



# Integrating gender into project-level evaluation

ECG reference document

# ECG

EVALUATION COOPERATION GROUP

## Acknowledgements

This reference document is the product of the collective work of an ECG Task Force composed of the following institutions and members, led by the AfDB, and supported by Michael J. Bamberger, Expert Consultant.

Task force members	
<b>Independent Development Evaluation (IDEV)</b> African Development Bank Group (AfDB)	Samer Hachem, Division Manager Sohna Ngum, Consultant (Office of the Special Envoy on Gender) Jessica Harris, Consultant (Office of the Special Envoy on Gender)
<b>Independent Evaluation Department (IED)</b> Asian Development Bank (AsDB)	Farzana Ahmed, Lead Evaluation Specialist Hyun Son, Principal Evaluation Specialist
<b>Evaluation Department (EvD)</b> European Bank for Reconstruction and Development (EBRD)	Shireen El-Wahab, Principal Evaluation Manager Beatriz Perez-Timmermans, Principal, Evaluation Manager
<b>Operations Evaluation (EV)</b> European Investment Bank (EIB)	Emmanuel Pondard, Evaluation Specialist
<b>Independent Evaluation Office (IEO)</b> Global Environment Facility (GEF)	Geeta Batra, Deputy-Director Anna Viggli, Senior Evaluation Officer
<b>Independent Evaluation Office (IEO)</b> IFAD	Catrina Perch, Evaluation Specialist Mark Keating, Evaluation Officer
<b>Independent Evaluation Office (IEO)</b> UN Women	Shravanti Reddy, Evaluation Specialist Sabrina Evangelista, Evaluation Specialist
<b>Independent Evaluation Group (IEG)</b> World Bank Group (WBG)	Gisela Garcia, Evaluation Officer Elena Bardasi, Senior Economist

© 2017 Evaluation Cooperation Group  
All rights reserved – Published June 2017

**Integrating gender into project-level evaluation**  
ECG reference document, June 2017

Publication coordinated by:

**Independent Development Evaluation (IDEV)**  
African Development Bank Group

AfDB Headquarters  
Avenue Joseph Anoma, 01 BP 1387, Abidjan 01, Côte d'Ivoire  
Phone: +225 20 26 20 41  
E-mail: [idevhelpdesk@afdb.org](mailto:idevhelpdesk@afdb.org)  
[idev.afdb.org](http://idev.afdb.org)



# TABLE OF CONTENTS

	Acronyms	v
	Introduction	vii
<b>Chapter 1</b>	<b>Addressing gender within the broader independent evaluation framework</b>	<b>1</b>
	1.1 Gender equality and women's empowerment are development priorities that must be included in IEO evaluation programs	1
	1.2 The parameters within which IEO evaluations are conducted: unique opportunities and challenges	2
	1.3 Assessing the value added of gender-responsive evaluations	3
<b>Chapter 2</b>	<b>Deciding which project-level evaluations should address gender and defining gender-focused questions to be addressed</b>	<b>6</b>
	2.1 Scenarios for incorporating gender into IEO evaluations	6
	2.2 Defining the scope and intensity of a gender-responsive evaluation	7
	2.3 Strategies for deciding which evaluations should have a more in-depth gender-focus	9
	2.4 Defining key gender evaluation questions	12
<b>Chapter 3</b>	<b>Designing gender-responsive evaluations at the project level</b>	<b>17</b>
	3.1 The gender-responsive approach to evaluation	17
	3.2 Designs to collect essential data for gender-responsive evaluations	19
	3.3 Designs for more in-depth gender-responsive evaluations	20
	3.4 Evaluability assessment	21
<b>Chapter 4</b>	<b>Tools for collecting data for gender-responsive evaluations</b>	<b>28</b>
	4.1 Defining evaluation questions and constructing a design matrix	28
	4.2 Methods of data collection	29
	4.3 Developing and using gender responsive checklists, indexes and indicators	31
	4.4 Integrating data collection for gender evaluations into ongoing surveys	33
<b>Chapter 5</b>	<b>Managing gender evaluations</b>	<b>37</b>
	5.1 The role of the evaluation manager in gender-responsive evaluations	37
	5.2 The role of the manager at each stage of a gender-responsive evaluation	37
	References	41

