

A WORLD BANK STUDY



Lessons Learned and Not Yet Learned from a Multicountry Initiative on Women's Economic Empowerment

Sara Johansson de Silva,
Pierella Paci, and Josefina Posadas



THE WORLD BANK

Lessons Learned and Not Yet Learned
from a Multicountry Initiative on
Women's Economic Empowerment

A WORLD BANK STUDY

Lessons Learned and Not Yet Learned from a Multicountry Initiative on Women's Economic Empowerment

Sara Johansson de Silva, Pierella Paci, and Josefina Posadas



THE WORLD BANK
Washington, D.C.

© 2014 International Bank for Reconstruction and Development / The World Bank
1818 H Street NW, Washington DC 20433
Telephone: 202-473-1000; Internet: www.worldbank.org

Some rights reserved

1 2 3 4 16 15 14 13

World Bank Studies are published to communicate the results of the Bank's work to the development community with the least possible delay. The manuscript of this paper therefore has not been prepared in accordance with the procedures appropriate to formally edited texts.

This work is a product of the staff of The World Bank with external contributions. Note that The World Bank does not necessarily own each component of the content included in the work. The World Bank therefore does not warrant that the use of the content contained in the work will not infringe on the rights of third parties. The risk of claims resulting from such infringement rests solely with you.

The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Nothing herein shall constitute or be considered to be a limitation upon or waiver of the privileges and immunities of The World Bank, all of which are specifically reserved.

Rights and Permissions



This work is available under the Creative Commons Attribution 3.0 Unported license (CC BY 3.0) <http://creativecommons.org/licenses/by/3.0>. Under the Creative Commons Attribution license, you are free to copy, distribute, transmit, and adapt this work, including for commercial purposes, under the following conditions:

Attribution—Please cite the work as follows: Johansson de Silva, Sara, Pierella Paci, and Josefina Posadas. 2014. *Lessons Learned and Not Yet Learned from a Multicountry Initiative on Women's Economic Empowerment*. World Bank Study. Washington, DC: World Bank. doi:10.1596/978-1-4648-0068-9. License: Creative Commons Attribution CC BY 3.0

Translations—If you create a translation of this work, please add the following disclaimer along with the attribution: *This translation was not created by The World Bank and should not be considered an official World Bank translation. The World Bank shall not be liable for any content or error in this translation.*

All queries on rights and licenses should be addressed to the Publishing and Knowledge Division, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2625; e-mail: pubrights@worldbank.org.

ISBN (paper): 978-1-4648-0068-9
ISBN (electronic): 978-1-4648-0070-2
DOI: 10.1596/978-1-4648-0068-9

Cover photo: © INPET, CAPLAB, and CELATS. Used with permission from INPET, CAPLAB, and CELATS.

Library of Congress Cataloging-in-Publication Data

CIP data have been requested.

Contents

<i>Acknowledgments</i>		<i>vii</i>
<i>About the Authors</i>		<i>ix</i>
<i>Abbreviations</i>		<i>xi</i>
Executive Summary		1
Background		1
Lessons on Impact		2
Lessons on Pilot Interventions		4
Note		5
Chapter 1	Background	7
	Introduction	7
	Why Focusing on Women’s Economic Empowerment and Why Using Pilots?	8
	The Nuts and Bolts of the RBI Pilots	10
	Notes	15
Chapter 2	What We Have Learned about RBI Intervention Impacts and What We Could Have Learned, But Did Not	17
	Economic Opportunities	22
	Human Capital Endowments and Assets	23
	Agency	24
	Note	24
Chapter 3	Lessons Learned on Pilot Design, Implementation, and Evaluation	25
	Risks at the Design Stage	25
	Issues Surrounding Impact Evaluation Methodology	29
	The Importance of Monitoring	38
	Notes	39

Chapter 4	Conclusions	41
Appendix A	Country Case Study Summaries	45
Appendix B	Technical Concepts in Impact Evaluation Design	59
Appendix C	Power Calculations for Mekong RBI	61
	<i>Bibliography</i>	63
Boxes		
1.1	When Is an Intervention a “Pilot”?	9
2.1	Key Findings on the Impact of the Interventions	22
3.1	Key Requirements for Successful Pilot Implementation	26
3.2	Guidelines for Designing IE Pilots	29
A.1	The World Bank’s Gender Equity Model	45
Figure		
3.1	Alternative Policy Recommendations When Control Group Comparisons Find No Statistical Significant Impacts	38
Tables		
ES. 1	Five Results-Based Initiatives: Objectives and Approaches	2
1.1	The Result-Based Initiatives: Objectives, Approach, and Outcomes	11
1.2	Commonalities and differences in Design, Implementation, and Impact Evaluation	12
1.3	Impact Evaluation Methods	14
1.4	Examples of Female Economic Empowerment Questions	14
2.1	Impacts on Economic Opportunities	18
2.2	Impacts on Human Capital Endowments and Resources	19
2.3	Impacts on Agency	20
3.1	Number of Beneficiaries	27
3.2	Project Delays	28
3.3	Budget and Time Constraints versus Quality of Impact Evaluation	31
3.4	Female Economic Empowerment Indicators	33
3.5	Observations in Treatment and Control Groups	35
3.6	Intervention and Data Collection Levels	37

Acknowledgments

This paper draws on the experiences, results, and lessons learned from the implementation of the Results-Based Initiatives (RBI). The RBI were among the first projects designed as part of the gender action plan (GAP). The RBI in the Arab Republic of Egypt, Kenya, Liberia, Mekong, and Peru were funded by a Development Grant Facility grant to the United Nations Women and the International Center for Research on Women (ICRW). UN Women (then called UNIFEM) was responsible for the design and implementation of these RBI, while ICRW was responsible for the impact evaluations; the World Bank had an overall supervisory role. Mayra Buvinic, Lucia Fort, Andrew Morrison, and Waafas Ofosu-Amaah from the World Bank conceived the initiative while developing the Gender Action Plan. Involved in the implementation of the Development Grant were Waafas Ofosu-Amaah (World Bank), Hiska Reyes (World Bank), Joanne Sandler (UNIFEM), Letty Chiwara (UNIFEM), Caroline Horekens (UNIFEM), Anne Golla (ICRW), and Anju Malhortra (ICRW). Several persons from the three organizations were involved in each of the pilots: Lorena Barba (UNIFEM; Peru), Helene Carlsson Rex (World Bank; Mekong), Carmela Chung (UNIFEM; Peru), Maria Elizabeth Dasso (World Bank; Peru), Izeduwa Derex-Briggs (UNIFEM; Liberia), Elisa Fernández (UNIFEM; Egypt, Kenya, Liberia, Mekong, Peru), Lucia Fort (World Bank; Peru), Anne Golla (ICRW; Egypt, Kenya, Liberia, Mekong, Peru), Caroline Horekens (UNIFEM; Egypt, Kenya, Liberia, Mekong, Peru), Zebib Kavuma (UNIFEM; Kenya), James C. Knowles (World Bank consultant; Mekong), Andrew Morrison (World Bank; Egypt), Maya Morsy (UNIFEM; Egypt), Sahar Nasr (World Bank; Egypt), Greg Ngungi (UNIFEM consultant; Kenya), Waafas Ofosu-Amaah (World Bank; Kenya, Liberia), Ruth Okoth (UNIFEM; Kenya), Ryratana Rangsitpol (UNIFEM; Mekong), Hiska Reyes (World Bank; Egypt, Kenya, Liberia, Mekong, Peru), Meredith Saggars (ICRW; Kenya, Mekong), Asa Torkelsson (World Bank; Kenya), Martin Valdivia (GRADE, World Bank consultant; Peru). This paper benefited from comments by Jesko Hentschel and Mattias Lundberg (peer reviewers), Stefan Agesborg, Elena Bardasi, Jeni Klugman, Andrew Morrison, Waafas Ofosu-Amaah, and Hiska Reyes.

About the Authors

Sara Johansson de Silva is an economist and an international consultant whose work experience and research interests have focused on the intersection of growth, job creation, skills development, and social protection policies. Between 1999 and 2003, she was a senior economist at the World Bank, working in the Middle East and North Africa Department on growth, employment, and poverty analysis. Before joining the World Bank, she was an economist at the Organisation for Economic Co-operation and Development from 1997 to 1999, where she worked on industrial and innovation policy and aid effectiveness. She holds a PhD in economics and a Master of Business and Economics from the Stockholm School of Economics.

Pierella Paci is a lead economist in the Office of the Poverty Reduction and Economic Management Vice President. An economist by training, Pierella began her career as a professor of economics at the University of Sussex in the United Kingdom and at City University in London. She has written extensively in the areas of labor economics, gender economics, inequality, and poverty. Since joining the World Bank she has been instrumental in developing the corporate agenda on jobs, migration, poverty, gender equality, and equity more broadly and has been the Manager of the Gender and Development Group. She holds a degree in economics from the University of Rome, in Italy, and a PhD in economics from the University of Manchester in the United Kingdom.

Josefina Posadas is an economist in the Gender and Development Group of the World Bank. Her area of expertise is labor economics, and since joining the World Bank in 2008 she has worked on issues related to gender equality, entrepreneurship, and poverty. Between 1996 and 2002, she worked at the Universidad Nacional de La Plata in Argentina where she was an associate professor for advanced microeconomics since 1999. During those years, Josefina also advised different government offices of Argentina, both at the local and at the national level, on employment and fiscal federalism matters. She holds a PhD in economics from Boston University and a Master in Economics from the Instituto Torcuato Di Tella in Argentina.

Abbreviations

AEDE	Agency for Economic Development and Empowerment (Liberia)
ARS	Agricultural Relief Services
BDS	business development skills
BH	bamboo handicrafts
BHP	bamboo handicraft product
CARI	Liberia's Central Agricultural Research Institute
CIMA	Centre for International Market Access
COFOPRI	Organismo de Formalización de la Propiedad Intelectual
DEN-L	Development Education Network Liberia
FAO	Food and Agriculture Organization
FGD	focus group discussion
GAP	gender action plan
GCWG	Ganta Concern Women's Group
GEM	gender equity model
GEME	gender equity model Egypt
GT	general training
HH	households
ICRW	International Center for Research on Women
IDI	in-depth interviews
IE	impact evaluation
IGA	income-generating activities
KGT	Kenya Gatsby Trust
NGO	nongovernmental organization
PI	Prosperity Initiative
RBI	results-based initiatives
RCT	randomized control trial
RPRCP	Real Property Rights Consolidation Project
TA	technical assistance
UN	United Nations

UNESCO	United Nations Educational, Scientific, and Cultural Organization
UNIFEM	United Nations Development Fund for Women
UPRP	Urban Property Rights Project

Executive Summary

Background

The results-based initiatives (RBI), launched in 2007, were a pioneering attempt to provide comprehensive, coherent, and rigorous evidence on effective interventions to foster the economic empowerment of women. Increasing women's access to earning opportunities, productive resources, and decision-making power can result in a more productive use of resources, more investment in children's welfare, and more representative public institutions. As such, female economic empowerment contributes to economic growth and poverty reduction, and jobs for women stimulate development.

The RBI comprised five small pilots with built-in impact evaluation designed to identify what works best in promoting better outcomes for women as entrepreneurs, wage earners or farmers, under different country contexts. The interventions shared a number of common elements so that general conclusions could be derived, but also differed in a number of ways so that different hypotheses could be validated. The interventions are summarized in table ES.1 and described more extensively in the appendix A. The program was a broad, collaborative effort between the United Nations Development Fund for Women (then UNIFEM, now UN Women), the International Center for Research on Women (ICRW), and the World Bank and country partners, including governments and nongovernmental organizations (NGOs).

The program was an innovative experiment in an important policy area. While there is a clear rationale for policy interventions to help remove constraints to women's economic empowerment, knowledge of what interventions work best in different settings remains limited. When the RBI were conceived, rigorous evidence in this area was close to nonexistent, because no systematic impact evaluations had been carried out in developing countries.

However, the RBI fell short of meeting several of their ambitious objectives, for reasons ranging from too much optimism and mistakes in design and budgeting to implementation failures and lack of continuous monitoring. While pilot interventions can fail to show impacts—indeed, they are riskier interventions applied at a small scale to test whether something works or not—these structural problems cast doubts on the reliability of some results and make it impossible to derive a coherent and generalizable message. In this respect, the RBI represent a

Table ES.1 Five Results-Based Initiatives: Objectives and Approaches***Egypt, Arab Rep.: Promoting Gender Equity and Productivity in Private Firms—the Gender Equity Model Egypt (GEME)***

Objective: Promote gender equity in private sector wage employment (working conditions, access to jobs, professional development, and participation in decision making in firms).

Approach: Help firms formulate and achieve gender equity goals and deliver targeted training.

Liberia: Value-Added Cassava Enterprise for the Ganta Concern Women's Group

Objective: Raise profitability of cassava production.

Approach: Give women's groups access to productive inputs (land, labor, tools, cassava cuttings), provide training on production and processing techniques, build a processing plant.

Kenya: Export Competitiveness of Women Beadworkers

Objective: Increase profitability of *Maasai* women's beadwork activities.

Approach: Provide training, mentoring, and support to (i) improve design, (ii) enhance marketing and business skills, and (iii) identify and access larger markets (including for export).

Mekong: Improving Bamboo Handicraft Value Chains for Women's Economic Empowerment

Objective: Enhance productivity and earnings of women in bamboo handicraft.

Approach: Provide training, mentoring, and support to (i) improve design; (ii) upgrade production techniques (including provision of tools) and organizational management; and (iii) identify and access larger markets.

Peru: Strengthening the Economic Empowerment of Women Microentrepreneurs in Lima

Objective: Enhance productivity of women microentrepreneurs in Lima and strengthen their power in household decision making.

Approach: Provide women microentrepreneurs (with land title) with training in business practices, networking, marketing, and life skills.

costly missed opportunity. Nonetheless, the RBI experiences provide many useful lessons for future pilots focusing on women's economic empowerment as well as for other policy areas.

This paper highlights lessons from the RBI with respect to both the impact of the interventions and dos and don'ts in the design and implementation of pilots. This paper focuses on three issues: How effective have different policy interventions been in terms of strengthening female economic empowerment? What are the main challenges involved in carrying out small-scale pilots with impact evaluations, especially with a gender focus? And what have we learned from the RBI that can help navigating these challenges more effectively in future interventions?

Of the eight RBI pilots initially envisaged, only five have so far been implemented and evaluated—in the Arab Republic of Egypt, Liberia, Kenya, Mekong, and Peru. As summarized in table ES.1, these programs focused largely on training, but also cover gender sensitization, improved business practices, and technical skills.

Lessons on Impact

The impact of the interventions on female economic empowerment has been limited and mixed. Using the framework of the *World Development Report*:

Gender Equality and Development (World Bank 2011a), the potential impact of an RBI on women's economic empowerment can be assessed along three dimensions:

- *Economic opportunities*, defined as
 - Higher productivity, turnover, and earnings, or
 - Higher wages and better career paths.
- *Human capital endowments and assets*, which contribute to higher income and poverty reduction over the short or long run and are defined as
 - Better labor market skills
 - Higher spending on health, nutrition, or education of children, or
 - Increased assets.
- *Agency*, which includes
 - More bargaining power within the household
 - Formation of and participation in groups and networks, and
 - Stronger self-perception of having skills, opportunities, and power, which is an objective in itself and a potential means of increasing access to earnings opportunities.

However, the impact evaluations suggest that, with the exception of Peru, the interventions did not significantly increase women's earnings and had little impact on other dimensions of their economic empowerment.¹

Economic opportunities. The interventions did not generally increase women's earnings. Sales and revenues of their enterprises did not grow even when the quantity and/or quality of their products increased, due to failure of the intervention to provide access to a broader market. And in Egypt, the female wage disadvantage did not decline. Only in Peru did general training combined with technical assistance succeed in bringing higher earnings to the beneficiaries.

Human capital endowments and asset control and ownership. Women who received training generally appreciated the access to new information and felt their skills had increased. In the Mekong Valley and Peru, training also resulted in improved business practices. However, in countries where the impact on investments in children's education/health or on relative bargaining power within the household was measured, there was no statistically significant effect.

Voice and agency. Across all countries, women who participated in the program perceived increases in their skills and/or their involvement in business associations and networks. In Peru, women's participation in household decisions increased, and in the Mekong Valley, the number of households involved in producer groups increased.

