data security and confidentiality. All data collected at the field level should be managed by the evaluation team and overseen by the E3 Analytics and Evaluation Project Home Office team. Data should be filed in the appropriate format and processed in parsimonious, machine-readable format as they are collected. Final datasets are expected to be submitted to USAID in a format consistent with ADS 579. Metadata should be generated in the form of codebooks and data summaries as necessary. To ensure transparency and replicability, all data should be submitted as annotated datasets clearly defined with codebooks and annotated analysis of files.

XV. Budget

The evaluation team will propose a notional budget in its Concept Paper for these evaluations, including cost implications of the methodological options proposed. Full detailed budgets will then be prepared and included in each Evaluation Design Proposal for USAID’s approval.

¹³ See <http://www.usaid.gov/sites/default/files/documents/1868/579.pdf>

ANNEX B: LITERATURE REVIEW

Existing approaches to supporting growth-oriented women entrepreneurs have been heterogeneous in their design and delivery, although they have provided some suggestive evidence on the key issues that should be considered to move forward (Cirera and Qasim 2014). Recent efforts recognize this and have attempted to provide a more unified approach to most effectively inserting women into the productive process and, at the same time, helping to maximize their contributions to the well-being of societies (Buvinic et al. 2013). In the context of the current empirical research on the barriers faced by women entrepreneurs and the existing literature, the WLSME activities identified (i) agency limitations, (ii) external constraints, and (iii) lack of relationships as critical issues that should receive support to remove crippling limitations to women's productive advancement and contribution to the economy.

Agency Constraints

Whereas most academic and development policy discussions about female entrepreneurs focus on credit constraints, many studies and discussions assume that entrepreneurs manage their businesses optimally. In fact, human capital is treated as fixed, with a focus on the process of infusion of financial capital into microenterprises, not human or managerial capital (since academics/researchers assume that entrepreneurs have the latter in optimal amounts) (Karlán and Valdivia 2012). Clearly this is not necessarily the case, as the relatively poor among the self-employed rarely have any formal training in business skills. In particular, it has been argued that one must develop “managerial capital” in order to help entrepreneurs affect their firm's business practices, including improving strategic and operational decisions, and enhancing productivity by assisting entrepreneurs in using the factors of production more efficiently (Bruhn et al. 2012). Managerial capital appears to be a fundamental constraint for microenterprise development; business training may enable entrepreneurs to better identify profitable business opportunities, leading to changes in business practices and, ultimately, to higher sales, profits, and happiness (Berge et al. 2012).

The USAID activities included under these impact evaluations try to reduce agency constraints by improving human capital of female entrepreneurs, with an emphasis on their managerial capabilities. The key question asked is thus: Is lack of managerial capital a first order impediment to firm results, profitability, and growth? It has been shown in other studies that small-firm entrepreneurs are constrained in the acquisition of these skills, particularly if they require formal training (Caselli and Gennaioli 2005). The design of the treatment arms in this activity follows a systematic pattern that tries to condense the approaches taken by a growing number of microfinance organizations attempting to build the human and managerial capital of micro-entrepreneurs, which have been vastly idiosyncratic and heterogeneous, and, as a consequence, have provided limited external validity. This is perhaps the reason why the current literature on human and managerial capital reflects a mixed record. For instance, Karlán and Valdivia (2012) and Cole et al. (2011) show that basic microenterprise training seems to affect the command of accounting practices for staff of microenterprises, but it has limited to no effects on actual firm outcomes and performance, including profits and sales. Similarly, Bruhn and Zia (2013) and Giné and Mansuri (2014) find that training in managerial capital leads to improvements in business practices, but has only limited effects on business performance and sales. On the other hand, Drexler et al. (2012) show that training programs increase in impact if they are targeted to the owner level, as training has significant impact on real outcomes for micro-entrepreneurs who had low educational attainment and poor business practices prior to the intervention. Along the same lines, Field et al. (2010) find positive treatment effects on upper-caste Hindus, but no such effects on lower-caste Hindus or Muslims.