## Part II: Case studies illustrating the application of the recommended GRE design framework

(separate document)

[Download Part II](#)

- Chapter 6 **Example of the application of GRE design framework: Evaluating a hypothetical groundnut infrastructure development project with defined gender objectives**
- Chapter 7 **Evaluating women's empowerment in a hypothetical village development project in Central Asia**
- Chapter 8 **Evaluating secondary and tertiary gender outcomes: A hypothetical public transport project in Eurasia**

## Appendixes

(separate document)

[Download Appendixes](#)

- Appendix 1 **The rationale for integrating gender into IEO evaluations (to be included)**
- Appendix 2 **Special challenges facing gender evaluations**
- Appendix 3 **Overview of GRE evaluation designs for basic and in-depth evaluations**
- Appendix 4 **Overview of data collection methods for GRE evaluation designs**
- Appendix 5 **A data collection planning matrix**
- Appendix 6 **The proposed GRE design for a hypothetical village development project in Central Asia**
- Appendix 7 **The proposed GRE design for a hypothetical public transport project in Eurasia**
- Appendix 8 **Links to resources for Chapter 2 Box 2.5**
- Appendix 9 **Strengthening the dissemination and use of gender-responsive evaluation findings**

## Acronyms

<b>ECG</b>	Evaluation Cooperation Group
<b>GN</b>	Guidance Note (Integrating gender into the evaluation programs of IEOs)
<b>GRE</b>	Gender-responsive evaluation
<b>ICT</b>	Information and communication technology (smartphones and other portable electronic devices)
<b>IEO</b>	Independent Evaluation Office
<b>IFAD</b>	International Fund for Agricultural Development
<b>IFI</b>	International financial institution
<b>PRA</b>	Participatory Rural Appraisal (now used as a generic term for any kind of participatory group planning or evaluation process)
<b>PSM</b>	Propensity score matching
<b>QCA</b>	Qualitative comparative analysis
<b>QED</b>	Quasi-experimental design
<b>RCT</b>	Randomized control trial
<b>SDGs</b>	Sustainable Development Goals
<b>TOC</b>	Theory of change
<b>UNEG</b>	United Nations Evaluation Group
<b>VAW</b>	Violence against women



# Introduction

The purpose of this ECG Reference Document is to provide practical approaches for Independent Evaluation Offices (IEOs) of multilateral development institutions to strengthen the treatment of gender equality and women's empowerment in their evaluation programs, and more specifically in evaluations of projects. It is intended for three main audiences: staff who are relatively new to development evaluation, and particularly to gender evaluation; staff with more experience in development evaluation as well as gender evaluation; and managers who must lead the gender evaluations. Recognizing the different levels of evaluation experience, the document is structured to provide guidance on both basic gender-responsive evaluation (GRE) tools and techniques required to cover the minimum requirements for incorporating gender into standard IEO evaluations; and to present more advanced tools and methods for evaluations where resources, time and management interest require the application of more advanced methods of GRE. A separate chapter is included on the management of gender evaluations as the evaluation manager plays an important role in ensuring: that the gender-focused evaluation questions are aligned with agency priorities; that support and cooperation of operations staff and senior IEO management is assured; and that the evaluation teams are able to work effectively with the new disciplines and approaches that might be required for the gender evaluations.

The Reference Document is structured around several key principles and messages.

First, GRE can use all the standard evaluation tools and techniques, but there are a set of guiding principles used in all GRE. These include:

- recognition that GRE does not just focus on women but examines the differential impacts of development interventions on women and men;
- the evaluations focus on removing barriers to equity and equality, human rights and economic and social empowerment of women;
- the use of a participatory and consultative approaches; and
- the use of mixed-methods designs that combine quantitative and qualitative approaches.

Given that women continue to suffer significant economic, political, legal, social and cultural disadvantages in almost all countries, evaluation strategies must balance the need to address these gender realities, while also focusing on the broader goals of understanding the complex relations between the sexes, and areas where some groups of men may suffer disadvantages, and how these affect development objectives.