However, it would be wrong to conclude that these interventions were not effective. The lack of robust positive impact may be due to the evaluations being conducted too soon to show fully the long-term effects of the interventions, or to problems in the design, implementation, or measurement of pilot outcomes. These issues are described in more detail in the following section.

Lessons on Pilot Interventions

The lessons learned from the RBI can be grouped under three headings: (i) risks at the design stage—including objectives and use of resources, (ii) issues related to the impact evaluation methodology, and (iii) the importance of adequate monitoring.

Three main messages emerge. First, for a pilot to be able to generate meaningful conclusions, it is essential to align resources with expectations. Having realistic expectations from the outset on costs, the risk of delays, and the potential impact of the intervention is essential to the success of any pilot. It is also important that these expectations are aligned with the available resources and that the time frame is long enough for the intervention to be fully implemented and its impacts observable. Meaningful rigorous impact evaluations require time and resources to bear fruit. Second, even minor methodological weaknesses in the design and implementation of the impact evaluation component may invalidate the findings and nullify its value. Combining quantitative data with qualitative information could considerably strengthen the potential for assessing impacts and for understanding the underlying transmission channels. Some conclusions can then be drawn even when more rigorous quantitative results fail. Finally, close and continuous monitoring during project implementation is essential to detect shortcomings in the design and implementation strategy. But it is also essential for the monitoring to be combined with an effective mechanism to feed the findings back into the design and implementation of the project. This is necessary to ensure that problems can be resolved early enough to give the intervention a fair chance at succeeding.

The amount of resources and time needed to deliver interventions and evaluate their impacts should not be underestimated. Pilots are potentially a cost-effective method of evaluation. They are quicker and cheaper than large interventions, and thus suitable to investigate new areas and test riskier interventions on a smaller scale. However, the evaluation component means that pilots are more resource intensive and have higher overhead costs than same scale programs without evaluation. The RBI experiences show the need for sufficient financial resources, capacity, and time. A good impact evaluation also requires an adequate sample and experienced teams of designers, implementers, and evaluators. Resources are needed for frequent monitoring during implementation and for ensuring wide dissemination of results—so that the knowledge generated can be used by others. Finally, as stated above, the time frame is critical, because the full impacts of complex interventions take time to emerge.

The RBI were overambitious regarding what could be achieved with a limited budget and a short time frame. Several pilots adopted a value chain approach and focused on achieving organizational change and fostering the adoption of new methods of doing business. These are long-term processes that require time to deliver outcomes, especially when targeted groups have very low capacity, as in Liberia and Kenya. Too short a time frame can thus prevent the intervention from achieving its full objective and compromise the findings of the evaluation.

Information should be collected at different times (immediate, intermediate, final) and at multiple levels (individual, group), include different types (quantitative, qualitative), and cover both direct and indirect outcomes. Although more expensive to collect, a multilevel approach provides a more complete picture of an intervention's transmission channels. Intermediate data collection and qualitative information can also provide a hedge in case of unforeseen problems in the final impact evaluation and can help to explain insignificant results of quantitative methodologies or point toward other unexpected impacts. Finally, collection of data for direct and indirect outcomes can mitigate measurement problems and should be comprehensive, ideally covering indicators for economic opportunities, endowments, and agency.

Project monitoring is fundamental to program implementation as an “early warning system,” and in evaluation as a complement to quantitative information. As mentioned above, monitoring data can be a useful complement to evaluation data to understand the mechanisms that caused a program to have a certain impact, or not. Lack of regular monitoring was a major weakness of the RBI, with two negative consequences: first, without this complementary information, it was difficult to determine whether the interventions failed because they were inherently ineffective or ill suited to the particular context, or because they were not implemented as planned. Second, without continuous monitoring, the teams were unable to identify implementation problems early enough to be able to modify the programs; Liberia is an example of such a case.

Note

1. “Significant” means that the impact evaluation did not show a statistically significant impact for program beneficiaries.

Background

Introduction

The *results-based initiatives* (RBI) began in 2007 as a pioneering attempt to test, across a range of country contexts, the potential impact of interventions to foster women's economic opportunities by enhancing their human and financial endowments. The program was a collaborative effort between the World Bank, the United Nations Development Fund for Women (then UNIFEM, now UN Women), and the International Center for Research on Women (ICRW). It consisted of country-tailored pilots implemented in close collaboration with country governments and local nongovernmental organizations (NGOs) and supplemented by rigorous impact evaluation.¹

The program was innovative in a number of ways. It represented a shift away from the traditional focus on gaps in human capital (health and education) that had characterized policy making in gender inequality, and toward direct attention on women's economic empowerment. It was also among the first attempts to generate rigorous evidence on what type of interventions works best in different economic and social contexts.

The idea that policy making should be based on clear evidence of what works, rather than ideological or theoretical considerations, is now central to the design and implementation of policies by development organizations and governments. The use of pilots to pretest the potential effectiveness of large-scale programs is consequently becoming standard practice, especially when the proposed intervention is innovative or is applied in new contexts. However, when the RBI started, results-based policy making was a relatively novel concept and the limited rigorous evidence that existed focused on training programs in developed countries and/or during economic downturns (World Bank 2011b).² In addition, there was almost no evidence on impacts from business training interventions (McKenzie 2010; McKenzie and Woodruff 2012). The RBI program provided a first attempt to systematically use small-scale *pilots* to evaluate the relative effectiveness of different interventions in the area of women's economic empowerment across

different countries. Given the scale and innovative character of the program, the RBI had the potential to provide useful lessons in two areas: (i) the **impact** of different interventions on female economic empowerment and (ii) the **design and implementation** of small-scale pilots, especially with a gender focus.

The main objective of this report is to compile these lessons to advise policy makers on *what works* and *what doesn't* in terms of interventions to increase the economic empowerment of women and to guide future researchers in the *dos and don'ts* of pilot design and implementation. The analysis draws on a set of background reports on individual pilots and on two summary reports prepared by UNIFEM and ICRW (Golla 2011; UN Women 2011). This report does not pretend to provide a full set of generalizable lessons, partly because it is too early to judge outcome sustainability. Instead, it aims to synthesize useful insights from these ambitious initiatives. The main messages on impact are broadly in line with a survey paper by McKenzie and Woodruff (2012), which was conducted in parallel and draws lessons from 13 recent pilots on business skills training, including the Peru RBI. However, the scope of this report goes beyond the focus on impact.

Why Focusing on Women's Economic Empowerment and Why Using Pilots?

Economic growth and development have reduced gender inequality globally across several dimensions, but women continue to face considerable disadvantages in a number of areas (World Bank 2011a). For example, despite increases in female labor force participation, the gender gap in economic opportunities continues to be substantial, especially in the developing world. This is clearly objectionable from a human rights point of view, but also from an economic perspective, because gender equality fosters economic growth and development (World Bank 2001, 2006, 2011a).

In particular, jobs for women are likely to be “better” for development than jobs going to men (World Bank 2012). Giving women more control over household resources results in more investments in the education and health of children (Atkin 2009; Bobonis 2009; Rubalcava, Teruel, and Thomas 2009; Duflo 2000; Duflo and Udry 2004; Lundberg, Pollack, and Wales 1997; Thomas 1990). Women are also more gender neutral in these investments (World Bank 2011a). Increasing female economic opportunities also strengthens women's bargaining power within the household, leading to more cooperation among household members and potentially higher levels of welfare (Doepke and Tertilt 2011). Gender equality in employment also implies better use of existing productive resources—female labor and human capital—and higher aggregate productivity (World Bank 2011a). Moreover, economic empowerment of women has important indirect effects on other dimensions of gender equality.

However, increasing economic opportunities for women requires a complex set of interventions that addresses the multiple constraints they face in entering the labor market. Depending on context, these include measures designed to

(i) ease women's *time constraints* by providing child care and improving infrastructure; (ii) improve women's endowments by enhancing their access to *productive resources*—especially to land and to credit—and *skills development*, predominantly through training; and (iii) tackle information problems and institutional biases that work against women (World Bank 2011a).

In this vein, the RBI program took a multipronged approach that addressed several aspects of female disadvantages. The program aimed at increasing access to endowments by raising women's **skills**. This increase was also expected to directly contribute to raising beneficiaries' labor earnings and agency, and indirectly improve the endowments of the next generation of girls. The RBI program also aimed to directly foster women's **economic opportunities** as farmers, formal nonagricultural workers, or entrepreneurs and to indirectly—through higher earnings and skills—raise their **agency** within the household and in society.

From the outset, the RBI were set up as pilots. Using the characterization in Gertler et al. (2011; box 1.1), compared to a large-scale operation, each initiative was **small** in scale and in budget in order to test higher-risk approaches on a smaller set of beneficiaries, evaluate impact, and feed lessons into larger programs. They were all **strategically relevant**, because they were intended to fill important knowledge gaps. They involved interventions that were **untested** and **innovative in delivery or content**, and thus provided new knowledge about a method or context. They were meant to be **replicable/scalable**, so as to be able to use the knowledge/approach acquired through them in another setting, and potentially **influential**, because they had the potential to help develop knowledge, thinking, and action around policy interventions.

Box 1.1 When Is an Intervention a “Pilot”?

Pilots are small-scale innovative interventions, but not all small-scale interventions are pilots. What defines a *pilot* is its innovative character and the fact that it incorporates an explicit learning component, framed in terms of a clear provision for monitoring and impact evaluation within a fixed time frame. This element is often absent even from very successful small-scale programs because their focus on small populations of beneficiaries limits the potential for generalizing the findings on impacts and reduces the rationale for costly impact evaluations.

Pilots also differ from innovative one-off policy experiments with impact evaluation, which are typically not designed to be replicated or scaled up. Stand-alone policy experiments are usually conducted by the research community to test *whether* a particular policy is effective in a given context, with little attention given to *how* its impact could be replicated or enhanced. By contrast, pilots usually go beyond assessing the impact of a policy: they are used by policy implementers (government, donors, NGOs) to evaluate the lessons learned and assess the extent to which a policy known to be effective in one setting is transferrable

box continues next page

Box 1.1 When Is an Intervention a “Pilot”? *(continued)*

to another (for example, a new geographic area or other changes to the target group). Of course, lessons from research experiments can, and often are, used for further policy replication or scale-up. However, the second stage is not implicit in the design of the policy experiments.

Source: Based on Gertler et al. 2011.

The pilot approach fitted the objectives of the RBI program as it is well suited to build evidence on what works to enhance female economic empowerment, to provide the rationale for a particular intervention, and to initiate the policy dialogue. Pilots can be designed to analyze the multidimensional causes and outcomes of gender inequality and the impacts of access to resources, training, and other interventions on these outcomes. They can also be used to test female targeting versus gender neutral policies and to improve the efficiency and effectiveness of program design and implementation by testing different alternatives and tweaking different aspects of ongoing or planned programs. However, as the experiences of the RBI show, not all types of interventions can be properly implemented and evaluated within the short time frame and limited budget that characterize most pilots.

The Nuts and Bolts of the RBI Pilots

The RBI program initially comprised eight country pilots, but to date, only five of these projects—the Arab Republic of Egypt, Kenya, Liberia, Mekong (Lao People’s Democratic Republic/Cambodia), and Peru³—have been fully implemented, evaluated, and analyzed. These five are the focus of this report and are described in table 1.1 and more extensively in appendix A.

To provide a comprehensive knowledge base, the interventions needed to have common elements, but also present some variations to give them validity across different contexts. They shared common objectives, design and approach, and experienced similar issues during the implementation and the impact evaluation (IE) stage. However, the RBI were also designed to differ in several aspects. The fact that they were tailored to the country context and institutional capacity in which they were implemented meant that they targeted different aspects of economic empowerment (table 1.1). The evaluation methodology was also adapted to each intervention and additional elements of variation were introduced by unforeseen developments during implementation and evaluation. Commonalities and differences across the pilots are summarized in table 1.2.

Design

The RBI shared many design elements. They were all small programs, targeting from a minimum of 250 women in Liberia to a maximum of 1,500 in Peru. They

Table 1.1 The Result-Based Initiatives: Objectives, Approach, and Outcomes

<i>Objectives</i>	<i>Approach</i>	<i>Outcomes</i>
<i>Egypt, Arab Rep.: Gender Equity Model</i>		
Improve gender equity in access to jobs, career development, working conditions, and labor force participation	Foster good gender equity practices in the private sector by helping firms formulate and achieve gender equity goals	No changes in firms' processes (hiring, training, promotion, and so forth), but increased employee satisfaction; program was institutionalized by the government
<i>Kenya: Strengthening Export Competitiveness of Women Beadworkers</i>		
Enhance productivity and earnings in beadwork by increasing access to export markets	Provide training and mentoring to improve design, marketing, and business skills and to help identify and access larger markets	No impact at either group level (organization, revenue, capacity) or at individual level (volume/revenue of sales, food security)
<i>Liberia: Value-Added Cassava Enterprise for the Ganta Concern Women's Group</i>		
Increase economic security and livelihoods, empower participants, and promote the cassava industry as a growth sector	Help women access productive inputs (land, labor, tools, cassava cuttings); provide training on production; provide land, plant construction, and equipment; and training on processing	Lack of access to markets for farina meant that the plant was operating much below capacity and at loss; women reported that they valued literacy skills
<i>Mekong: Improving Bamboo Handicraft (BH) Value Chains for Women's Economic Empowerment</i>		
Enhance productivity and earnings of women in the BH trade	Assisting with organizational build-up, machinery, and training in technology, business skills, design, and so forth	No significant effect on BH income; but positive effects on BH sales and production
<i>Peru: Strengthening the Economic Empowerment of Women Microentrepreneurs in Lima</i>		
Enhance productivity of women microentrepreneurs and their bargaining power at household level	Providing women microentrepreneurs (with land title) with training in business practices, marketing, and life skills	Sales increased and there was modest progress in participation in networks and in access to credit; no impact on decision making within the household

Source: Based on background documents.

had limited budgets of about US\$300,000 (operational costs) in total and were expected to be relatively quick to implement and evaluate (about 2 years).⁴ They were planned to include rigorous impact evaluations and provide measurable and statistically valid evidence on program impact. Although scale-up and replication were at the heart of the RBI program, the design of the interventions did not always identify the larger population to which the intervention could be applied if successful. With the exception of Egypt, all the pilots had a common focus on (i) increasing women's access to productive resources and skills development—endowments—and (ii) strengthening links between production, processing, and marketing, for example via value chains.⁵

However, there were also many ways in which the interventions differed from one another. The program spanned very different cultural, geographic,

Table 1.2 Commonalities and Differences in Design, Implementation, and Impact Evaluation

<i>Commonalities</i>	<i>Differences</i>
Objectives	
Foster female economic empowerment by increasing endowments, earnings, and agency	Removing institutional constraints to wage gaps (Egypt, Arab Rep.) versus raising productivity and earnings of self-employment/groups (other countries)
Design and setup	
Small programs with limited budget and accelerated time frame	Cultural, geographic, social, and economic context:
Pilot approach with impact evaluation included	<ul style="list-style-type: none"> • Three regions (Asia, Africa, Latin America), low- and middle-income countries, poor and nonpoor target groups
No implicit next stage planned (scale-up, replication, dissemination)	<ul style="list-style-type: none"> • Rural (Liberia, Kenya, Mekong) versus urban/capital city (Egypt, Arab Rep., Peru)
Focus on training/capacity building	<ul style="list-style-type: none"> • Low capacity (Liberia, Kenya, Mekong) versus higher capacity (Egypt, Arab Rep., Peru)
Value chain approach (except Egypt, Arab Rep.)	
Implementation	
Targeted beneficiaries limited in number	Different capacity of implementing agencies (high in Egypt, Arab Rep., Peru and Mekong, lower in Liberia and Kenya)
Need to adjust target groups because of limited interest or availability	Degree of government involvement
No monitoring of process	
Delays in implementation	
Impact evaluation	
Need to adjust approach due to implementation problems	Evaluation methodology: randomization with matched pairs (Egypt, Arab Rep., Kenya, Mekong), before/after (Liberia), randomized control trial (Peru)
Delays in collection of end-line data	Only Mekong and Peru measured household effects
No dissemination	

Source: Based on background documents.

and socioeconomic situations across three different regions: low-income and middle-income countries, rural and urban settings, with and without links to agriculture, poor and nonpoor groups, and across different levels of capacity of beneficiary groups. These differences were reflected in the project design. For example, the projects in Kenya, Liberia, Mekong, and Peru focused on increasing the earning capacity of self-employed women. By contrast, the Egypt pilot targeted wage employees in large firms and aimed at reducing gender wage gaps directly (through policies working toward equal pay) and indirectly (by addressing institutional constraints and demonstration effects that limit women's economic opportunities). The country specificity of the design was a plus from the point of view of making the intervention more effective at country level, but reduced the potential for learning global lessons from the RBI program, because the differences in the interventions introduced too many degrees of variation.