External Constraints

In spite of the importance of human capital and managerial capital, a consistent finding in recent academic research is that business training is vastly more effective for male entrepreneurs than female entrepreneurs (Berge et al. 2012; de Mel et al. 2014; Gine and Mansuri 2014). The differences are striking. Even though female entrepreneurs benefit from training in terms of business knowledge, researchers have been unable to find a positive effect on their business-related outcomes. Whereas it has been found that there are no differences in business knowledge between males and females, the former report better business practices, lower business failures, higher investment, and even higher household expenditures (Gine and Mansuri 2014). These findings point to the need for more comprehensive measures to promote the businesses of female entrepreneurs, as any positive effect of the business training is contingent on gender.

Furthermore, it appears that factors deeper than a lack of business knowledge seem to constrain the development of female-owned microenterprises. Female and male entrepreneurs fundamentally differ in terms of mindset and household constraints, which may indicate that more comprehensive measures are necessary to promote development of female entrepreneurs, with greater attention paid to their motivation(s) for being involved in business activities and to external constraints that may limit their opportunities (Berge et al. 2012). In this context, the WLSME activities aim to remove external constraints by promoting an enabling environment and a more positive attitude toward women entrepreneurs, with an emphasis on social norms. This constitutes the most plausible explanation for the gender differences between the roles of men and women in the workplace.

Relationship Constraints

Despite the fact that informal social mechanisms, such as word of mouth, may help reduce external constraints, there is a limit to them. For instance, it has been shown that a significant share of women say their (male) spouses are responsible for most of their business decisions, suggesting that female-owned businesses show no improvement because women have little decision-making control. Female entrepreneurs are less willing to share income information with their spouses than male entrepreneurs, which may suggest that female entrepreneurs are “taxed” by their husbands and, thus, may have less to gain from expanding their businesses (Berge et al. 2012). In this context, it is important to develop specific, formal, direct channels through which women entrepreneurs are able to interact with all the actors involved in the productive process. Research findings indicating that women may be less willing than males to compete suggest they may have less of an entrepreneurial mindset focused on business competition and growth (Berge et al. 2012).

The WLSME initiative aims to reduce information and social gaps in the productive process for women entrepreneurs by facilitating effective relationships between women and the value chain actors, and increasing cohesion in the productive process. It is also expected that the components included in the activity can help increase women entrepreneurs’ sense of empowerment, especially women with specific leadership skills. Thus, through women’s increased economic activity and greater control over income resulting from access to a larger network (Mayoux 2001; Kulkani 2011), these activities can help enhance the status of women entrepreneurs within the community, and this is reinforced by the formation of the networks that are also part of this activity. This approach is consistent with an empowerment paradigm that advocates for explicit strategies that support women’s ability to protect their individual and collective gender interests (Mayoux 2001).

ANNEX C: EMPIRICAL METHODS

It is well known that a randomized controlled trial approach is the best possible empirical method to answer questions of causal attribution, as eligible participants are randomly selected to treatment and control groups. Typically, both groups have, on average, balanced observable characteristics and are deemed to be statistically identical in the absence of the intervention. When well designed, the treatment and control groups only differ in exposure to the intervention, so any difference in outcomes at the end of the evaluation can be attributed to the intervention itself. However, by limitations in the design and other implementation issues, a “clean” RCT approach cannot be applied in the WLSME India activity for several reasons. First, the sample sizes of the corresponding treatment arms are rather small, which may compromise the randomness of the corresponding treatment samples. Second, the control group was not randomized from the same sample as the treatment groups. Third, while not critical for the purposes of a standard RCT, the survey baseline data for this evaluation cannot be used. The baseline data were not labeled by treatment arm, and the unique identification numbers did not match the records kept by CARE India. Given that the baseline did not contain any personally identifiable information (nor was this information kept separately), it was not possible to link the survey observations to specific respondents to merge the baseline with the follow-up data.