Second, while many projects have defined gender objectives, these are often quite narrow in scope, often focusing on only a few (but important) quantitative indicators of participation and access to project services. The proposed evaluation approach broadens the analysis to go beyond the *primary gender objectives* defined in the project design, to identify and assess a wide range of *potential secondary outcomes* referring to broader effects on the target population, and in some cases *tertiary outcomes* referring to spillover effects on households and communities not participating in the project. This broader focus is critical because projects may have a much greater and wider impact than normally addressed in current GRE approaches. The decision on the scope of each GRE must, of course be negotiated with IEO and operations management.

Third, GRE evaluation teams may have an advocacy and “marketing” role to convince stakeholders inside IEO and in other parts of the agency on the value added of GRE, and showing that there are practical and economical ways to incorporate gender into the evaluations.

Fourth, most project evaluations are conducted retrospectively meaning that IEO evaluations are not able to employ many of the conventional evaluation techniques. In particular, it is usually not possible to use most of the experimental and

quasi-experimental designs, requiring pre-test post-test comparisons. Consequently, creativity is required to achieve the highest degree of evaluation rigor while working under these constraints.

The reference document is organized in two parts. Part I contains five chapters laying out the proposed approach and methodology. Part II presents three case studies illustrating how the proposed evaluation framework and methodologies could be applied in typical IEO evaluation contexts. There are also nine appendixes that provide more detailed material relating to the topics covered in the main report.

Chapter 1 addresses gender within the IEO evaluation framework. It stresses the importance of gender in all IEO evaluations while recognizing the unique challenges and opportunities presented by the parameters within which IEO evaluations are conducted. The chapter also emphasizes the need to demonstrate the value added of gender-focused evaluations.

Chapter 2 provides guidelines for deciding which project-level evaluations should address gender and with what level of intensity. There are different scenarios depending on whether, and how far, gender has been incorporated into the agency’s evaluation strategy. The question of why and how to examine gender dimensions for projects that do not have defined gender objectives is addressed. The need to go beyond the evaluation of defined gender objectives and consider the inclusion of secondary and tertiary outcomes is introduced. Finally, guidance is given on how to select the key evaluation questions to address. The proposed structure for GRE reports is introduced, and the application of this framework is illustrated in the three case studies presented in Part II.

Chapter 3 reviews the main evaluation approaches that can be used for GRE. This can draw on all standard evaluation designs, but all GRE evaluations apply a distinct gender-responsive lens – including the application of a gender analysis framework. The distinction is made between evaluation designs to collect essential data for standard GRE evaluations, and designs for more in-depth and advanced GRE. Two tables summarize all the basic and more advanced designs, their strengths and potential weaknesses and (for the basic designs) tips for strengthening the designs. A third table cites examples of how each evaluation approach has been applied in a gender-focused evaluation.

Chapter 4 discusses the data collection tools that can be used for each design, and a table summarizes the methods and offers tips on how to avoid common pitfalls and strengthen applications. It is also recommended that all GRE should incorporate a mixed-methods approach to understand the complex processes of behavioral change and empowerment and to improve the validity of the data and the depth of the interpretation through triangulation. Reference is made to the wide range of gender

checklists, indexes and indicators that can strengthen data collection. Finally, guidelines are offered for reducing the costs and time of data collection through coordination with other planned or ongoing studies (“piggy-backing”).

Chapter 5 discusses the important roles of the manager of the gender-responsive evaluations. Some of these roles include: ensuring the gender-focus is aligned with agency objectives, ensuring the evaluation addresses the key gender-related questions of concern to key stakeholders, and team-building to incorporate effectively new professional expertise. The manager must also coordinate with operations departments to ensure access to required gender-related data (which can sometimes be more difficult and time-consuming to extract and where there may be some reluctance on the part of busy operations staff), and s/he will often take the lead in promoting dissemination and use of the findings.