Implementation

The interventions had a number of common characteristics. They shared a high degree of government and local NGO involvement, and they had UNIFEM as

the executing agency and relied on local implementation agencies. During implementation, virtually all interventions experienced problems recruiting beneficiaries and/or with attrition. Project teams found themselves with fewer beneficiaries than expected and therefore a smaller sample size for evaluation (see chapter 3 for more details). Furthermore, all programs experienced delays in implementation, reflecting a common, overly optimistic time frame with no allowance for slippage.

However, the approach to solving these problems varied from country to country, which resulted in differences during the implementation stage. In Egypt, recruiting firms proved difficult partly because of skepticism on the potential benefits of gender training. In Kenya, beneficiaries with experience of group-level (cooperation/collective organization) beadwork could not be identified, and so less experienced groups had to be recruited. The Peru project required a second recruitment drive, and in the Mekong Valley, the scale of the project could never reach the desired sample size. The teams in Kenya, Liberia, and Mekong did not provide adequate monitoring information. As a result, the interventions could not be adapted to mitigate difficulties encountered during implementation, which made them less effective.

Other differences, more difficult to assess, relate to the competence and experience of the implementing agency and the extent of government involvement. For example, in the case of Peru, the same implementing agency worked successfully with part of the evaluation team in another intervention. In Egypt and Liberia, the government welcomed the RBI intervention from early stages.

Impact Evaluation

All RBI interventions were designed to have a rigorous impact evaluation, but all had to undergo some methodology adjustments during implementation. In Egypt, due to delays in the recruitment process, treatment for the first 10 recruited firms had begun while control firms were still being recruited.⁶ In Kenya, the program had to draw the control and treatment groups from different geographical areas for ethical reasons as well as because of recruitment problems. In Liberia, working groups were not divided into treatment and control groups, so the final evaluation had to focus on individual-level impact. The initial geographical coverage of the Mekong and Peru impact evaluation had to be redesigned. In all countries, the final sample sizes were too small to provide reliable indications of changes, and the monitoring was insufficient to support the impact evaluation. Finally, the RBI program as a whole did not include a strategy for disseminating knowledge generated by the individual pilots.

These adjustments led to a range of different approaches to the impact evaluations because the roll-out and final methodology varied with country context. Table 1.3 summarizes the method used in each RBI pilot, the level at which the randomization was conducted (where applicable), and the stratification variables used. All RBI pilots, except Liberia, used an IE method with a treatment group (beneficiaries) and a valid control group. Egypt, Kenya, and Mekong used randomization of matched pairs to sort out treatment and control groups, using a

Table 1.3 Impact Evaluation Methods

<i>Country</i>	<i>IE method</i>	<i>Randomization/ matching at the level of</i>	<i>Stratification/matching variables</i>
Egypt, Arab Rep.	Randomization of matched pairs	Firm	Economic sector, size, gender composition of workforce, and gender policies
Kenya	Randomization of matched pairs	Village	Village size, distance to a main road, beadwork, and IGA experience
Liberia	First difference (comparing results before and after)	n.a.	n.a.
Mekong	Randomization of matched pairs	Village	Village population, district location, percentage of households in BH, type of BH products, type of traders
Peru	Randomized control trial	Individual	Districts/neighborhoods

Source: Based on background documents.

Note: IGA = income-generating activity; IE = impact evaluation; BH = bamboo handicraft; n.a. = not applicable.

Table 1.4 Examples of Female Economic Empowerment Questions

<i>Household decision making</i>	<i>Gender roles</i>	<i>Group and community participation</i>
In your household, which family member is responsible for deciding how to spend money?	Did [woman] participate in <ul style="list-style-type: none"> • Childcare • Going to the market • Processing food • Herding cattle • ... activities in the last year?	Why did you originally join the group? <ul style="list-style-type: none"> • Wanted support of other woman • Social • Wanted economic benefits • Wanted to help community • Wanted protection.
Do you and your husband talk about the following with each other (1) often, (2) sometimes, (3) never? <ul style="list-style-type: none"> • Things that happen in at his work/on the farm? • Things that happen at home? • What to spend the money on? • Things that happen in the community? 	Did [man] participate in <ul style="list-style-type: none"> • Childcare • Going to the market • Processing food • Herding cattle • ... activities in the last year?	Are you satisfied with your participation in the group?
Who in the household has the final say on the following: <ul style="list-style-type: none"> • Making large purchases? • Making household purchases for daily needs? • What kind of work you do and where you work? • How much to save for the household? • Whether or not to take a loan for the household and how much? • Children's schooling? • Children's health care? 	If you have a daughter, <ul style="list-style-type: none"> • Which education level would you like her to attain? • At what age should she marry? • How many children should she have? "If the woman works outside the home, it is very likely that there would be family problems." Do you <ul style="list-style-type: none"> • agree • neither agree or disagree, or • disagree with the above statement?	Do you plan to continue being a member of the group?

Source: Based on Kenya, Liberia, Mekong, and Peru questionnaires.

comprehensive set of variables. This type of IE method allows the evaluator to have control over the characteristics of the sample, and when combined with the appropriate weighting, can help minimize the variance of observable characteristics (Bruhn and McKenzie 2009). In Peru, randomization was at the individual level, through a public lottery.

However, difficulties in the implementation stage resulted in higher degree of heterogeneity in the IE methodology across pilots than originally envisaged. The most extreme case was Liberia, where the originally planned randomized control trial approach had to be replaced with other methods. Adjustments also had to be made in Kenya and in the Mekong Valley. The range of indicators and areas of impact also differed across interventions. For example, only the Peru and Mekong questionnaires addressed impacts on bargaining power within the household. Table 1.4 presents a few examples of the questions used.

Notes

1. Later on, three more pilots were included in the RBI program: Ghana, Nicaragua, and Tanzania. In the last two, the end-line data were collected and the analysis will be completed during 2013.
2. More evidence is now available, for example, on programs that provide start-up capital (de Mel, McKenzie, and Woodruff 2008; Fafchamps et al. 2011), those that provide business development training (Karlan and Valdivia 2011), and those that give a combination of the two.
3. The RBI pilots in Egypt, Kenya, Liberia, the Mekong, and Peru were funded by a Development Grant Facility to be executed by UN Women, the ICRW, and The World Bank. UN Women (then called UNIFEM) was responsible for the design and implementation of these RBI, while ICRW was responsible for the impact evaluations. The World Bank had an overall supervisory role.
4. See UN Women (2011) for details on other type of budget entries.
5. The Egypt pilot is somewhat unique because it addressed institutional constraints and demonstration effects that limit women's economic opportunities.
6. The treatment group consists of the intervention beneficiaries (those receiving "the treatment"). The control group is ideally made up of potential beneficiaries that share characteristics with the beneficiaries, except that they do not receive the treatment. Comparison between the groups is used to isolate the effects of the intervention.

What We Have Learned about RBI Intervention Impacts and What We Could Have Learned, But Did Not

The results-based initiatives (RBI) pilots could have had impacts across three different dimensions: (i) human and physical endowments, (ii) economic opportunities, and (iii) agency. Drawn from these impacts are three important questions: first, **what impact** has each pilot had on some, or all, of these dimensions? Second, **how reliable** are these findings, based on statistical significance and robustness of results, as well as on the validity of the impact evaluation design? And third, is there a **consistent message** emerging across countries? Do the RBI show similar messages regarding the impacts, despite differences in the design and implementation of the intervention and in the country context? Findings from the individual impact evaluations are presented in tables 2.1, 2.2, and 2.3.

The observed impacts of the RBI have been mixed (box 2.1). The most positive results were experienced in Peru, where the intervention led to a significant improvement in business earnings. The other pilots had no observable impact on wages, earnings, or business outcomes (revenues and turnover). There was no evidence of increased investment in children's education and health, but some mildly positive impact was observed on agency—via participation in networks and associations.

However, the diverse paths taken by individual pilots in response to structural problems with design and/or implementation cast doubts on the reliability of the results and reduce the ability to draw robust and generalizable conclusions. In addition, the relatively short time frame of the impact evaluation (IE) studies may fail to capture the full impact of the pilots, which were designed to initiate long-term processes of transformation.

In this respect, the RBI represent a costly missed opportunity, because it is impossible to ascertain whether the lack of discernible impact was due to ineffective or inadequate design, or due to incorrect implementation or evaluation, or to lack of impact. The next section focuses on the lessons learned from the RBI that can help guide the next generation of pilots in this area.

Table 2.1 Impacts on Economic Opportunities

	<i>Women's labor force participation</i>	<i>Technology of production</i>	<i>Wages and sales</i>	<i>Career</i>	<i>Statistically reliable results?</i>
Egypt, Arab Rep.	Absenteeism increased for both men and women in the GEME-treated firms. However, this result can be driven by seasonality effects.		No effects on gender wage gap.	No conclusive results at firm level on hiring, training, promotion, or representation of women in management.	YES With caution due to (i) some reliance on self-reported information, and (ii) no robust statistical specification.
Kenya	No new workers attracted to bead handicrafts. Beneficiaries initiated other income-generating activities.	Groups did not embrace suggestion to purchase in bulk to lower input price. No effect on the introduction of new designs.	No overall effect on sales strategies or number or links with traders at the group level. Small effect at individual level. No effect on revenues for the beneficiaries in either bead-only work or in groups that combined beadwork with other activities.		NO Weak design: recruitment and contamination; weak implementation: training curricula.
Liberia	Very few paid jobs created, and these jobs were only available to those living near Ganta.	Class training on farming practices had no effect, but onsite field demonstrations were more effective; plant did not operate at full capacity (maximum production 60 kg/day compared to capacity of 1 metric ton per day, and most of the time produced much less than 60 kg).	Increased access to land and labor and improved techniques led to higher cassava production in group farms, but high variability of impacts across groups; However, cassava was sold only to the processing plant, which could not process it, and many fields remain unharvested.		NO Weak design: no valid control group; weak implementation: underestimated constraints, such as literacy.
Mekong	No change in the probability of households producing bamboo handicraft products.	Increase in quality and new designs; bamboo splitting machine not used; no effect on access to raw bamboo.	No effect on the importance of bamboo as a source of household income; sales during the low season increased, but only in Cambodia.		NO Design: lack of statistical power to control for clusters.
Peru	A few beneficiaries closed their businesses, possibly because they realized business was not viable, but this may also reflect normal high exit rate of small firms.	Women who received both training and targeted assistance introduced some good business practices—bookkeeping and executing innovations—and increased reliance on formal credit. Use of credit—mostly from informal sources—increased for all beneficiaries.	Sales increased by 19 percent among women who received training and targeted assistance.		YES Robust design and implementation and credible impact evaluation.

Source: Based on Golla (2011), Knowles (2011), and Valdivia (2011a, 2011b).

Note: GEME = gender equity model Egypt.

Table 2.2 Impacts on Human Capital Endowments and Resources

	<i>Women's labor market skills</i>	<i>Children's health, nutrition, and education</i>	<i>Assets</i>	<i>Statistically reliable results?</i>
Egypt, Arab Rep.	Women in treated firms received and benefitted from on-the-job training.			YES Based on self-reporting by workers.
Kenya	Training in new designs, but no changes were observed in the beads produced; women valued exposure to trade fairs/ marketplaces.		Reduction in group's savings and material assets.	NO Weaknesses in design and implementation.
Liberia	Mixed reactions to training for agricultural practices; most women not interested in classroom training, but women valued the literacy training.			NO Weaknesses in design and implementation.
Mekong	Training of the trainers succeeded, and 1,354 group participants, mostly women, received the training and perceived it as useful; most of the trainers were also women.	Increase in spending for boys' education, but only in Lao PDR; decrease in spending in personal care items for men; no effect on child labor.		YES With caution, as based on self-reporting by producers.
Peru	Increase in women's time devoted to studies for those who received both training and targeted assistance; decrease for those who received only training.	Decrease in the time other females in the household spend in business and at other work, for beneficiaries of both training and targeted assistance; no effect on time spent on studies by children (aged 7–13 years).	Increase in the use of informal credit (though some crowding out of formal credit).	YES Robust design and implementation, credible impact evaluation.

Source: Based on Golla (2011), Knowles (2011), and Valdivia (2011a, 2011b).

Table 2.3 Impacts on Agency

	<i>Household decision process</i>	<i>Groups and networks</i>	<i>Perceptions</i>	<i>Statistically reliable results?</i>
Egypt, Arab Rep.			Employee satisfaction increased for both men and women. Awareness of gender equality increased for both men and women. Mixed results on workplace discrimination (increased and only among women) and experiences of a hostile work environment (decrease of sexist jokes).	YES With caution: outcomes rely on self-reported information; Weak design: lack of power to control for clusters.
Kenya		No change in the number of groups registered or the average number of group members.	Most young women with some education perceived the project as beneficial. Market visits and trade fairs found useful to learn about marketing strategies.	NO Weak design: recruitment and contamination plus lack of power.
Liberia	Some women reported more confidence and higher self-esteem; families' perceptions of the group's work were mixed due to extensive time commitment of beneficiaries with little economic return.	Membership in group did not change (but information unreliable). Limited ability to manage the plant independently. No sense of ownership/responsibility toward the enterprise. Diagnostic of the organizational problems not enough to encourage the group to implement changes. Leadership training, although valuable for the individuals, did not change group governance or increase transparency.	Overall sense of dissatisfaction with the intervention. Group leaders would have preferred guidance with the management of Ganta and not the introduction of the farina plant. Some satisfaction with training in farming techniques.	NO Based mostly on qualitative data from focus groups.

table continues next page

Table 2.3 Impacts on Agency (continued)

	<i>Household decision process</i>	<i>Groups and networks</i>	<i>Perceptions</i>	<i>Statistically reliable results?</i>
Mekong	<p>Significant positive effects on likelihood of bulk purchases (Cambodia only).</p> <p>No significant effects on gender roles in BHP-related decisions.</p> <p>No effect on gender roles within the HH, but decrease in spending in personal care items for men.</p>	<p>Increased number of households working with a producer group.</p>	<p>Perceived constraints are:</p> <ul style="list-style-type: none"> • Access to inputs (too expensive) • Access to traders • Low product prices. 	<p>NO</p> <p>Weak design: lack of statistical power to control for clusters.</p>
Peru	<p>Increased women's participation in HH decision process for women who received both GT+TA.</p>	<p>Increased participation in business associations for all beneficiaries.</p>	<p>Increased acceptance of women working.</p> <p>No effect on perception of gender roles within the household.</p>	<p>YES</p> <p>Robust design and implementation and credible impact evaluation.</p>

Source: Based on Golla (2011), Knowles (2011), and Valdivia (2011a, 2011b).

Note: HH = household; BHP = bamboo handicraft product; GT = general training, GT+TA = general training and technical assistance.

Box 2.1 Key Findings on the Impact of the Interventions

Overall, the RBI provide few statistically significant results in terms of overall impact on economic opportunities and on the relative effectiveness of alternative program designs. Further, there was no consistent evidence of positive impact on children's education and health, or on access to assets and women's decision making within the household, or the business did not appear to have improved significantly.

Some interesting findings nonetheless emerge:

- The interventions appear to have had positive impacts on self-perceptions of labor market skills, which may result in higher income over time, or in stronger agency.
 - Training increased economic opportunities in Peru, but there was little impact on the wage gap or on earnings elsewhere.
 - Peru and Mekong show encouraging evidence of participants adopting more efficient business practices.
 - Women's participation in associations increased in the Mekong Valley and in Peru, possibly as a result of increased collective voice in society as a result of the intervention.
-

Economic Opportunities

The RBI addressed important barriers to gender equality in economic opportunity in each country. For example, in the Arab Republic of Egypt, the intervention targeted institutional constraints in the formal labor market to overcome gender discrimination in wage employment. The other pilots focused on promoting parallel economic activities to help women diversify sources of income and balance household chores with income-generating work (Kenya and Mekong) or increase their return as farmers (Liberia) and entrepreneurs (Peru).