The issues above pose several challenges the team tried to overcome by using different statistical methods. Both the primary and secondary questions are answered using the three empirical methods described below.

- **Difference in Means:** This simple method was used instead of the more robust difference-in-differences method because only one round of survey data was usable. To answer the hypothesis, this method estimates the treatment effect by comparing the outcomes between two groups after an intervention. Given the non-randomized selection of the comparison group in this activity, this method is very limiting in its ability to answer the primary question, as there may have been observable differences between the treatment and comparison groups before the intervention started that cannot be controlled for. This method is also weak in its ability to answer the secondary question. Even though the treatment arms were randomized, the small sample size may have compromised the randomness of the corresponding treatment samples. Therefore, for this evaluation, the difference in means method is not appropriate to objectively answer the evaluation questions from an attribution perspective. However, the results of this method are reported as a referential measure to make comparisons with the results of the other two methods considered below. The evaluation team computed the treatment effect through an ordinary least squares (OLS) regression of the form:

$$Y_i = \beta_0 + \beta_1 D_i + \delta X_i + \varepsilon_i$$

Here, Y_i is the outcome variable, D_i is the treatment status dummy, X_i is a matrix of relevant covariates, and ε_i is the error term. Specifically, X_i contains the following variables: age, marital status, literacy, business ownership, and participation in previous trainings. Cashew processing activity fixed effects were also included. In this model, β_1 is the treatment effect.

- **Propensity Score Matching (PSM):** This method is used to calculate the treatment effect by comparing two groups while accounting for the covariates that predict receiving the treatment, following the methodology developed for non-experimental studies by Rosenbaum and Rubin (1983) and extended to other studies such as Heckman, Ichimura, and Todd (1998). PSM attempts to reduce the bias due to confounding variables by comparing outcomes between treated units versus a weighted average of non-treated units with similar observable characteristics. This technique is particularly useful in cases of non-randomized selection, which is the case for the comparison group

in this study. However, this matching technique is limited when sample sizes are small. For this study, the evaluation team increased the size of the comparison group by 100 new SMEs to address this issue. The team estimated the treatment effect using PSM by first computing a predictive model of the probability of being assigned to a particular treatment given a set of observable covariates (propensity score). A propensity score results from this probabilistic model and is used to identify the degree of comparability (in terms of the variables included in the model) among treatment and comparison observations. In this way, the problem of multidimensionality is reduced since the matching and the treatment effect estimation is calculated conditionally only to a scalar (the probability of being treated) rather than using a matching in a vector or covariates. The predictive model is defined as a model of discrete choice, Probit, in which the binary dependent variable, $D_i \in \{0,1\}$, takes the value of one if the observation “i” receives the treatment and zero in the other case. The regression of the Probit model includes a set of independent variables that correlates with the probability of the observation “i” receiving the treatment, including age, marital status, literacy, business ownership, and participation in previous trainings. Having these considerations, the model is defined as:

$$\Pr(D_i = 1|X) = F(X_i, \beta)$$

Where D_i is the treatment variable, X_i is the matrix of relevant variables and $F(\cdot)$ would assume a normal standard cumulative distribution function expressed as the integral given by

$$F(z) = \Phi(z) = \int_{-\infty}^z \phi(v) dv$$

Where $\phi(v)$ is the normal standard distribution. The marginal effect computed from the β parameters show the correlation between the independent variables and the probability of being treated. The team then used this propensity score to match treated and untreated units and estimated the treatment effect by comparing treated units to a weighted average of the matched untreated units. For the results in this report, we compute the kernel estimator for obtaining the average treatment effect since it avoids the loss of observations and improves the efficiency of the matching process by including all individuals in the comparison group as counterfactuals, weighted by the distance between the propensity score values. Formally, the weights are determined by

$$W_{N_0 N_1}(i, j) = \frac{K\left(\frac{p(X_i) - p(X_j)}{h_n}\right)}{\sum_{k \in \{D=0\}}^{N_0} K\left(\frac{p(X_i) - p(X_j)}{h_n}\right)}$$