Finally, while there are a set of GRE guidelines that can be widely applied, it is recognized that each IEO works within a unique organizational context, and that each IEO must adapt the GRE guidelines to their own organizational mandate. ■



# Chapter 1

## Addressing gender within the broader independent evaluation framework

### 1.1 Gender equality and women's empowerment are development priorities that must be included in IEO evaluation programs

According to the UN Evaluation Group (UNEG), “gender equality<sup>1</sup> entails the concept that all human beings, both men and women, are free to develop their personal abilities and make choices without the limitations set by stereotypes, rigid gender roles, or prejudices. Therefore, needs of women and men's rights, responsibilities and opportunities will not depend on whether they are born male or female. Gender equity in the development context usually implies measures to compensate for the historical and social disadvantages of women<sup>2</sup>.”

It is important to recognize, as does the UNEG definition, that gender equality (and consequently gender analysis) concerns limitations imposed by society on the ability of both women and men to achieve their full potential. Many organizations tend to focus their gender strategies on addressing the severe historical and social

disadvantages of women and tend not to analyze the differential effects of social and economic change, and of project interventions on both women and men. Given resource constraints and the clear challenges facing women in most societies, strategic choices must be made on how resources for promoting gender equality are to be allocated. However, it is important that the GRE examine the consequences of agency policies and projects on both women and men. It is of course the prerogative of policy makers to decide how the evaluation findings will affect future policies and projects.

Two areas where the importance of this broader definition of gender analysis are coming to be recognized are the multiple causes and consequences of gender-based violence (for example, research on masculinity and how young males are socialized to believe that

1 There continues to be a lack of clarity concerning the relationship between *equity* and *equality* both in general and how they relate to gender analysis. For example, UNICEF often uses the term *equity* (Bamberger and Segone 2011 Section 1) to describe what some other agencies would define as equality. For a discussion of these concepts from the perspective of a developmental economist see Kumar (2013)

2 Source: <http://www.uneval.org/document/detail/1616>

violence is the only appropriate response in many conflict situations and that fathering a number of children is an essential proof of manhood); and the effects of changing employment patterns on both women and men (for example, how unemployment and the inability to fulfill his role as bread-winner can result in male depression, alcoholism and increased domestic violence).

All international financial institutions (IFIs) now have policies recognizing gender equality as a development objective. However, many of these policies are relatively new and still in the process of being operationalized. Consequently, many IEOs are still developing policies and strategies to

integrate gender into their evaluation programs. So, the IEO teams responsible for promoting and implementing gender-responsive evaluations (GREs) may have to play an advocacy role to make the case both inside and outside their own office why GRE is important to achieving agency development objectives. Given resource constraints, they will also have to demonstrate the value-added from allocating scarce evaluation budget resources to address gender. There is extensive literature, much of it produced by the IFIs, on the critical linkages between gender, equity, human rights and development objectives (see Box 1.1). The SDGs also emphasize the importance of gender and equity in the evaluation of development interventions.

### Box 1.1 Illustrative references on gender and development and the contribution of GRE

- World Development Report (2012). Gender Equality and Development.
- UNEG (2014). Integrating Human Rights and Gender Equality in Evaluations.
- Organisation for Economic Cooperation and Development Development Assistance Committee (OECD-DAC) (2016). Handbook on the OECD-DAC Gender Equality Policy Marker.
- Independent Evaluation Group (IEG), World Bank (2016). Integrating Gender into IEG Evaluation Work.
- International Fund for Agricultural Development (IFAD) (2017). What Works for Gender Equality and Women's Empowerment: A Review of Practices and Results. An Evaluation Synthesis.
- UN Women and the Sustainable Development Goals. <http://www.unwomen.org/en/news/in-focus/women-and-the-sdgs>

## 1.2 The parameters within which IEO evaluations are conducted: unique opportunities and challenges

IEOs operate within a mandate that defines the scope of their evaluations, the time frames within which they operate and the evaluation methodologies they use. Given these parameters, the IEO approach to evaluation in general, and GRE in particular, limits the application of many approaches discussed in most evaluation textbooks (see Appendix 2). IEOs were established to ensure

their independence so that they can provide an objective assessment to the Board of Directors and other key stakeholders on the extent to which projects and other interventions have achieved their objectives, whether they have used their resources in the most efficient way and whether they have complied with the administrative procedures defined by their respective agencies.