Overall, the RBI did not appear to have had a statistically significant impact on main indicators of labor market outcomes. In Egypt, there was no increase in the hiring of women, either in treatment or control firms. In Kenya, Liberia, Mekong, and Peru, the number of women engaged in any of the activities promoted did not increase, and in Liberia, the cassava transformation plant did not generate additional sustainable jobs. Table 2.1 also shows that the impacts on economic returns to labor were mostly weak, although positive when significant, as in Peru. In Egypt, women's wages did not increase and the gender gap in both wages and access to training remained constant in both treatment and control firms. Self-employed women in Kenya, Liberia, and Mekong did not see an increase in revenues as a result of any of the RBI interventions. These results are consistent with the findings of other impact evaluations of business development training and can be partly due to the difficulties in adequately measuring sales, costs, and revenues (McKenzie and Woodruff 2012).

However, some encouraging findings emerged. In Peru, business practices improved—for example, bookkeeping was introduced—and sales increased by

nearly 20 percent for the microentrepreneurs who received combined general and tailored technical assistance in business development. In the Mekong pilot, the quality of products increased and new designs were introduced. However, returns remained unchanged; the higher quality did not translate into higher prices in the local market. In Egypt, job satisfaction among women workers appears to have increased and male workers displayed stronger awareness of existing gender inequalities and of the importance of supervisors' fairness and more gender equality in career development, training, and job opportunities.

Human Capital Endowments and Assets

In the Mekong, Peru, and, to a lesser extent, Kenya and Liberia pilots, the impacts of the RBI on endowments were assessed along three dimensions: skills of the participants, investment in education and health of their children, and assets.

All RBI had a training component, and across all pilots this led to improved self-perception among individual women of their own labor market skills. Although the magnitude of the impacts and their statistical significance vary from case to case, women consistently reported that they benefited from training. In Egypt, women in beneficiary firms reported benefitting from increased on-the-job training. In Liberia, some, but not all, women who participated in focus groups pointed to literacy training and training on farming practices as the most (or only) useful part of the intervention (table 2.2). However, not all training translated into practice. In Liberia, for example, the suggestion by the trainers to make purchases in bulk to take advantage of better prices was never adopted. In the Mekong Valley, the splitting machines acquired by the project were barely used because they produced bamboo strips that were too large for small handicrafts.

For other potential impacts, there is little evidence in general, and observed impacts are not necessarily consistent with predictions. Impacts on investment in children were only measured in the Mekong and Peru pilots. There was a statistically significant positive impact on boys' education in the Lao People's Democratic Republic—not for girls, as would have been expected—but no impact was observed in Cambodia. However, some expenditure and time-use reallocations took place within the household. In Mekong, expenditures on male care items decreased, and in Peru, other female members of the household spent less time in the female family business.

The impact on asset ownership was tested only in Kenya, where the beadwork groups experienced a reduction in savings and in the stocks of production assets (and not an increase, as might have been expected). However, the impact was not statistically significant and, furthermore, the results are questionable because savings and assets might have been used to initiate other income-generating activities, something which was not captured by the evaluation. In Peru, there was an increase in the use of informal credit, but this might have crowded out part of the formal credit of microentrepreneurs.

Agency

While agency can be measured over different outcomes (for example, societal voice and household decision making), the RBI impact evaluations, by monitoring how decisions are made within the household, focused almost entirely on control over resources.¹ Participants in Cambodia and Peru reported participating more actively in decisions over large household expenses, but this effect is not reported in Lao PDR. In the Mekong Valley, both women and men identified themselves—inconsistently—as the main household decision maker, both before and after the intervention.

In the Mekong Valley and Peru, programs increased women's participation in groups and networks, and they perceived that their socioeconomic environment improved when new, potential income-generating opportunities were introduced. In the Mekong Valley, women moved from home-based to group production of bamboo handicrafts and reported that the group approach allowed them to learn from each other. In Peru, more women joined business associations and informal lending networks. However, such positive impacts were not observed in Kenya and Liberia.

Note

1. Responses in this area are almost invariably subjective, and, in fact, men and women usually give different answers to the same question, which means that it is difficult to find valid results (Bardasi et al. 2010).

Lessons Learned on Pilot Design, Implementation, and Evaluation

A key purpose of the results-based initiatives (RBI) was to provide new and reliable information on the effectiveness of interventions to promote women's economic empowerment in different country environments. However, design weaknesses and implementation problems seriously undermined the validity of the results.¹

Three main messages emerge from this section's summary of the RBI' lessons learned (box 3.1): **first**, for a pilot to be able to generate meaningful conclusions, it is essential to align resources with expectations. Realistic expectations from the outset on potential impact, costs, and time required are essential to the success of any pilot. But it is also important that expectations are aligned with the available resources and that the time frame of the pilot is long enough for full implementation and for impacts to be observable. Meaningful, rigorous impact evaluations require time and resources to bear fruit. **Second**, even small deviations from methodological recommendations in the design and implementation of the impact evaluation component may invalidate the findings and nullify their value. Combining quantitative data with qualitative information considerably strengthens the potential for assessing impacts and for understanding the transmission channels through which they materialized. Some conclusions can then be drawn even when more rigorous quantitative results fail, and results are more credible and more informative for policy makers. **Finally**, close and continuous monitoring during project implementation is essential to detect shortcomings in design and implementation strategy at an early stage. Moreover, monitoring needs an effective mechanism to feed findings back into the design and implementation of the project.

Risks at the Design Stage

Pilots should be small, strategically relevant interventions that are innovative in delivery or content, replicable, and influential (Gertler et al. 2011). However, a number of challenges may arise from the tensions between these characteristics.

Box 3.1 Key Requirements for Successful Pilot Implementation

A number of key factors need to be in place for a pilot to successfully meet its objectives:

- The overall objectives of the intervention need to be realistic and be aligned with available time and resources.
 - Interventions need time to show impacts; measuring end-line results prematurely can be very misleading.
 - Rigorous quantitative impact evaluations without adequate time and resources for planning, implementation, and evaluation are pointless. Alternative methods, such as a good qualitative study, may then provide a more effective option.
 - Quantitative data collection needs to be supported by qualitative information, and a wide range of process and/or outcome indicators should be used to obtain a complete picture of results and to hedge (with qualitative data) against potential data collection and other measurement failures.
 - An adequate monitoring system is essential to ensure (i) effective implementation, by providing early warnings when it is necessary to adapt the program to unforeseen circumstances, and (ii) effective evaluation, by providing critical information on intermediate changes and processes.
-

Such strains are especially likely in the context of interventions for female economic empowerment, which are complex due to interacting institutions and cultural dynamics and constraints. The RBI experiences highlight the need for pilots to address these tradeoffs explicitly from the outset to ensure success.

Budgets Need to Sustain Monitoring, Evaluation, and Dissemination Costs

The budget envelope for each RBI, except for the Mekong Valley, was approximately US\$300,000.² This proved too small for complex programs that spanned a range of countries and contexts and involved building value chains, training in a low-capacity environment, and full-scale impact evaluations. Lack of funding reduced the capacity to derive reliable and generalizable conclusions on the effectiveness of the interventions by imposing trade-offs between focusing on effective implementation and meeting the demands for rigorous impact evaluations. The failure to earmark resources for monitoring and dissemination activities added to the challenge; these are potentially expensive but critical elements of any successful pilot.

The first lesson is that although pilots may be designed to be small in size, the cost per beneficiary is likely to be higher than similar small program interventions that do not incorporate an evaluation component. A pilot is also likely to use a relatively larger share of budgets for overhead, administration, and monitoring and evaluation than small programs that are less innovative and do not have a learning component. Beyond what is spent on beneficiaries, large and mostly

Table 3.1 Number of Beneficiaries

<i>Number of beneficiaries</i>	
Egypt, Arab Rep.	10 firms, number of employees ranging from 15 to 148 per firm
Kenya	23 producer groups, number of members ranging from 7 to 70 per group
Liberia	246 women of Ganta's producer group
Mekong	27 villages (or bamboo handicraft groups), 986 BHP households
Peru	1,418 female microentrepreneurs

Source: Based on background documents.

Note: The number of beneficiaries differs from the number of observations used in the impact evaluations because some firms, villages, and households had to be discarded for the analysis. BHP = bamboo handicraft product.

fixed costs are added because of (i) due diligence and preparation, (ii) detailed monitoring and evaluation, (iii) impact evaluation, and (iv) dissemination.

Small Size Needs to be Consistent with Overall Targeting Objectives

The RBI pilots had a relatively small number of direct beneficiaries, ranging from about 250 women in Liberia to almost 1,500 in Peru (Table 3.1) and generally relied on geographical or socioeconomic targeting to identify beneficiaries. The advantages of small groups are lower costs and easier implementation. Targeting can further enhance these benefits; geographical targeting also provides a control for external shocks because constraints and opportunities are expected to be similar across beneficiary groups, which facilitates design and adaptation.

However, several of the RBI pilots proved too small to provide statistically reliable results. Geographical targeting in Kenya and Liberia also led to a high degree of contamination³ because there was no clear distinction between treatment and control groups. The distance between villages may not have been large enough to prevent contamination from one treatment village to a control village.

The second lesson is that, due to the multistage character of pilots, beneficiaries have to be chosen so that the conclusions from the intervention are valid/causal for those eligible to receive the treatment (internal validity), and the choice of beneficiaries also needs to be consistent with some generalizable lessons that can be applicable to a larger population (external validity). The size and composition of the treatment group needs to be consistent with the objective of replicability and/or scale-up.

Time Frames Need to Allow for Planning, Implementation, and Impacts

The RBI were expected to show quick results—generally within one to two years of project initiation. A short time frame for planning and design, implementation, and evaluation is a typical feature of pilots, especially when they are conceived as a first phase of a larger program. However, an accelerated time frame can condense the initial stages of pilot development, including the client consultation stage. It can also shorten the time dedicated to understanding the theory behind the intervention: its expected impacts, the mechanisms through which the impacts will be achieved, and its potential distributional effects. Short time frames can further reduce the scope of the exploratory work, including necessary

Table 3.2 Project Delays

	<i>Planned months</i>	<i>Actual months</i>	<i>Gap</i>	<i>Reason for delays</i>
Egypt, Arab Rep.	26	34	8	Slow recruitment of firms
Kenya	30	42	12	Change in approach and in target group, subsequent lack of suitable beneficiary groups, civil unrest, drought
Liberia	34	46	12	Acquiring cassava production and land, identifying local consultants for different technical assistance activities
Mekong	41	45	4	Administrative and technical (bamboo-splitting machines) delays, IE design
Peru	24	24	0	Slow recruitment and assignment into treatment and control groups, high drop-out rates

Source: Based on background documents.

Note: IE = impact evaluation.

simulations to gauge the characteristics of target population, the eligibility criteria that will minimize the number of noncompliers, and the potential impacts.⁴ Critically, complex interventions need time to have full effect. Premature evaluation because of time pressures may thus miss the potential impacts of programs, or even reverse the conclusions if the trajectory of impact is nonlinear.

Each RBI pilot was based on extensive feasibility studies, which served as the basis for project selection and design and facilitated implementation. In Peru and the Mekong Valley, the questionnaires were also pretested. Nevertheless, in the Mekong Valley, Kenya, and Liberia, the population of potential beneficiaries turned out much to be smaller than suggested by the feasibility studies, and all initiatives incurred difficulties in identifying and recruiting a meaningful sample of beneficiaries. Because implementation was delayed, four pilots ran for longer time than initially expected (table 3.2), and only in Peru was the response to the unexpected recruitment problems swift enough for the project to remain on track. Generally, the initial delays resulted in a very short time period between the end of the intervention and end-line data collection.

Overall, the RBI objectives may have been too broad to be transferable into quick and effective interventions. Increasing women's earnings and their bargaining power within households are long-term objectives that call for an entire transformation to effect change. The capacity-building efforts required, especially when targeting low-income groups with low levels of education, are resource and time consuming. This became obvious in Liberia, Kenya, and the Mekong Valley, where organizational transformation, behavioral changes, and skills acquisition were all needed for the projects to have a significant impact. By contrast, in Peru, the implementation phase was conducted under considerable time constraints due to delays in the design and planning stage. This resulted in a good impact

evaluation, but possibly at the expense of too short a time between the implementation and the collection of end-line data, reducing the possibilities of measuring final and full impacts of the intervention.

The **third lesson** is that pressures to produce quick results must be weighed against the benefits of allowing time for good planning, careful implementation and adaptation to produce results, and for a solid impact evaluation to be carried out. Even with a small number of beneficiaries, if the objectives are ambitious and transformational (of behaviors, capacity, or management), the implementation process can be lengthy for pilots to be successful. When prior evidence is scarce, it is important to conduct focus group discussions, pretest the questionnaire, or pretest part of the intervention. If the intervention requires behavioral changes, then training, institutional support activities, and a demanding process of learning by doing for both trainers and trainees are required. These are time consuming activities, but help to deliver more effective interventions and derive robust evaluations of effectiveness. In some cases, existing work can feed into the process. For example, in Peru, an experiment focusing on RBI-related training was launched just before the program, and served to inform and improve the program.⁵

Issues Surrounding Impact Evaluation Methodology

Box 3.2 summarizes 10 basic guidelines for designing and conducting impact evaluation (IE) pilots that can effectively evaluate if, and how, specific interventions work, and show why. Table 3.3 maps the main characteristics of the RBI

Box 3.2 Guidelines for Designing IE Pilots

1. A pilot needs to be **justifiable** and **ethical** in terms of learning potential: (i) the intervention had to affect a larger number of people; or (ii) evidence of its potential impacts is lacking.
2. The pilot and the expectations of what can be achieved through the impact evaluation need to be guided by an underlying **theory of change**: that is, the hypotheses to be tested and their transmission mechanisms. It is good practice for the theory of change to include (i) theoretical models, (ii) logic models, and (iii) result chains.
3. The IE **questions** that respond to the hypotheses developed by the theory of change need to be clearly spelled out.
4. The choice of the IE method should adjust to the intervention and have a **valid control group**.
5. **Performance indicators** need to be specific, measureable, attributable, realistic, and targeted.
6. It is good practice to **monitor** the implementation of the intervention by collecting information along the way, using mid-term surveys, focus groups, and other tools.

box continues next page

Box 3.2 Guidelines for Designing IE Pilots (*continued*)

7. Clear and transparent identification of **eligibility criteria** for the target population and **targeting** rules for the beneficiaries are essential to avoid **contamination** between treatment and control groups.
8. A **minimum and representative sample** size is necessary. The minimum sample size for statistical robustness can be determined through power calculations (that take into account mean and variance of performance indicators) and should consider possible attrition during intervention or as part of the data collection process in that country context.
9. **Data** should be collected at the level at which the intervention changes behavior; mixed methods that combine quantitative and qualitative data are recommended. Mixed methods can provide a better understating of the transmission channels through which impacts operate and can help mitigate the risk that poor quantitative data collection invalidates the results of the evaluation.
10. **Sufficient time** should pass between the intervention and the collection of the end-line data to measure changes in the performance indicators selected. Changes in outcomes related to changes in behavior and social norms need time to occur and premature evaluation can yield misleading conclusions.

Source: Based on Gertler et al. (2011), Khandker, Koolwal, and Samad (2009), Imbens and Wooldridge (2009), Todd (2012), Duflo, Glennester and Kremer (2008), Ravallion (2009), and Bamberger, Rugh, and Mabri (2006).

impact evaluations to these good practice principles. It is clear that the RBI pilots complied with some but not all of the good practice principles in box 3.2. All pilots—except Liberia—had valid control groups that reflected the choice of the IE method, although there were some problems with selection and randomization. Rich quantitative data were collected at the level of the intervention and beyond to capture changes in group dynamics and even in social norms. The Mekong pilot and, to a lesser extent those in Peru, Kenya, and Liberia, also collected rich qualitative data, and innovative performance indicators were used to profile household decision making, group dynamics and agency, areas where quantitative indicators and data collection expertise are scarce.