Where N_0 is the number of control observations, N_1 the number of treated observations, $p(X_i)$ the propensity score for the observation “i”, $K(\cdot)$ the kernel function and h_n the bandwidth of smoothing parameter. With this kernel, the matching estimator for the outcome Y takes the following form (Lee, 2005):

$$T = N_1^{-1} \sum_{i \in \{D=1\}} \left\{ Y_i - \frac{\sum_{k \in \{D=0\}}^{N_0} K\left(\frac{p(X_i) - p(X_j)}{h_n}\right) Y_j}{\sum_{k \in \{D=0\}}^{N_0} K\left(\frac{p(X_i) - p(X_j)}{h_n}\right)} \right\}$$

- **Entropy Balancing (EB):** This methodology, proposed by Hainmueller (2012), preprocesses the data with the aim of achieving balance in the value of the moments (mean, variance and skewness) of a vector or independent variables of interest. To do so, the method generates a vector of weights to conduct a weighted least squares regression. Some advantages of the EB in comparison to the PSM are that the former does not assume a model for the probability of being treated, or rely on the results of the common support, and the weights obtained can be used in any regression method.

Formally, we have a binary treatment variable $D_i \in \{0,1\}$, that takes the value of one if the observation “i” receives the treatment and zero in the other case. With this, the average treatment effect would be estimated as a difference of the outcome variable (Y) for the treatment and control groups as $ATE = E[Y(1) - Y(0)|D = 1]$ which can be expressed as $ATE = E[Y(1)|D = 1] - E[Y(0)|D = 1]$. Since we do not observe the counterfactual denoted by $E[Y(0)|D = 1]$, we would need an estimator for it. The EB methodology generates the following estimator for the counterfactual:

$$E[\overline{Y(0)}|D = 1] = \frac{\sum_{\{i|D=0\}} Y_i \omega_i}{\sum_{\{i|D=0\}} \omega_i}$$

Where Y_i denotes the outcome variable and ω_i represents the weight of the observation “i”. The weight vector is obtained as a solution to the following minimization problem:

$$\min H(\omega) = \sum_{\{i|D = 0\}} h(\omega_i)$$

Restricted to the following conditions to assure the balance of the independent variables between the treatment and control group:

$$\sum_{\{i|D = 0\}} \omega_i c_{ri}(X_i) = m_r ; r \in 1, \dots, R$$

$$\sum_{\{i|D = 0\}} \omega_i = 1 ; \omega_i \geq 0$$

Where $h(\omega_i)$ is a function that measures the distance between the distribution of weights of the control group (ω_i) and the distribution of the base weights which are computed as $q_i = 1/n_0$, where n_0 is the size of the treatment group. Additionally, $c_{ri}(X_i) = m_r$ imposes a series of “R” restrictions give for the “R” moments that wish to be balanced between the groups.

Hainmueller (2012) recommends the use of the entropic divergence function in the minimization problem, which is denoted by $h(\omega_i) = \omega_i \log(\omega_i/q_i)$. Under these conditions, the objective function is convex and the Lagrangian equation solution would be the weight for each observation given by:

$$\omega_i = \frac{q_i \exp(-\sum_{r=1}^R \lambda_r c_{ri}(X_i))}{\sum_{\{i|D=0\}} q_i \exp(-\sum_{r=1}^R \lambda_r c_{ri}(X_i))}$$

Using this weight vector, we compute a weighted ordinary least squares regression of the form:

$$Y_i = \beta_0 + \beta_1 D_i + \delta X_i + \varepsilon_i$$

Where Y_i is the outcome variable, D_i is the treatment status dummy, X_i is a matrix of relevant covariates, and ε_i is the error term.