This mandate provides both unique opportunities and challenges for conducting GREs. The opportunity arises from the fact that IEOs report directly to the Board of Directors, and that there are: (i) procedures to ensure the objectivity and independence of the evaluations and hence their credibility; and (ii) institutionally defined mechanisms for the dissemination and use of evaluation. Challenges arise from the fact that IEO evaluations are conducted ex-post after projects have closed, so it is usually impossible to influence the kinds of baseline or implementation data that are collected on the projects being evaluated. Consequently, it is impractical to use the kinds of pre-test post-test experimental and quasi-experimental evaluation designs discussed in evaluation textbooks. These challenges are particularly important for GRE where women's empowerment is generated through social mechanisms and processes of behavioral change that are even more complex than those affecting

men as many societies have developed religious, economic, legal and social mechanisms specifically to control and regulate the actions of women. Ideally, these processes should be observed over time rather than assessed through recall at the end of the project as they are so deeply engrained that many women and men are not aware of their existence or how they affect their lives. This document also refers to new information technologies (smart phones and big data) that make it possible to reconstruct baseline and longitudinal data, broadening the range of evaluation methodologies (Bamberger 2017).

Given these considerations, two GRE approaches will be presented: (i) those that are already being used by IEOs to collect essential gender-related data; and (ii) innovative approaches that could be applied to a sub-set of evaluations where the mandate, time and resource make it possible to dig deeper.

### 1.3 Assessing the value added of gender-responsive evaluations

GRE are often more expensive and time-consuming than conventional evaluations in particular due to more expensive data to be collected. Consequently, IEOs will need to demonstrate that the additional evaluation information is aligned with organizational policies, that it provides an improved estimate of project outcomes and that the value added of the analysis justifies the additional time and costs<sup>3</sup>. This will require:

- An estimation of the additional time, cost and organizational effort required to conduct the GRE and guidelines on how these can be reduced.

- Developing different indicators to estimate the value-added in terms of, for example, increased project beneficiaries (particularly in terms of the poorest and most vulnerable groups), increased benefits to target groups, higher rates of return, and improved project efficiency. It is also important to assess new technical understanding and knowledge products. GRE may also contribute to the “Do no harm” policy endorsed by many agencies, by identifying and addressing negative outcomes. For GRE, this may include addressing gender-based violence and eliminating social, economic, legal and

3 Although it does not specifically address gender, Levin and McEwan (2001) *Cost-effectiveness analysis* provides a useful overview of the different approaches that can be used for assessing cost-effectiveness.

other barriers to women's participation and empowerment.

- Methodologies to assess the value-added of GRE for projects with and without defined gender objectives. In the first scenario, gender-related objectives have already been defined and the challenge is to measure them. However, in the second scenario the case has to be made as to why secondary and tertiary gender outcomes (that were not identified in the project document) are important, as well as to measure them.

An additional challenge is to develop gender metrics that will be convincing to program managers who may not be aware of the importance of many of the gender-related outcomes such as women's empowerment, greater freedom of movement and feelings of security that are often proposed by feminists and women's advocates as gender outcomes. This means the indicators and analysis must be aligned both with women's needs and with government policies, and ideally agency organizational gender goals should embrace both. ■



# Chapter 2

## Deciding which project-level evaluations should address gender and defining gender-focused questions to be addressed

### 2.1 Scenarios for incorporating gender into IEO evaluations

While all IFIs now have a gender policy or strategy, for many agencies the policies are still quite new and often they have not yet been fully operationalized. Consequently, most IEOs do not incorporate gender into all their project

evaluations, and policies are still being developed to decide when gender should be incorporated. Three scenarios can be identified, each with different implications on how the GRE should be designed and how it will be used (Table 2.1).