However, methodological weaknesses in the IE design severely limited the validity of the results. In Liberia, due to the lack of a valid control group, the impacts had to be measured by comparing the pre- and postintervention value of the indicators among female farmers belonging to the treatment group. The intervention was delivered only to the Ganta Concern Women's Group, an organization comprising 11 community-level women's groups located in Nimba County. To control for time effects, the evaluators also measured changes over time for the same indicators among a nontreatment group. However, to avoid conflict among Ganta members, they decided to select the nontreatment group among nonmembers. This choice invalidated the control group since participation in Ganta in itself indicates different attitudes toward income-generating activities.

Table 3.3 Budget and Time Constraints versus Quality of Impact Evaluation

	<i>Justifiable</i>	<i>Theory of change</i>	<i>IE questions</i>	<i>Performance indicators (examples)</i>
Egypt, Arab Rep.	Can affect a large number of workers if scaled up. Similar interventions implemented in other countries.	Gender equitable practices in the work place increase productivity by giving women more job satisfaction and by providing a better match between tasks and skills.	Gender equitable practices reduce gender discrimination in recruitment, provision of on-the-job training, career development, and prevention of sexual harassment. The prospect of gender equity certification provides incentives to implement gender equitable practices in the workplace.	<i>Firm level:</i> Proportion of women employees, proportion of women in newly hired (six months), proportion of women in management. <i>Employee level:</i> Employee satisfaction measures, perception of (i) openness of work environment, (ii) willingness to promote, and (iii) fairness of supervisor, compensation (benefits + salary), and training opportunities.
Kenya	Knowledge gap on effect of training on women's business skills	Training and mentoring contribute to developing women's business skills and thus increase productivity. Women in groups have more access to inputs' and outputs' markets since there are economies of scale in markets.	Training and mentoring increase women's business skills. Business skills increase productivity. Group participation increases market awareness. More control over income increases women's bargaining power within the household.	<i>Group level:</i> Business practices of the group (sales, trading, and so forth), revenues. <i>Individual level:</i> Income-generating activities of women, group participation.
Liberia	Little evidence of effect of training on female farmers' skills	Training and mentoring help develop business skills and thus higher productivity. Access to startup capital increases productivity.	Training increase women's farming skills. Access to capital increases women's productivity.	
Mekong	Knowledge gap on effect of training on business skills of women.	Training and mentoring help develop business skills and thus increase productivity. Increased productivity in bamboo production results in more diversified household income.	Business skills increase productivity. Larger control over income increases women's bargaining power within the household. Women's strengthened control over income increases (gender neutral) investment in children's education and health.	Participation in bamboo production Sales Children's education attainment Food consumption Decision making by household members.
Peru	Knowledge gap on effect of training on women's business skills.	Training and mentoring help develop women's business skills and thus increase productivity.	Business skills increase productivity; more control over income increases women's bargaining power within the household.	Business practices Sales Time spent in enterprise Access to credit.

table continues next page

Table 3.3 Budget and Time Constraints versus Quality of Impact Evaluation (continued)

	<i>IE method</i>	<i>Targeting</i>	<i>Sample size</i>	<i>Time frame</i>	<i>Data collected</i>
Egypt, Arab Rep.	Randomization of matched pairs of firms. Matching done on a few characteristics and with very few firms → control group could differ in many unobservables.	<i>Clear definition of eligibility rules:</i> Medium and large exporting firms in greater Cairo <i>Recruitment problems:</i> Very few firms replied to initial call → concerns about selection and external validity.	No power calculations and very small sample size, particularly for having cluster corrections.	Short time to observe change in behavior by managers and cultural attitudes by workers.	Quantitative worker survey Quantitative firm survey.
Kenya	Randomization of matched pairs of villages.		No power calculations.		
Liberia	Before and after.	A specific group of producers.	No power calculations.		
Mekong	Randomization of matched pairs of villages.	Three provinces identified through feasibility study. Villages with producer groups and bamboo traders, but not all control villages had producer groups.	Below the recommended size by power calculations for cluster corrections by village or producer groups.		Household quantitative data. FGDs with producer groups. IDIs with traders and village leaders (or key informants).
Peru	Randomized control trial of female microentrepreneurs.		Within the recommended sample size by power calculations.	BDS and TA span over three-month period each, and about six months after end-line data were collected	Interviews with female micro-entrepreneurs. FGD with beneficiaries. IDIs with municipal officials.

Note: BDS = business development skills; FGD = focus group discussion; IDI = in-depth interviews; TA = technical assistance; IE = impact evaluation. Appendix C contains the power calculations performed for the Mekong and Peru RBI.

Similarly, in Kenya, ideal randomization would have been among groups producing beads in each village and across villages. Because of ethical concerns, however, the groups were sorted into 13 geographic clusters distant enough from each other to minimize the possibility of conflict arising from some clusters participating while others did not. The teams then faced the considerable challenge of balancing clusters to produce similar enough treatment and control sets in terms of geographical location (urban/rural), distance to a main road, work experience (beadwork and income-generating activity experience), and membership size.

The recruitment problems and the short time span between the end of the intervention and the collection of the end-line data were major limitations. Another critical flaw was the lack of power observed in four of the five RBI pilots.

Selection of Performance Indicators for Agency Particularly Complex

Selecting appropriate performance indicators is a challenging but critical step for evaluation success. The choice was particularly challenging for the RBI, given the little evidence available during the design stage on barriers and engines to female economic empowerment and on effective ways of measuring impacts in this area. The ambitious attempt to quantify impact on a number of dimensions of access to economic opportunities and agency and to experiment with innovative performance indicators further added to the challenge (table 3.4). Many of the innovations in measuring women's agency relied on questions that are innately subjective and in some cases showed inconsistent answers over time or between men and women, while others proved difficult to measure.⁶

The fourth lesson learned is that it is essential to select indicators early in pilot design because this choice influences the measurement not only of the impact but also of the multiple channels through which impacts are occurring.

Lessons Not Generalizable without Consistency in Design and Methodology

The experience with the RBI as a set of interventions highlighted the challenge of balancing trade-offs between adjusting to country contexts and ensuring that intervention impact is robust in different circumstances on the one hand (which requires some variation), and obtaining generalizable lessons on the other. As shown earlier in table 1.2, the pilots varied considerably in design, targeted population, implementation, and methodology of the impact evaluation. In retrospect, the number of factors that changed was so high that no conclusions

Table 3.4 Female Economic Empowerment Indicators

<i>Economic opportunities</i>	<i>Agency</i>
<ul style="list-style-type: none"> • Access to employment, livelihoods • Access to credit • Gender wage gaps in firms • Female share of managers 	<ul style="list-style-type: none"> • Control over resources within the household • Contribution to family support • Ownership of assets • Time use and division of labor within the household

Source: Adapted from ICRW (2007).

could be generalized across pilots. If this was an objective of the program, more weight should have been given to having consistency in design, implementation, and methodology.

The **fifth lesson learned** is that differences in implementation and IE design should be minimized to be able to generalize the conclusions of a program based on several pilots. Sufficient differences will arise from tailoring the implementation to the country context, variations in the capacity of implementers, and unforeseen problems that need different solutions. Therefore, interventions should start off with a consistent approach and IE methods; using different methodologies introduces unnecessary uncertainty on the measured impact.

Recruitment of Beneficiaries and Sampling Issues

All RBI interventions faced difficulties in recruitment, except Liberia, where the group was chosen from the outset and where there was no control group. There were problems in both recruiting enough firms (the Arab Republic of Egypt, Kenya, and Peru) and finding enough villages (Mekong). In Peru, few firms responded to the first round of recruitment in the North Cone of Lima. Qualitative surveys, as well as the information coming from other sources, such as Organismo de Formalización de la Propiedad Intelectual (COFOPRI),⁷ showed that the implementing agency did not reach the target population. Failure to deliver timely and appropriate intervention information to the target population is a common reason for poor recruitment in pilots. The implementing team in Peru decided to expand geographically instead of improving the information delivery, and expanded the intervention to the South Cone. This solved the poor recruitment issue, but introduced an unplanned and unnecessary element of variation.

Recruitment problems can indicate a design failure—as in Kenya’s pilot. Initially, the intervention was expected to target group beadwork enterprises with at least two years of beadwork experience. However, this population was much smaller than foreseen, and a few months into the program, the local implementing agency had identified only 11 groups. While in the field, the evaluation teams worked with the agency to quickly identify new groups, but in the end, only 42 of the 70 groups in the evaluation frame had any previous group-level beadwork experience. Of the remaining 38 groups, few could be classified as having a group beadwork enterprise of the kind envisioned (in which the group collectively made decisions about production and sales). However, nearly all members made beadwork products, and 80 percent of the women interviewed had sold beadwork in the last 12 months.

Addressing recruitment issues early is fundamental, because failures in recruitment mean small sample sizes that threaten the validity of the impact evaluation.⁸ Once the sample size is below the minimum threshold, the distance to the threshold becomes irrelevant and results can no longer be measured. This problem worsens if the impact evaluation demands cluster corrections. Table 3.5 shows the sample size for each RBI pilot. In Egypt, approximately 43 percent of respondents interviewed at baseline could not be interviewed at end line. The

Table 3.5 Observations in Treatment and Control Groups

	<i>Treatment</i>	<i>Control</i>	<i>Randomization by</i>
Egypt, Arab Rep.	8 447	8 519	firms workers
Kenya	23 175	36 294	producer groups women
Liberia	246	239	women
Mekong	14 612	15 496	villages households
Peru	1,418	565	female microentrepreneurs

Source: Based on background documents.

Note: For Egypt, Arab Rep., the number of workers interviewed at the baseline and end line (both male and female); many workers were present only at the baseline or only at the end line (Golla and Selim 2011, table III.2). Only 16 of the 19 recruited firms were included in the impact evaluation, 1 treatment firm was left out since it could not be matched, and 2 other matched firms (one in treatment and one in control) were also left out because the treated firm refused to participate in the end line.

most cited reasons were that the employee no longer worked at the firm (56 percent and 67 percent for treatment and control nonrespondents, respectively) and that employees were on leave (26 percent and 8 percent, respectively). Where employees were not available for the follow-up interview, team supervisors interviewed substitutes of the same gender and department. The Mekong RBI demanded cluster corrections among individuals within a group/village because this was the level at which the intervention was delivered.⁹

The RBI pilots' design did not allow for potential failures (or simply delays) in the implementation stage. However, during implementation, it became clear that it takes time for potential beneficiaries to learn about the intervention and that the target population may have difficulty in physically accessing the intervention. Limited resources also lead to undersampling. With a more generous budget, the evaluation team could have worked from the outset with a larger target population, which likely would have led to a larger target sample despite the recruitment problems faced once in the field. Similarly, a more flexible time frame would have allowed the team more time to inform potential beneficiaries.

The sixth lesson learned is that sufficient time and resources need to be dedicated to the recruitment stage to prevent problems later. If there is high risk that the quantitative impact evaluation will lack power, it is important to develop an alternative plan that builds on the feedback of regular monitoring. This can include good qualitative studies, and/or designing the intervention in a way that allows subsequent roll-out to other regions—and in this way expanding the sample—if additional funding is leveraged. Another way forward may be collaboration to test several interventions with one sample (crossover—see appendix B on technical concepts), since a certain sample may be underpowered for one intervention but enough for other.¹⁰

Time Frame of Impact Evaluation Can Drive Results

The optimal time frame between the end of the intervention and the collection of the end-line data depends on the theory of change—that is, on the type of changes

expected from pilots—and on the type of performance indicators chosen. Thus, there is not one simple rule on the optimal time frame for impact evaluations.

Most of the RBI pilots collected the end-line data within six months of the end of the intervention. However, this time frame was too short to be able to observe the full impacts of the interventions, especially given their ambitious objective and the performance indicators selected. In Peru, for example, the evaluation could not pick up changes in sales by female entrepreneurs, and to remedy this problem, the evaluation team had to collect a second round of data.

Seasonality may also affect the measurement of impacts. For example, in the Mekong Valley, the team chose to collect the end-line data soon after the intervention was completed to avoid entering into the rice farming season, a time of the year when bamboo production comes to a halt.

The **seventh lesson learned** is that it is important to choose indicators that change at different times and to plan for several rounds of end-line data collection. Indicators of final impacts should be complemented with immediate and intermediate impact indicators—such as bookkeeping or enrollment in business or financial associations—that can be monitored within shorter time frames and tend to be easier to measure. These intermediate indicators are also useful to gain information on the transmission channels of the impacts and on emerging barriers to progress. They help avoid a “black box” evaluation and provide useful information even if final impacts do not materialize.

Data Collection and Indicators

The trade-offs between quality and budget and between quality and time frame also apply to data collection. Good quality data often require a large sample to minimize measurement error. Collecting quality data can also involve long questionnaires that demand more expertise from local agencies and enumerators, and include variables that are more difficult to measure. Good-quality data also require time, because the questionnaires need to be pretested and some variables need more time to be collected or to show changes. For the Mekong RBI, the qualitative data were well integrated with the quantitative data and used to build up the story. In all cases, the qualitative data were intended to complement the quantitative data. The **eighth lesson learned** is that good qualitative data can be collected with a limited budget and that using mixed methods provides a broader understanding of the processes.

The RBI experiences also show that it pays to focus on indicators for which there is more room for improvement. For example, in the initial (baseline) data collection for the Liberia RBI, over 90 percent of women reported making decisions about what to plant on their own or jointly with their husbands. Although the project was explicitly aimed at increasing female decision making, there was little room for the share of women reporting decision making to increase. The **ninth lesson learned** is that measurable changes can only be expected in areas where there is scope for improvement. Qualitative data collected prior to the baseline and thorough testing of the survey instruments can help evaluators understand what the baseline levels of key indicators are likely to be.

Table 3.6 Intervention and Data Collection Levels

<i>Impact measured at level of</i>	<i>Individual</i>	<i>Household</i>	<i>Producer groups</i>	<i>Firm</i>	<i>Village</i>
Egypt, Arab Rep.	Data			Intervention data	
Kenya	Data		Intervention data		
Liberia	Data		Intervention		
Mekong	Data		Intervention data		Data
Peru	Intervention data	Data			

Source: Based on background documents.

The RBI pilots collected data at several levels and in most cases measured the change at the level it occurred (Table 3.6). Impacts from the Egypt pilot, for example, were measured at the firm level and data were collected using a questionnaire administered to the human resources department of each firm and an employee questionnaire for each worker. However, in Kenya and Liberia, due to problems during the implementation stage, the evaluations were limited to measuring impact at the individual level, although the interventions were implemented at a group level. An important planning question was thus whether the groups would create true “group enterprises” or simply provide support for individual entrepreneurs. In other words, would direct change take place mainly at the group or the individual level? After much discussion, the evaluation in both countries focused on group-level outcomes. However, many of the groups in Kenya did not have much experience running a beadwork enterprise, and many were not interested in doing so. The project had to be redesigned to focus more on individual training. If more had been known upfront about the participant groups, the intervention and the evaluation could have been better designed. In Liberia, the evaluation focused on the impact of a group enterprise on returns to the individual. How group profits would be shared with members was not clear in the planning stage—developing such a system was a goal of the project. In practice, the intended organizational development efforts started only at the end of the project, and the group never made payouts to individuals. As a result, the end-line survey had to be dropped from the evaluation plans because it was clear there had been no individual impact for the key indicators.

The tenth lesson learned is that it is essential to collect data at the level at which changes in the indicators occur, even though this might increase the minimum sample size required. If the project works primarily to create change at an enterprise or group level, enterprise or group impacts should be measured and data collected at that level. It may be possible to reconstruct the group-level measure from individual or household impacts, but this is a more indirect measure.