ANNEX D: DATA COLLECTION AND QUALITY ASSURANCE

Data Collection Process

Quantitative data collection for the first follow-up endline was carried out by Market Xcel, a local survey partner subcontracted by Management Systems International, with close collaboration, supervision, and quality oversight provided by the evaluation team. The surveys were administered face to face at the participants' homes, which are usually where the businesses are set up. The survey took between 30 and 45 minutes to complete for each participant. The enumerators obtained oral informed consent from each participant prior to the start of the survey to confirm willingness to participate. A small token of appreciation (equivalent in value to USD \$3) was provided to each respondent after the survey was completed. Prior to the start of data collection, the survey instrument was pre-tested and enumerators were trained over the course of three days. Data quality assurance processes were put in place internally by Market Xcel, and independently by the evaluation team.

Baseline data collection conducted by FHI 360 took place between August 2013 and June 2014, and was completed prior to randomization of the treatment arms. The first follow-up survey conducted by the E3 Analytics and Evaluation Project took place six months after the end of the WLSME activity, between May and June 2017, with a 95 percent response rate. Non-response rates were evenly distributed between treatment and comparison groups. Simultaneously to the start of the first follow-up survey, the villages from the comparison group were visited to create a listing of 500 eligible cashew processing SMEs from which 100 were then randomly to be included in the sample. The survey team visited these new respondents at the same time as the rest of the comparison group, and the same survey instrument was used.

Qualitative data collection was led by Swarna Rajagopalan, a qualitative researcher from the E3 Analytics and Evaluation Project, with logistical support from Market Xcel. A subsample of WLSME participants from each treatment arm was randomly selected to participate in the focus group discussions (FGDs). The number of participants recruited was inflated to account for a no-show rate of 50 percent. The FGDs were held in a convenient and central location in the Panruti block. Participants' transportation expenses were covered and light refreshments were provided. Prior to the start of each FGD, each participant received and signed an informed consent form ensuring confidentiality and voluntary participation. The FGDs lasted 60 minutes and consisted of open-ended, guided questions. Each FGD was audio recorded and transcribed verbatim into Tamil, then translated into English. Translations were audited independently by another member of the evaluation team.

Data Quality Assurance

Standard operating procedures for data collection were followed by the survey firm, including verification procedures conducted on site. Specifically, the survey firm employed the following set of quality control procedures:

- The field manager and supervisors constantly managed the workflow to ensure all enumerators followed the agreed upon timeline and procedures. Field managers were in contact with the evaluation team to find proper solutions to any unexpected challenges.
- Supervisors accompanied enumerators to at least 10 percent of the interviews conducted.
- Logic checks and skips were automated in the electronic survey to minimize surveyor data entry error.

- Data from completed surveys were uploaded and reviewed for completion daily.
- Datasets and progress reports were submitted to the evaluation team twice per week. The progress reports included the number of contact attempts for pending surveys and reasons for pending status.

The evaluation team provided additional oversight and monitoring of the quality of data collected:

- The evaluation team accompanied enumerators during interviews at regular intervals. The evaluation coordinator observed the enumerators' familiarity with and comprehension of the questionnaire and clarity in asking questions.
- Each week, the evaluation team conducted additional checks to compare each enumerator's average performance to the total sample averages in terms of interview length, number of completed codes, number of "do not knows," scale usage, section skips, and ranges of numerical values. No significant outliers were found.

ANNEX E: TEAM COMPOSITION

A five-person team carried out core activities for this evaluation, supported by the Home Office teams from the E3 Analytics and Evaluation Project. The specific qualifications and roles for each team member are listed below. Each evaluation team member signed a conflict of interest disclosure statement that is retained by the MSI home office and available upon request. In addition, the Indian survey research firm, Market Xcel, conducted the survey data collection and provided logistical support on the FGDs.

Principal Investigator

Dr. Alberto Chong, an evaluation specialist external to USAID, holds a Ph.D. in Economics and is currently a professor in the Department of Economics at the Andrew Young School of Policy Studies at Georgia State University. Dr. Chong has worked extensively with the Inter-American Development Bank and the World Bank, has published several academic papers, including impact evaluation results with respect to gender, microfinance, and the private sector, and has applied advanced econometric and quantitative evaluation techniques to data collected in developing country contexts. Dr. Chong was primarily responsible for improving the quality of the evaluation design and minimizing its limitations, particularly with respect to the evidence to be obtained regarding causality and the attribution of outcomes to the project.

Evaluation Coordinator

Irene Velez holds a Master's Degree in International Development Policy and has over seven years of global work experience designing and implementing impact evaluations. She has technical knowledge of different experimental and quasi-experimental evaluation methods, as well as practical experience managing the execution of these evaluations. She has conducted large-scale data collection efforts, including hiring, training, and supervising survey teams, as well as providing supervision and quality assurance oversight to subcontracted local survey firms. Ms. Velez was primarily responsible for overseeing and coordinating the execution of the evaluation design, ensuring efficient and timely reporting, and monitoring fidelity of the evaluation design. She is also the main technical point of contact on the evaluation team for USAID and the implementing partner.

Local Qualitative Researcher

Swarna Rajagopalan is a qualitative researcher with a Ph.D. in Political Science, and broad research experience on economic and political gender issues. Dr. Rajagopalan has prior experience conducting focus groups in the region and is fluent in Tamil, Hindi, and English. Dr. Rajagopalan was primarily responsible for the qualitative component of this evaluation, including facilitating the focus groups discussions, conducting the data analysis, and summarizing key findings.

Quantitative Data Analyst

Angelo Cozzubo holds a Master's Degree in Economics and has five years of research experience in applied microeconomics and conducting econometric analysis of panel data and poverty assessments. For this evaluation, Mr. Cozzubo supported the analysis of quantitative data collected for this impact evaluation.

Survey Research Firm

Following a competitive procurement process, MSI subcontracted with Market Xcel to provide survey research services for the evaluation team in India. This included conducting the survey field work and logistical support for the FGDs.¹⁴ Market Xcel is a New Delhi based company, with operations across India. A regional office in Chennai provides a wide range of data collection and analysis services for international and donor organizations. Market Xcel is part of the Market Research Society in India.

Home Office Support

Home Office support by the E3 Analytics and Evaluation Project team was also provided to the core evaluation team, including technical reviews, research assistance, qualitative data analysis, administrative oversight, management of the survey research firm, and logistical support.

¹⁴ MSI subcontracted a different company from the baseline survey data collection phase conducted by FHI360. The baseline survey data firm was Sigma Research and Consulting.