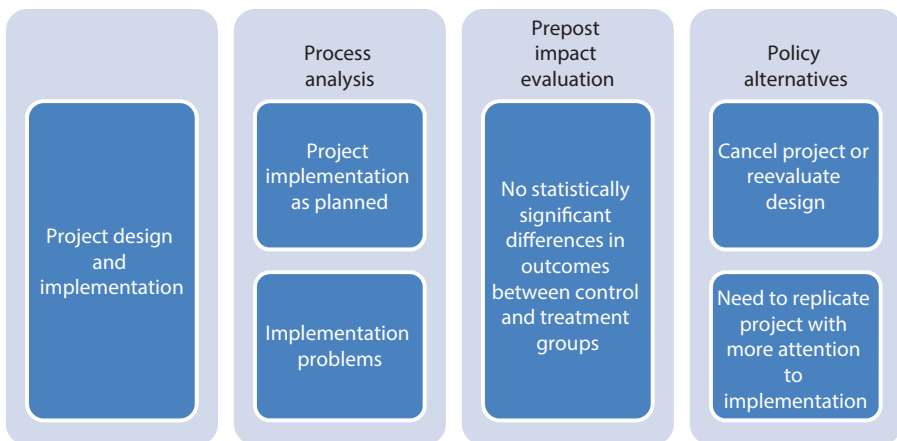
Finally, all the RBI pilots, with the exception of Egypt, include questions related to the decision-making process within the household. The RBI included indicators to show that unintended undesirable impacts have not occurred. For

example, the Peru RBI evaluation included indicators of women’s decision making in their businesses to be sure that women maintained control of their businesses as they expanded. The **11th lesson learned** is that it is possible and important to include indicators that measure perceptions related to agency because they are an effective way of measuring an intention of change in the behavior of the affected population, and may be easier to capture under time constraints.

The Importance of Monitoring

Rigorous impact evaluations are essential to assess *whether* a program had the intended effects. However, they do not explain *why* interventions succeed or fail. Regular monitoring and reporting thus serve two main purposes in program execution and evaluation. First, they allow for early detection of problems and adaptation to improve performance. This may be more easily done in pilot-type programs, which are small in size and therefore potentially more flexible than larger programs. Second, monitoring facilitates understanding of why something had an effect or not. A project may have failed because it was not implemented as planned, but might otherwise have been effective. If an evaluation finds no statistically significant differences between the change in indicators of intended outcomes between the project and comparison group, and there are no monitoring data, conventional evaluation designs cannot distinguish between two alternative explanations: there are weaknesses in the project logic and design in this context and the project should not be replicated, or at least not until it has been redesigned (alternative policy recommendation A in figure 3.1), or significant problems or deviations during project implementation prevented testing the logic of the project design—then another pilot project could be funded with more focus on implementation (alternative policy recommendation B).

Figure 3.1 Alternative Policy Recommendations When Control Group Comparisons Find No Statistical Significant Impacts



Source: Bamberger, Rao, and Woolcock 2010.

For the RBI there was no clear assignment of monitoring responsibility, neither at the partnership level nor at the more decentralized project implementation level. While project monitoring was part of the technical design, there was no ex ante agreement on the different information needs, useful monitoring indicators, how to integrate monitoring into impact evaluations, or on measures to use monitoring for project management underway. Moreover, no budget line was assigned to monitoring, and teams were not specifically trained in monitoring techniques and did not understand the need or use for the data. As a result, project monitoring fell significantly short of what was expected from the technical documents, and what was needed to help adapt programs already underway and provide a fuller understanding of why programs did or did not work. These shortcomings significantly impaired the impact evaluations.

The 12th lesson learned is that monitoring processes and performance along the way is an essential component of any form of program evaluation. Thus, pilot programs—like other programs—must have a clear monitoring plan that identifies indicators to be collected, who will collect the data, how it will be recorded and reported, and when and who will be supervising and analyzing the data. Budgets need to reflect this and should have a specific line for monitoring costs. Monitoring of training programs, for example, could include the number of sessions delivered, the number of women attending, the quality of training provided, who attended (target group or others), and why eligible (target) groups did not attend.

Notes

1. Technical concepts relating to impact evaluation are explained in appendix B.
2. Additional in-kind contributions (for example, staff time) from ICRW, UNIFEM, and the World Bank were also made and are not reflected in this budget envelope. The Mekong intervention involved not only multiple activities but also a splitting machine for each bamboo producer group. At the end of the intervention, each producer group had the option of buying the bamboo-splitting machine.
3. Contamination refers to the possibility of the nontreatment groups benefitting from the interventions through simple association with the target group.
4. Noncompliers are individuals who are either excluded from the program, although they should have been included, or included in the program although they should not have been.
5. Documented in Karlan and Valdivia (2011).
6. There is recently a renewed interest among economist on subjective measures showing the benefits of these variables (for example, see Ravallion [2012]).
7. COFOPRI (Organismo de Formalización de la Propiedad Informal) is the land-titling organization in urban Lima, Peru.
8. A power calculation quantifies the sample size required to detect with certain power or probability the impact of an intervention, measured by the difference in outcome variables between a treatment and a control group. See appendix B for definitions.

9. The lack of power was mostly due to the existence of correlation within cluster. Only the Peru pilot had a large enough sample. In the Mekong Valley, the final number of observations per cluster and the number of clusters were not that far away from the required minimum.
10. <http://blogs.worldbank.org/impacetevaluations/one-evaluation-one-paper-getting-more-for-your-money>.

Conclusions

Under the results-based initiatives (RBI) program, the World Bank, United Nations Development Fund for Women (UNIFEM), and International Center for Research on Women (ICRW) partnered to launch several small pilot intervention programs focusing on the economic empowerment of women. The purpose was to provide measurable and credible evidence on how small, quick, and targeted policy interventions could help foster better outcomes for women as entrepreneurs, wage earners, or farmers. The RBI thus aimed to fill an important knowledge gap: evidence on what types of interventions were most effective was very limited. Moreover, pilot interventions with impact evaluations had not been applied in a multicountry, gender-focused initiative. The lessons learned from the RBI program in terms of what works in promoting female economic empowerment and what does not could help to fill the knowledge gap. The 12 lessons learned from the design, implementation, and evaluation of the RBI presented in this paper also highlight missed opportunities, areas where robust and generalizable conclusions could not be derived due to shortcomings in the pilots' implementation design. As the use of pilots is becoming commonplace in different forms of development assistance, the RBI experiences can inform the design and implementation of other small-scale pilots with impact evaluations, especially those with a gender focus. The RBI program experiences provide four key messages to guide future pilots.

First, the RBI interventions had limited impact on the ground. On the positive side, women who received training generally appreciated the access to information and felt their skills had benefitted. In the Mekong Valley and Peru, some women also improved their business practices in response to training and joined different forms of networking platforms or associations. Each of these effects could, with time, help empower women economically. Yet, in only one out of five cases—Peru—did the intervention result in measurable improvements in business earnings for participants. In the other cases, wage gaps remained the same (the Arab Republic of Egypt) and earnings did not increase (Kenya, Liberia, and Mekong). Moreover, where measured, there was no impact on investments in children's education and health or on household bargaining power. More

generally, the varied types of programs and different country contexts along with the lack of statistically significant results (or contradictory results) make it difficult to draw general conclusions about the types of interventions tested. In addition, the overambitious program objective and the many problems experienced during implementation and IE stages raise serious questions about the reliability of findings. Thus, it is not clear whether the lack of impacts is due to measurement problems, implementation problems, or because the approach was inherently ineffective.

Second, pilots need adequate funds and time. Pilots are potentially a cost-effective means of testing policy effectiveness. They are quicker and cheaper than large innovations and as such are suitable to investigate new theories of change and test riskier interventions on a smaller scale. However, because of the evaluation component, pilots are more resource demanding and have higher overhead costs than a small program reaching the same number of beneficiaries but without evaluation. Sufficient financial resources are needed to conduct evaluations and to plan programs that may be riskier or scaled up. Sufficient capacity is needed among all teams (designers, implementers, and evaluators), especially to plan the pilot and evaluation and satisfy the information requirements for impact evaluation. Sufficient time is needed for processes to work through (organizational change, putting learning into practice, and so forth). Pilot objectives need to be realistic about what can be achieved with limited budgets and time frames; evaluation objectives/methods also need to be consistent with such constraints.

Third, information should be collected at different times (immediate, intermediate, and final) and at multiple levels (individual and group), and using different types of data (quantitative, qualitative). Although the cost of more comprehensive data collection may be higher, a multilevel approach provides a more complete picture of mechanisms through which policies have an impact (or not). Intermediate data collection and qualitative information also provide a hedge in case unforeseen problems arise in the final impact evaluation. If quantitative impacts cannot be ascertained, qualitative information and monitoring data can provide answers, or may give evidence of other unexpected positive or negative impacts.

Fourth, continuous project monitoring is fundamental, as both an early warning system during program implementation and to complement quantitative analysis at the evaluation stage. A major weakness of the RBI was the lack of continuous program monitoring. As a result, it was not possible to identify implementation problems (for example, in Liberia) early enough to be able to modify the program accordingly. There were also no monitoring data to complement evaluation data to explain why the expected impact did not materialize or was different from what had been expected. Overall, the lack of monitoring makes it impossible to ascertain whether a particular program failed because of the intervention itself or because it was not well implemented.

The use of pilots is now considerably more common than when the RBI program was launched in 2007—and the RBI can help inform current and future research efforts. For example, a survey conducted in parallel to this report by McKenzie and Woodruff (2012) on business development training interventions could only

identify 15 other studies, of which only a handful had been completed by 2012. There are also more interventions aiming to increase gender inequality in economic opportunities. Yet, there remains an important quest for knowledge about what works and what does not. The lessons from the RBI can be helpful in developing this research agenda further. One of the main messages from this report is the need for targeted interventions with adequate funding and time frames long enough to allow impacts to develop. However, the RBI experiences also suggest that more information could be derived from a less heterogeneous set of interventions, with less variation in the type of intervention, country context, and type of beneficiary to permit researchers to draw generalizable conclusions.

Country Case Study Summaries

Promoting Gender Equity and Productivity in Private Firms in the Arab Republic of Egypt: The Gender Equity Model¹

Context and Setup

Women are disadvantaged in the Egypt workplace. They have less access to jobs, especially in the private sector, and have lower earnings than men, beyond what can be explained by differences in education levels. The results-based initiatives (RBI) program took place in a context of high unemployment and large gaps between male and female unemployment rates. In 2007, the unemployment rate for women was three times higher than men's (19 versus 6 percent) and the gaps have been increasing; only one in four female adults is active in the labor market, compared to three in four of adult male. These labor market gaps may be driven partly by norms and values that limit women's opportunities.

In this context, the *Promoting Gender Equity and Productivity in Private Firms—The Equity Model Egypt* (GEME) program aimed to promote gender equity in the private sector by helping firms formulate and achieve gender equity goals and instituting a certification program to recognize good practices through the Gender Equity Seal. The GEME program was a replication of a World Bank program tried in Latin America (box A.1). The program focused on equity in working conditions, women's access to jobs, professional development and training opportunities, and women's participation in decision making in the firms.

Box A.1 The World Bank's Gender Equity Model

Since 2001, the World Bank has been promoting a certification model for gender equity globally. This is a public-private partnership initiative that was tested in Mexico and then replicated in Argentina, Chile, Colombia, the Dominican Republic, Egypt, and recently Turkey.

The World Bank's Gender Equity Model (GEM) promotes gender equity in the private sector and has increased participating firms' productivity and reputations by facilitating greater

box continues next page

Box A.1 The World Bank's Gender Equity Model *(continued)*

diversity and a more motivated workforce. The model builds on a series of good practices in four key areas: recruitment, career development, family-work balance, and sexual harassment policies. Firms in both the public and private sectors adopt the model voluntarily and are provided with technical assistance and training during a four-stage certification process comprising (i) self-assessment, (ii) action plan design and implementation, (iii) preauditing and auditing, and (iv) certification. Each GEM is designed and implemented with country characteristics and needs in mind, and builds on existing capacity. A systematic validation process prior to implementation is carried out in each country to ensure adequacy of the model and ownership by the country.

The model is successful in generating changes in employees' perceptions: both men and women report seeing a reduction in salary gaps and discrimination, improved communications, a greater number of women in leadership and management positions, and a better overall work environment.

Source: World Bank 2009.

Project Implementation

The GEME program was implemented between March 2008 and June 2010 and was organized around several actors:

- United Nations Development Fund for Women (UNIFEM)-Egypt was the lead implementing organization over four consultancy firms that implemented the program and served as technical advisers, with each firm responsible for one of the four components (technical assistance, training, certification, and marketing and communications strategy).
- The government of Egypt, through the Ministry of the Interior, provided inputs into design and implementation and awarded the Gender Equity Seal.
- The World Bank provided financing, technical assistance, and overall supervision.
- The International Center for Research on Women (ICRW) and the Social Research Center at the American University in Cairo conducted the design and implementation of the impact evaluation.
- The advisory committee, with members from the Ministry of Investment, Ministry of Manpower and Emigration and the National Council of Women, encouraged private sector participation, provided links to other government initiatives, and launched dissemination activities.

Program activities fell into four main categories:

1. *Provision of technical services to firms to analyze gender policy and develop a gender action plan:* The implementing partners helped each firm assess its gender policy and practices and identify areas for improvement. Next, they

guided each firm in the development of a gender action plan designed to incorporate gender equitable practices within the firm's organizational structure, build on strengths identified in a preliminary assessment, and address any weaknesses.

2. *Training in areas identified by the plan:* The implementing partners developed a tailored training program aligned to its gender action plan to address the needs of each firm. Training modules covered the practical importance of gender equity for firms, provided an overview of tools to promote gender equity, and discussed recruitment, training, career development, and the prevention of sexual harassment. Attendance and completion rates for the training sessions were high; about 95 percent of those starting the training completed it.
3. *Certification for gender equity:* The program developed a Gender Equity Seal and certification process to recognize firms that succeeded in meeting gender equity standards. The seal enabled them to capitalize on their achievements when seeking new business, marketing products, and attracting new employees.
4. *Development of a marketing and communications strategy for the GEME approach:* To raise awareness of and support for the program, a social marketing firm was hired to develop and implement a marketing and communication plan for employees. This plan was intended to raise awareness of and support for the program and to develop recognition for the Gender Equity Seal. Examples of the marketing strategy include articles on equal work printed in the firm magazine, brochures about the GEME model to distribute among company employees, and information meetings.

The GEME worked with 16 firms recruited between November 2007 and November 2008, plus three additional firms recruited up to July 2009. Selected firms were based in or near the Greater Cairo area, from the private sector, medium to large in scale (number of employees), and had a human resources department. Most selected firms were large: 11 of the 16 firms had more than 500 employees and all of them had more than 100 employees.

Firm recruitment proved difficult because of low awareness of the potential benefits; the GEME also faced low level of initial support from management within firms, which delayed program progress. However, the program was completed as planned in all firms, with training conducted and the Gender Equity Seal approved for all participants. By engaging in discussions with staff, framing the program within existing company policies, emphasizing the value of equal opportunities for all (rather than "just" women), incorporating gender equity as compatible with Islam and cultural values and through the quality of the training program, the implementation teams managed to strengthen internal and external support. Several firms asked the Women Business Development Center to conduct additional training at the firms' expense on specific modules. The program is also expected to be institutionalized and expanded by the government of Egypt in partnership with the National Council for Women.

Recruited firms were from the following sectors: banking and finance (two), manufacturing (six), clothing (four), and pharmaceutical (four). Women have different representation in each of these sectors. While female participation is high in the clothing and pharmaceutical sectors (57 and 46 percent, respectively), it is considerably lower in manufacturing (11 percent). The participation in banking and finance is close the national average (24 percent).

Impact Evaluation

The impact evaluation employed a quasi-experimental design, with control firms selected to match the characteristics of treatment firms. The analysis is based on data from 16 of the 19 firms recruited, 8 treatment firms and 8 control firms. Comparison of the effects between participating and control firms suggested mixed results. No significant changes were found in terms of hiring, training, promotion, or representation of women in management or in augmenting productivity. Part of the problem may be the limited time frame of the program and limited data availability. However, the program did significantly increase employee satisfaction in a number of areas, and employees' perception of gender equity commitment by their firm was strengthened.

Kenya: Strengthening Export Competitiveness of Women Beadworkers²

Context and Setup

Traditional artisanal work is often singled out as a potentially profitable export niche for developing countries and as a vehicle for increasing earnings and income security for women artisans. The potential arguably hinges on the quality of the work, on the entrepreneurial skills of the workers and firms, on the access to international markets, and on an understanding of the requirements for successful exports. Beadwork is a traditional skill among Maasai women. They make colorful, elaborate beaded jewelry and accessories, and virtually all women learn beadwork at a young age from their female relatives. Some sell traditional beadwork items within their community or to tourists and Kenyans from outside the Maasai community. Beadwork is usually not the women's primary economic activity, but rather supplements the livelihoods households earn from livestock. Beadwork is highly flexible, portable work that women work on throughout the day, as their time permits.

The Kenya RBI project focused on increasing the profitability of Maasai women's beadwork activities to generate employment and livelihoods for low-income women. In particular, the project aimed at strengthening export competitiveness and hoped to replicate the strategy in other areas of production, using the lessons learned from the project. The project was implemented in an area of high poverty and low education—45 percent of the inhabitants live below the poverty line, and one in two women cannot read.

The project worked with women's groups involved in beadwork by:

- Providing training and mentoring to develop the group's business and entrepreneurial skills

- Launching activities to help members understand market demand, introduce them to more profitable product designs and training, and improve production skills and quality.
- Helping market groups' products and facilitate large product orders, by helping groups identify niche markets locally and internationally, develop materials, and identify buyers.
- Promoting investment in women's enterprise development and market-based methodologies.

Project Implementation

The project was implemented from October 2008 to September 2010, with the involvement of the following organizations:

- UNIFEM was in charge of the overall coordination and supervision (through a program officer).
- Kenya Gatsby Trust (KGT) and the Centre for International Market Access (CIMA) were the implementation agencies in charge of designing and conducting training, assisting in developing and producing new designs, linking to buyers, and providing mentoring.
- Two national- and local-level steering committees, the Agency for International Health and Development and Rudestat, filled advisory and review roles in design and implementation of the project.
- The World Bank supervised and provided advice.

The program activities included:

- *Training and mentoring to develop the groups' business and entrepreneurial skills.* The specific curriculum was to be developed after assessing group needs. After initial training sessions were finished, the project planned to provide business mentoring on an as-needed basis throughout the project. The project planned extensive follow-on mentoring to help groups once they began receiving orders.
- *Activities to help members understand market demand, introduce groups to more profitable product designs and training, and improve production skills and quality.* At baseline, KGT identified women's weak understanding of markets as a constraint to profitability. A seven-module training contained lessons on bookkeeping, marketing, negotiation skills, and business planning. It was intended to be delivered in four sessions. Therefore, activities were planned to increase women's awareness of markets, develop new products, and improve quality and consistency. After a product and market assessment, the project planned to work with the groups to select and learn more marketable designs. KGT/CIMA designers would mentor women as needed to continue identifying new designs and to understand and meet buyers' expectations. The project also planned to take group members to national and international beadwork

exhibitions, allowing them to interact directly with buyers and to learn about markets.

- *Activities to market groups' products and facilitate large product orders.* The project team aimed to identify niche markets locally and internationally where groups' work could be promoted, to develop marketing materials and stories, and to help groups find buyers of large orders. KGT's marketing arm, CIMA, offers a service that links producers and buyers and was seen as one way that groups could continue to market their products after the project had ended.
- *Advocacy activities to promote investment in women's enterprise development and market-based methodologies to strengthen women's economic empowerment.* The project planned to produce media packages to promote public awareness of the project, hold policy and feedback forums with key national stakeholders, and facilitate links with relevant Kenyan entities to influence the wider policy environment and create more business potential for groups in the future.

The RBI project included some 70 groups, about 1,500 individuals, of which 30 were treatment groups and 40 were control groups (not benefiting from the program). All the groups of Maasai women were in the Kajiado district, a rural area in Kenya's Rift Valley. The groups were already established and nearly two-thirds had existed for more than five years. Although the project was designed to work with groups that had developed group beadwork experiences, in practice, few, if any, of the groups included could be classified as "beadwork groups," where the group would make joint decisions on production and sales. However, nearly all members did some form of beadwork.

The project implementation plan had to be changed along the way for two reasons. First, groups with experience in group-level beadwork could not be identified as expected; as a result, the project had to change to using the group as a means to reach individual beadworkers. Second, a severe drought struck the region, with detrimental effects on household livelihoods, obliging women to search for food and water and reducing the time available for beadwork. Moreover, training was prepared in Swahili, which many of the women did not speak, and required skills in literacy and numeracy that some of the women did not possess.

These events meant that some of the training ended up being less useful to participants, for example, on how to run an organized business, because many of the women were informal operators for which this was not relevant, or could not take in the concepts due to lack of numeracy/literacy skills or language barriers. Because of the drought, less than 10 meetings were held with groups on average, especially toward the end of the project life cycle. Moreover, local demand for beadwork may have fallen during this time.

Market visits and exhibitions provided exposure to market demand and other types of design and business models. They also provided contacts with potential customers and helped women in different groups to connect. The market visits reportedly resulted in orders for beadwork.

Impact Evaluation

The project was evaluated using a pre/posttreatment and control design. Random treatment was not possible, and treatment and control groups could not be in the same geographical area. The primary focus was on group-level outcomes—whether the group was involved in beadwork, changes in sales, revenues, and orders, and so forth—and secondly on individual outcomes. At the final evaluation, 23 treatment and 36 control groups were included. The instruments included interviews on group and individual activities and collected both quantitative and qualitative data.

The analysis suggested no significant positive program impact at the group level, neither in the ways that the group worked (involvement in beadwork, group role in beadwork activities), revenues, or group capacity and sustainability. The only significant program effects were reduced group participation in other forms of activities and reduced cash savings. At the individual level, no significant impact was noted on whether women sold beadwork (at an individual level) or not, on revenue from beadworks, or on household food security.

Focus group discussions suggested market visits and trade fairs had been the most useful, while the impression of training was more mixed. In addition, the project was not complemented with raw materials or start-up capital, and this was considered a disadvantage.

Liberia: Value-Added Cassava Enterprise for the Ganta Concern Women's Group³

Context and Setup

Women produce the majority of agricultural products in Liberia, many of them self-employed working their own land, but they are generally confined to subsistence agriculture and locked out of more profitable cash crop farming. This is particularly true for Nimba county, which remains a poor and underdeveloped area, and has been deeply affected by the civil war. Although the conditions for agriculture are very favorable, economic development is held back by underdeveloped markets, weak infrastructure services, and low human capital, which all combine to limit economic development and entrepreneurship.

The Liberia RBI project set out to raise the profitability of cassava production for local women, both by increasing yields and by promoting some processing of cassava into farina to increase value added and help its marketing. The objectives were to (i) increase economic security through strengthening livelihoods; (ii) empower participants to become independent entrepreneurs, take a leadership role, and control resources at the household level; and (iii) promote the cassava industry as a growth sector and poverty reduction strategy. Apart from the macro level constraints in terms of a postconflict society, low human capital, and poor physical, economic, and financial infrastructure, the project complexity was compounded by the lack of sufficiently skilled persons to work as consultants for the implementation, and of the inexperience of the nongovernmental organization (NGO) sector (in terms of organization and governance) as a whole in Liberia.

Project Implementation

The project was implemented between 2007 and December 2010, and involved the following key actors:

- UNIFEM was in charge of the overall supervision, including one full-time officer based in Monrovia.
- Agricultural Relief Services (ARS), a local agricultural development NGO, conducted implementation.
- The Development Education Network Liberia (DEN-L) provided training and mentoring on organizational development, leadership, and management.
- The Food and Agriculture Organization (FAO) provided tools and grater, and the United Nations Educational, Scientific, and Cultural Organization (UNESCO) provided literacy training.
- Liberia's Central Agricultural Research Institute (CARI) provided improved cassava cuttings.
- National and local steering committees provided guidance.
- ICRW and the Agency for Economic Development and Empowerment were in charge of designing and implementing the impact evaluation.
- The World Bank provided financing and advice.

The project worked with an umbrella organization, the Ganta Concern Women's Group (GCWG), which coordinated 11 community-level women's groups in the Nimba area. The project was intended to both help women's groups increase cassava production and help them process cassava to sell as a higher value-added product:

- A set of activities was developed to increase cassava production on group farms, including helping the community groups that make up the GCWG negotiate for more land for group farms; providing tools; paying for labor to help clear the land; providing higher yielding new varieties of cassava; training women on improved techniques for growing cassava; and providing ongoing extension.
- The centerpiece of the project was construction of a small-scale cassava processing plant for the group. The plant would allow the women to add value and make more profit than selling cassava tubers. It also would address a major constraint to commercial marketing of cassava; taking the water out of the cassava and increasing the value of the final product was expected to make it more economical to transport cassava, in the form of farina, to more distant markets where prices were higher. The plant was expected to produce 1 metric ton of farina daily.
- The project plan also called for training the women to process cassava into farina and assisting the GCWG with marketing the farina. Specific activities to promote marketing and who would perform them were not detailed in advance. The project team expected to find buyers and markets outside of Ganta, including in Monrovia and on the international market. They also

hoped to find institutional buyers such as colleges, who would purchase regularly in bulk for their cafeterias.

- Finally, a set of activities was designed to strengthen the GCWG's organizational governance and management; build its financial, marketing, and gender-sensitive capacity; and promote women's human rights.

For the cassava production, the project helped women negotiate with local government for access to more land for group farms, provided labor and tools for clearance (tools were distributed in December 2007 and in 2008), and provided cuttings of improved Cassava varieties (from CARO) and training on cassava production. For the cassava processing and plant operation, the project purchased land for the plant, organized and paid for plant construction, and procured the equipment.

The project ran into several problems. Construction of the cassava plant was delayed by difficulties in purchasing the land and obtaining the title. Qualified consultants for strengthening the governance of the GCWG and assisting in implementation were extremely difficult to find. Women cassava producers could not or would not work at the processing plant, and full-time plant staff were doing all the processing. The marketing of farina to nonlocal markets or large-scale clients failed because the farina was not priced competitively and the group lacked marketing capacity.

Impact Evaluation

The evaluation strategy was originally planned to measure household-level changes in cassava production, sales, food security, and women's empowerment and provide a baseline survey among group members, and therefore a control group was formed. However, as the project progressed, it became clear that the individual-level impact was much lower than expected because income was lower (with processing much below capacity) and income was not shared among group members. The final evaluation was instead focused on expanded qualitative analysis and reviews of plant records.

The evaluation results suggested that there was very little project impact. Women had appreciated and absorbed the training in new cassava planting techniques, and the cassava production increased. However, for some groups, much of the production could not be sold; none was sold to outside buyers (all went to the plant), the plant was operating much below capacity, and at loss for a variety of reasons including poor equipment; lack of manpower; weak coordination around deliveries and processing resulting in wastage; lack of operational capital; inconsistent quality of the farina; and poor recordkeeping. Lack of plant ownership, financial and governance transparency in the GCWG, and overall lack of monitoring and supervision during project execution appear to have contributed to these outcomes. There appears to be an inconsistency between the relatively quick project that RBI should have been, and the organizational and market change that the cassava project would require. However, the slow but ongoing organizational development of the GCWG may prove beneficial to its members over time.

Mekong: Improving Bamboo Handicraft Value Chains for Women's Economic Empowerment

Context and Setup

Women are often second earners in the household, income that provides a buffer in case of shocks or supplements the main source of income. Bamboo handicraft products are the second largest source of income, after agriculture, for many households in the Mekong area. Women are the main producers of these bamboo products. Although the income from this activity is low, it is perceived as good given the flexibility in time and place, and being “light” work. As a result, 90 percent of households around the Mekong River produce bamboo handicrafts for sale. Producers generally retain less than 50 percent of their value added, and the rest of the share goes to traders and collectors. Producers are usually in a weak market position because they do not have access to final markets and lack of information about demand. As a result, producers are reluctant to make investments that result in improvements and resistant to introducing new designs or improving quality, let alone investing in machinery so they can stop doing the work manually with the only help of a knife.

Against this background, the RBI pilot was implemented in the province of Chhnang in Cambodia, and Vientiane and Champasak provinces in Lao PDR. Its main purpose was to help increase the productive capacity of women in the bamboo handicraft trade by helping them strengthen their organizations, provide them with machinery and technical knowledge, and assist with training in technical, design, and business skills.

Project Implementation

The project began in September 2009 with the formation of producer groups and ended in December 2010. The following organizations were involved in this project:

- The World Bank supervised implementation and the impact evaluation.
- ICRW and the World Bank jointly designed the impact evaluation.
- UNIFEM was the executing agency, and Prosperity Initiative (PI) was the implementing organization.

The project consisted of many simultaneous activities:

- Producer groups formed (usually one of two per village), each receiving a bamboo-stripping machine.
- Groups received training in bamboo handicraft production, marketing, and basic business skills.
- Each producer group leader was part of a “study tour” to markets in Vietnam and Thailand, as well as support for local trade fairs.
- Traders also received study tours to Vietnam and Thailand.

Impact Evaluation

The project was evaluated using a randomized control group after matching villages in pairs on the basis of their district location, their size, the percentage of bamboo handicraft (BH) producer households, the types of BHPs made, and the channels used to market BHPs. One village in each pair was randomly selected to participate in the project, and the other served as control.⁴ The Champasak region was excluded from the impact evaluation because the final decision to include it in the project was made after the baseline survey had been fielded.

Baseline data were collected in March/April 2009 and the end-line data in April/May 2011. Households were matched on the basis of their baseline characteristics using propensity score matching. Information was collected on both BHP and non-BHP households in treatment and control villages (24 treatment villages, 14 in Lao PDR and 10 in Cambodia). The questionnaire was very comprehensive, exploring many aspects including income-generating activities and internal household dynamics (gender roles and household decision making). The quantitative data were complemented with qualitative data coming from focus groups with producers, traders, and in-depth interviews with village leaders.

The impact evaluation focused on three general areas: first, intermediate project effects on BHP outcomes; second, impact on gender empowerment; and third impact on household living standards.

There were no clear significant effects on female economic empowerment. There were positive effects on BH sales and production, both in Cambodia and Lao PDR (increase in the percentage of households producing BHP, aggregate sales indicators, and percentage of households working with producer groups and on its own). For example, in Cambodia, the number of households working with producer groups to make bamboo handicrafts increased by 28 percent between 2009 and 2011. However, there was no evidence from either country that the project had a significant effect on the relative importance of income from BHP relative to either farming or other household income, or in their cash income. In Cambodia, there was a modest increase in relative and absolute BHP income in households of both treatment and control villages, while in Lao PDR, there was a sharp decrease in the relative importance of BHP in households of both control and treatment villages.

Strengthening the Economic Empowerment of Women Microentrepreneurs in Lima⁵

Context and Setup

Lack of entrepreneurial or business skills and lack to access to credit are often identified as key constraints that prevent women from becoming entrepreneurs. However, there is only limited evidence that adopting best practice entrepreneurial methods improves micro- and small businesses' performance, and hardly any to understand how to efficiently transmit business skills. Access to credit is another critical constraint for female business operators, with lack of collateral as the main obstacle. In Peru, a recent land titling program benefited many women

living in urban Lima, providing a unique opportunity for testing how business development training can affect female entrepreneurs when access to credit is not hindered by lack of collateral.⁶

The Peru RBI project targeted female microentrepreneurs living in urban Lima who recently benefited from the land titling of their parcels in the North Cone, Comas and Independencia, and in the South Cone, Villa el Salvador and San Juan de Miraflores. The intervention aimed to bring a better understanding of the effect of business practices on profits when credit constraints are not necessarily binding, but sociocultural constraints possibly are, as well as on what types of training modes may be more effective at delivering results. The training focused on improving business skills and on strengthening networks, but also on strengthening life skills to improve bargaining power in the household and social/personal skills in the market.

Project Implementation

The project was implemented between March 2009 and November 2010. The following organizations were involved:

- The World Bank was responsible for overall coordination and supervision.
- GRADE, in consultation with ICRW, designed the impact evaluation and the survey, and collected the data as well.
- The project was implemented through a consortium of three organizations with expertise in training urban microentrepreneurs: Centro de Servicios para la Capacitación Laboral y el Desarrollo (CAPLAB), which coordinated the other agencies and was responsible for overall project management and communication with project partners and UNIFEM, along with INPET (Instituto de Promoción del Desarrollo Solidario) and CELATS (Centro Latinoamericano de Trabajo Social).