ANNEX F: REFERENCES

- Angrist, Joshua D., and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Kartika Sari Juniwyaty, and Bertil Tungodden (2012). "Business Training in Tanzania: From Research Driven Experiment to Local Implementation." *Journal of African Economics*, forthcoming.
- Bruhn, Miriam, Dean Karlan, and Antoinette Schoar (2012). "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico." Yale Economics Department Working Paper 100.
- Bruhn, Miriam, and Bilal Zia (2013). "Stimulating Managerial Capital in Emerging Markets: The Impact of Business Training for Young Entrepreneurs." *Journal of Development Effectiveness*, Taylor & Francis Journals 5, no. 2 (June): 232-266.
- Buvinic, Mayra, Rebecca Furst-Nichols, and Emily Courey Pryor. 2013. *A Roadmap for Promoting Women's Economic Empowerment*. New York: United Nations Foundation.
- Caselli, Francesco, and Nicola Gennaioli. 2005. "Credit Constraints, Competition, and Meritocracy." *Journal of the European Economic Association* 3(2-3): 679-689.
- Cirera, Xavier, and Qursum Qasim. 2014. *Supporting Growth-Oriented Women Entrepreneurs: A Review of the Evidence and Key Challenges*. Innovation, Technology, and Entrepreneurship Policy Note 5. Washington, D.C.: World Bank Group.
- Cole, Shawn A., Thomas Sampson, and Bilal Zia. 2011. "Prices or Knowledge? What Drives Demand for Financial Services in Emerging Markets?" *Journal of Finance* 66, no. 6: 1933–1967.
- De Mel, Suresh, David McKenzie, and Christopher Woodruff (2014) "Business Training and Female Enterprise Start-up, Growth, and Dynamics: Experimental Evidence from Sri Lanka." *Journal of Development Economics* 106: 199-210.
- _____. (2009) "Measuring Microenterprise Profits: Must We Ask How the Sausage is Made?" *Journal of Development Economics* 88(1): 19-31.
- Dehejia, Rajeev (2005). "Practical Propensity Score Matching: A Reply to Smith and Todd." *Journal of Econometrics* 125(1): 355-364.
- Drexler, Alejandro, Greg Fischer, and Antoinette Schoar (2012) "Keeping it Simple: Financial Literacy and Rules of Thumb", Manuscript, London School of Economics.
- Field, Erica, Seema Jayachandran and Rohini Pande (2010) "Do Traditional Institutions Constrain Female Entrepreneurship? A Field Experiment on Business Training in India." *American Economic Review Papers and Proceedings* 100(2): 125-29.
- Gathani, Sachin, and Dimitri Stoelinga (2013) "Export Similarity Networks and Proximity Control Methods for Comparative Case Studies." *Journal of Globalization and Development* 3(2): 1-39.
- Gine, Xavier and Ghazala Mansuri (2014) "Money or Ideas: A Field Experiment on Constraints to Entrepreneurship in Rural Pakistan." Policy Research Paper, World Bank 6959, Washington, DC.

- Hainmueller, Jens (2012). "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies" *Political Analysis* 20(1): 25-46.
- Heckman, James J., Hidehiko Ichimura, and Petra Todd (1998) "Matching as an Econometric Evaluation Estimator." *The review of Economic Studies* 65(2): 261-294.
- Heckman, James J., Hidehiko Ichimura, and Petra Todd (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *The Review of Economic Studies* 64(4): 605-654.
- Hejazi, Seyed Amir Mousavi, Mohamad Mehdi Maleki, and M. J. Naeji (2012) "Designing a Scale for Measuring Entrepreneurial Leadership in SMEs." International Conference on Economics Marketing and Management, 28: 71-77. Singapore: IACSIT Press.
- Karlan, Dean and Martin Valdivia (2012) "Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions." *The Review of Economics and Statistics* 93(2): 510-527.
- Kulkani, Vami (2011) "Women's Empowerment and Microfinance: An Asian Perspective Study." Occasional Paper 13, International Fund for Agricultural Development, Washington, DC
- LaLonde, Robert J. (1986) "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review* 604-620.
- Lee, Myoung-Jae (2005) "Micro-econometrics for Policy, Program, and Treatment Effects." Oxford University Press on Demand.
- Mayoux, Linda (2001) "Microfinance and the Empowerment of Women: A Review of the Issues." International Labor Organization, Geneva.
- McKenzie, David (2012) "Beyond Baseline and Follow-up: The Case for More T in Experiments." *Journal of Development Economics* 99(2): 210-221.
- McKenzie, David and Christopher Woodruff (2014) "What Are We Learning From Business Training and Entrepreneurship Evaluations Around the Developing World?" World Bank Research Observer, Forthcoming.
- Rosenbaum, Paul R., and Donald B. Rubin (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1): 41-55.
- Rubin, Donald B. (1974) "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5): 688.
- Smith, Jeffrey A., and Petra E. Todd (2005). "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125(1): 305-353.
- Waldinger, Fabian (2014) "Differences-in-Differences." Lecture Notes, University of Warwick, United Kingdom.
- Zaslavsky, A. (1988) "Representing Local Reweighting Area Adjustments of Households." *Survey Methodology* 14(2): 265-288.

U.S. Agency for International Development
1300 Pennsylvania Avenue, NW
Washington, DC 20004