The RBI project included 1,983 eligible women who expressed interest in the program. The women participants were recruited through local radio and newspaper ads and by word-of-mouth at local markets. Of these, 1,418 were randomly assigned in two different treatment groups and 565 in the control group (not benefiting from the program). The businesses involved were generally very small and in the services sector, many food and grocery businesses. Many of the women entrepreneurs were heads of households (main breadwinners), and the majority had comparatively low levels of education. The program suffered from serious attrition, as is common for training programs who were offered the training declined despite having expressed interest and availability at the recruitment. Further dropouts occurred during training. Moreover, implementation was delayed by difficulties in reaching agreement on the impact evaluation method, which in turn resulted in shorter than anticipated time for the project to run (and have an impact).

The training was organized in two components: a general training (GT) component delivered to all women in the treatment group, and a technical assistant

(TA) component delivered only to half of the women who completed the general training. The content of general training component consisted of best practices associated with successful female microentrepreneurs. It was organized in three modules: personal development, business development, and management and productivity improvements. The first module focused on the strengthening of women's self-esteem, social skills, and tools for life planning to empower women as individuals, but also as members of their families and communities. The second module focused on tools to plan new businesses, process innovations in current ones, marketing and sale strategies, and costing. The technical assistance also was divided in the same three modules, but tailored to the particular needs of each woman entrepreneur. It combined individual visits to businesses with group sessions for similar businesses.

Impact Evaluation

The method of the impact evaluation was randomization, allowing for estimation of the causal impact of the training by comparing the outcomes of businesses of women in treatment and control groups (intention-to-treat effects). Almost all of the women were reinterviewed at the end line (1,627 of the 1,983 women interviewed in the baseline, that is, 82 percent). The results suggested that combined GT and tailored TA had more positive effects than GT alone across several dimensions. Combined modes of training had a significant positive effect on sales, and also increased the time dedicated to studies by mothers, male adults, *and* adolescent children in the household. Those with GT did not significantly increase their sales compared to control group. There was no impact on any of the variables that reflect the decision process within the household: household saving decisions, borrowing, home accounting, the number of children, or when to have children.⁷

Notes

1. Information of this section is mainly based on Golla and Selim (2011).
2. This section is mainly built from Golla and Sagers (2011).
3. This section builds on Golla and Jones-Demen (2011).
4. In addition, some villages were nonrandomly assigned to the treatment or control group at the request of the project team because of commitments made to the villages during the process of project design.
5. This section is based on Valdivia (2011a, 2011b).
6. The land administration programs in which eligible women participated are: the Urban Property Rights Project (UPRP) implemented from 1998 to 2004, and the Real Property Rights Consolidation Project (RPRCP) currently under way.
7. These results are only discussed in the Valdivia (2011b).

Technical Concepts in Impact Evaluation Design

- **Impact evaluation (IE)** is a particular type of evaluation that aims at measuring the changes in key outcomes that are directly attributable to an intervention.
- **Theory of change** refers to the expected channel through which an intervention is to deliver the desired results.
- **Randomized control trial:** IE method consisting of a treatment and a control group that are randomly selected from the target population, using only stratification rules to ensure representation of relevant subgroups of the population. Successful randomization results in unobservables equally distributed in treatment and control groups.
- **Regression discontinuity:** IE method consisting of a treatment and a control group result from a cutoff, which can be a consequence of nature or of the design. Successful cutoffs result in unobservables equally distributed around the cutoff point.
- **Differences-in-differences:** IE method consisting of comparing a treatment and a comparison group (first difference) before and after the intervention (second difference).
- **Matching:** Control group is matched to the treatment group by using the propensity score (predicted probability of participation given observed characteristics).
- **Type I error** is made when an evaluation concludes that a program has had an impact, when in reality it had no impact.
- **Type II error** is made when an evaluation concludes that the program has had no impact, when in fact it has had an impact.
- The **power** is the probability of observing an impact when there is one. A high power implies a low probability of type II error.
- **Minimum sample size** requirements increase if the evaluation wants to avoid type II errors, if the predicted impact of the intervention is small, if the variance of the outcome indicator is large, and if there is interest in measuring heterogeneous effects among beneficiaries.

- **Mixed methods** approaches combine quantitative approaches that permit estimates of magnitude and distribution effects and generalization tests of statistical differences with qualitative approaches that permit in-depth description, analysis of processes, and patterns of social interaction (Bamberger, Rao, and Woolcock 2010).
- **Measure context:** These are economic context; policy environment and politics; institutions and operational context; physical infrastructure; and socio-economic and cultural characteristics of the affected populations.
- **Eligibility or targeting:** These are the set of rules that define who is eligible for program benefits.
- **Eligible population or target population** is defined by those individuals, firms, or groups that could benefit from the intervention and for whom the evaluators are interested in knowing the impact of an intervention. The target population is usually defined by a clear set of rules that allow separating treatment and control groups.
- **Crossover design** occurs when the same sample includes more than one independent intervention. Several independent lotteries are conducted at once on one sample of beneficiaries.
- **Monitoring** is a continuous process that tracks what is happening within a program and uses the data collected to inform program implementation and day-to-day management decisions.

APPENDIX C

Power Calculations for Mekong RBI

Kampong Chhnang province, Cambodia

Increase in income (%)	<i>Alpha = 0.05, Power = 0.90</i>			<i>Alpha = 0.10, Power = 0.80</i>		
	<i>Minimum sample size for a simple random sample</i>	<i>Minimum number of clusters with 20 observations per cluster</i>	<i>Minimum number of observations per cluster with 12 clusters</i>	<i>Minimum sample size for a simple random sample</i>	<i>Minimum number of clusters with 20 observations per cluster</i>	<i>Minimum number of observations per cluster with 12 clusters</i>
10	2,198	739	n.a.	1,294	390	n.a.
25	408	138	n.a.	240	81	n.a.
50	130	40	n.a.	76	26	n.a.
75	72	25	n.a.	44	15	n.a.
100	50	17	n.a.	30	12	8

Vientiane province, Lao PDR

Increase in income (%)	<i>Alpha = 0.05, Power = 0.90</i>			<i>Alpha = 0.10, Power = 0.80</i>		
	<i>Minimum sample size for a simple random sample</i>	<i>Minimum number of clusters with 20 observations per cluster</i>	<i>Minimum number of observations per cluster with 18 clusters</i>	<i>Minimum sample size for a simple random sample</i>	<i>Minimum number of clusters with 20 observations per cluster</i>	<i>Minimum number of observations per cluster with 18 clusters</i>
10	2,110	459	n.a.	1,242	270	n.a.
25	392	86	n.a.	232	51	n.a.
50	126	28	n.a.	74	17	13
75	70	16	11	42	10	4
100	48	11	5	30	7	2

Source: Cambodia (2004 CSES); Lao PDR (2002/03 LECS III); exploratory work (Jim Knowles, September 2008).

Note: n.a. = not applicable.

Bibliography

- Atkin, David. 2009. "Working for the Future: Female Factory Work and Child Health in Mexico." Unpublished manuscript, Yale University.
- Bamberger, Michael, Vijayendra Rao, and Michael Woolcock. 2010. "Using Mixed Methods in Monitoring and Evaluation." Policy Research Working Paper No. 5245, World Bank, Washington, DC.
- Bamberger, Michael, Jim Rugh, and Linda Mabry. 2006. *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Thousand Oaks, CA: Sage Publications.
- Bardasi, Elena, Kathleen Beegle, Andrew Dillon, and Pieter Serneels. 2010. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania." Policy Research Working Paper No. 5192, World Bank, Washington, DC.
- Bobonis, Gustavo J. 2009. "Is the Allocation of Resources within the Household Efficient? New Evidence from a Randomized Experiment." *Journal of Political Economy* 117 (3): 453–503.
- Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics* 1 (4): 200–32.
- de Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." *Quarterly Journal of Economics* 123 (4): 1329–72.
- . 2009. "Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns." *American Economic Journal* 1 (3): 1–32.
- Doepke, Matthias, and Michele Tertilt. 2011. "Does Female Empowerment Promote Economic Development?" Policy Research Working Paper No. 5714, World Bank, Washington, DC.
- Duflo, Esther. 2000. "Child Health and Household Resources: Evidence from the South African Old-Age Pension Program." *American Economic Review Papers and Proceedings* 90: 191–202.
- . 2003. "Grandmothers and Granddaughters: Old-Age Pensions and Intra-household Allocation in South Africa." *World Bank Economic Review* 17 (1): 1–25.
- Duflo, Esther, and Christopher Udry. 2004. "Intra-household Resource Allocation in Côte d'Ivoire: Social Norms, Separate Spheres Accounts and Consumption Choices." NBER Working Paper No. 10498, National Bureau of Economic Research, Cambridge, MA.

- Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics*, Volume 4, Chapter 61, edited by Paul Schultz and John Strauss. Amsterdam: North Holland.
- Fafchamps, Marcel, David McKenzie, Simon Quinn, and Christopher Woodruff. 2011. "When Is Capital Enough to Get Female Enterprises Growing? Evidence from a Randomized Experiment in Ghana." Policy Research Working Paper No. 5706, World Bank, Washington, DC.
- Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington, DC: World Bank.
- Golla, Anne Marie. 2011. "Impact Evaluation for the Small-Scale Project: What Do Managers Need to Know? Lessons Learned from the Results-Based Initiative (RBI) Program." International Center for Research on Women, unpublished.
- Golla, Anne Marie, and Annie Jones-Demen. 2011. "RBI Program Liberia: Value-Added Cassava Enterprise for the Ganta Concern Women's Group in Liberia." International Center for Research on Women, unpublished.
- Golla, Anne Marie, and Meredith Saggars. 2011. "Kenya RBI Project: Strengthening Export Competitiveness of Women Bead Workers." International Center for Research on Women, unpublished.
- Golla, Anne Marie, and Mona Selim. 2011. "Egypt Results-Based Initiative: Promoting Gender Equality and Productivity in the Private Firms in Egypt; The Gender Equity Model Egypt." Impact Evaluation Report, ICRW, unpublished.
- ICRW (International Center for Research on Women). 2007. "Impact Evaluation of RBI." ICRW-UNIFEM-World Bank consultation, unpublished.
- Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5–86.
- Jha, Saumitra, Vijayendra Rao, and Michael Woolcock. 2007. "Governance in the Gullies: Democratic Responsiveness and Community Leadership in Delhi's Slums." *World Development* 35 (2): 230–46.
- Karlan, Dean, and Martin Valdivia. 2011. "Teaching Entrepreneurship: Impact of Business Training in Microfinance Clients and Institutions." *Review of Economic and Statistics* 93 (2): 510–27.
- Khandker, Shahidur R., Gayatri B. Koolwal, and Hussain Samad. 2009. *Handbook on Quantitative Methods of Program Evaluation*. Washington, DC: World Bank.
- Knowles, Jim. 2011. "Final Report: Impact Evaluation of the Mekong Results-Based Intervention." World Bank, Washington, DC, unpublished.
- Lundberg, Shelly, Robert A. Pollack, and Terence J. Wales. 1997. "Do Husbands and Wives Pool Their Resources? Evidence from the United Kingdom Child Benefit." *Journal of Human Resources* 32 (3): 463–80.
- McKenzie, David. 2010. "Impact Assessments in Finance and Private Sector Development: What Have We Learned and What Should We Learn?" *World Bank Research Observer* 25(2): 209–33.
- McKenzie, David, and Christopher Woodruff. 2012. "What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?" Policy Research Working Paper No. 6202, World Bank, Washington, DC.

- Rao, Vijayendra, and Michael Woolcock. 2003. "Integrating Qualitative and Quantitative Approaches in Program Evaluation." In *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*, edited by Francois J. Bourguignon and Luiz Pereira da Silva, 165–90. New York: Oxford University Press.
- Ravallion, Martin. 2008. "Evaluating Anti-Poverty Programs." In *Handbook of Development Economics*, Volume 4, edited by Paul Schultz and John Strauss. Amsterdam: North Holland.
- . 2009. "Evaluation in the Practice of Development." *World Bank Research Observer* 24 (1): 29–53.
- . 2012. "Poor, or Just Feeling Poor? On Using Subjective Data in Measuring Poverty." Policy Research Working Paper No. 5968, World Bank, Washington, DC.
- Rubalcava, Luis, Graciela Teruel, and Duncan Thomas. 2009. "Investments, Time Preferences, and Public Transfers Paid to Women." *Economic Development and Cultural Change* 57 (3): 507–38.
- Thomas, Duncan. 1990. "Intra-Household Resource Allocation: An Inferential Approach." *Journal of Human Resources* 25 (4): 635–64.
- Todd, Petra. 2012. "Effectiveness of Interventions Aimed at Improving Women's Employability and Quality of Work. A Critical Review." Policy Research Working Paper No. 6189, World Bank, Washington, DC.
- UN Women. 2011. "Results-Based Initiative Program Evaluation: Report on Findings." Authored by Gary Woller, unpublished.
- Valdivia, Martin. 2011a. "Training or Technical Assistance for Female Entrepreneurship? Evidence from a Field Experiment in Peru." World Bank, Washington, DC.
- . 2011b. "Training or Technical Assistance for Female Entrepreneurship? Evidence from the Second Follow-Up for the Peruvian Field Experiment." World Bank, Washington, DC.
- World Bank. 2001. *Engendering Development: Through Gender Equality in Rights, Resources, and Voice*. World Bank policy research report, World Bank, Washington, DC.
- . 2006. *Conducting Quality Impact Evaluations under Budget, Time and Data Constraints*. Washington, DC: Independent Evaluation Group, Knowledge Programs and Evaluation Capacity Development (IEGKE).
- . 2009. "The World Bank's Gender Equity Model (GEM): Country Success Stories." World Bank, Washington, DC.
- . 2011a. "Gender Equality and Development." World Bank, Washington, DC.
- . 2011b. "Gender Equality as Smart Economics: World Bank Group Gender Action Plan." Four-Year Progress Report (January 2007-December 2010), May 2, 2011. http://siteresources.worldbank.org/INTGENDER/Resources/4year_progress_report_May5.pdf.
- . 2012. "Jobs." World Bank, Washington, DC.

Environmental Benefits Statement

The World Bank is committed to reducing its environmental footprint. In support of this commitment, the Publishing and Knowledge Division leverages electronic publishing options and print-on-demand technology, which is located in regional hubs worldwide. Together, these initiatives enable print runs to be lowered and shipping distances decreased, resulting in reduced paper consumption, chemical use, greenhouse gas emissions, and waste.

The Publishing and Knowledge Division follows the recommended standards for paper use set by the Green Press Initiative. Whenever possible, books are printed on 50 percent to 100 percent postconsumer recycled paper, and at least 50 percent of the fiber in our book paper is either unbleached or bleached using Totally Chlorine Free (TCF), Processed Chlorine Free (PCF), or Enhanced Elemental Chlorine Free (EECF) processes.

More information about the Bank's environmental philosophy can be found at <http://crinfo.worldbank.org/wbcrinfo/node/4>.



Launched in 2007, the Results-Based Initiatives (RBI) aimed to provide comprehensive, coherent, and rigorous evidence on effective interventions that foster the economic empowerment of women. The RBIs comprised five small pilots with a built-in impact evaluation designed to identify the best method to promote better outcomes for women as entrepreneurs, wage earners, or farmers—under different country contexts.

The program was an innovative experiment in an important policy area. Although there is a clear rationale for policy interventions to help remove constraints to women's economic empowerment, knowledge remains limited on the interventions that work best in different settings. When the RBI were conceived, rigorous evidence in this area was close to nonexistent because no systematic impact evaluations had been carried out in developing countries. However, the RBI fell short of meeting several ambitious objectives.

Lessons Learned and Not Yet Learned from a Multicountry Initiative on Women's Economic Empowerment highlights lessons from the RBI with respect to the impact of the interventions and the dos and don'ts in pilot design and implementation. Regarding the impact on economic opportunities, the interventions did not increase women's earnings, except in the Peru pilot. In general, women who received training appreciated the access to new information and expressed an increase in their skills and involvement in business associations and networks.

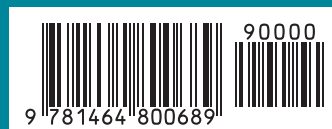
However, it is incorrect to conclude that these interventions were not effective. The lack of robust positive impact could be due to evaluations being conducted too soon and being unable to fully show the long-term effects of the interventions. In particular, an early warning system to synchronize the corrections in the interventions with the design of the impact evaluation is clearly needed. The RBIs were overambitious regarding achievement potential on a limited budget and short time frame.



THE WORLD BANK



ISBN 978-1-4648-0068-9



SKU 210068