Table 2.1 Different evaluation scenarios for GRE

Scenario	Focus	Design considerations
<b>1 A one-time evaluation that is not expected to be replicated</b>	<ul style="list-style-type: none"> <li>• An issue of concern to senior management and partners</li> <li>• Often evaluating the gender impacts of a major economic or man-made crisis</li> </ul>	<ul style="list-style-type: none"> <li>• A special evaluation design will often have to be developed</li> <li>• An external consultant might be contracted if the IEO does not have in-house expertise</li> </ul>
<b>2 A pilot GRE to test the value-added of incorporating GRE into the IEO multi year evaluation program</b>	<ul style="list-style-type: none"> <li>• Identifying “quick wins” that can demonstrate the value-added, and the practical feasibility of incorporating gender</li> <li>• Identifying sectors and pilot project evaluations where the GRE design could be replicated in similar projects.</li> <li>• Identifying findings that are directly relevant to operations staff</li> </ul>	<ul style="list-style-type: none"> <li>• Making maximum use of existing data</li> <li>• Identifying gender outcomes that are easy to measure</li> <li>• Minimizing the time burden on operations staff</li> </ul>
<b>3 Gender is already integrated into IEO the multi-year evaluation program</b>	<ul style="list-style-type: none"> <li>• The IEO will prioritize gender as part of its planning cycle (usually 1-3 years)</li> </ul>	<ul style="list-style-type: none"> <li>• The GRE will use one of the standard designs that have been (or are being) developed</li> <li>• Often the IEO will have developed designs with different levels of intensity</li> </ul>

## 2.2 Defining the scope and intensity of a gender-responsive evaluation

Most IEOs do not have the resources or the political support to conduct an intensive analysis of gender issues in all their evaluations. There are three dimensions that determine the scope, depth and complexity of a GRE.

**A. Depth of the analysis.** It is useful to define three levels of depth and intensity of the evaluation.

- **Level 1: *Diagnostic of the current situation.*** Collecting gender disaggregated data for key indicators. If a project has defined gender objectives, gender-responsive data may have been collected, but for projects without defined gender objectives this is often not the case.
- **Level 2: *Pre-test post-test comparison.*** This centers on the differential needs of women and men, and how well these have been met. Have things changed over the life of the project – or over a longer period where there is a focus on sustainability? The evaluation may include a comparison group if this can be constructed economically. For example, this may be achieved by comparing school enrolment or graduation rates between project schools with records from other district schools.
- **Level 3: *Analysis of causality and transformative change.*** This is usually the most intensive level as time and resources may be required to identify and collect data for a counterfactual comparison, as well as to study processes of social control and behavioral change. In addition, this will often require a larger sample size as well.

**B. Defining the boundaries of the evaluation.** How boundaries are defined has a major effect on the types of analysis that are required (see Figure 2.1). Options include:

- Only estimating the direct (primary) gender objectives defined in the project design, on the target population (e.g. increased female employment by a bus company). Often “gender” is understood as only referring to the situation of women. Gender analysis should address the differential project effects on both women and men and on the interaction between the sexes.
- Estimating in addition potential secondary outcomes on the target population, not defined as project objectives. For example, comparing access enjoyed by women groundnut farmers in the project to different stages of the value chain. While each of the three cases focuses on promoting women’s equality, both secondary and tertiary outcomes can also affect men and ideally the evaluation should have this broader focus. Table 2.1 includes examples of possible effects on men that could be assessed.
- Estimating in addition potential tertiary outcomes beyond the target population. This can include individuals or households in the community, neighboring communities or at the regional or national levels (for example impacts of a campaign to raise awareness about violence against women (VAW) on discussion of VAW in the media and among policy makers).

At each level the evaluation should focus on outcomes and not just outputs. This is important

because many evaluations tend to be limited to assessing outputs as these occur within the project implementation time-frame, are easier to measure and are usually defined in the results framework.

**C. Time horizons.** The longer the time horizon the better the understanding of the broader impacts of a project, but the more complex and expensive the evaluation. Time horizons can be classified into:

- The period of project implementation
- The time horizon over which project effects are intended to continue
- Historical antecedents. These may cover the period immediately preceding the project launch (for example, people with political contacts forcing women to sell land at a low price to benefit from rising land values), or it may cover a longer period to assess the influence of earlier projects or important events in the community history.

For projects without defined gender objectives IEO must address two questions: (i) why should gender be addressed? and (ii) how can this be done within budget and time constraints? Some gender advocates in IEOs may feel they should play an advocacy role to promote the inclusion of a gender focus. Appendix 1 summarizes the main arguments for assessing differential effects of projects on women and men. Gender analysis can contribute to enhancing project efficiency and outcomes. It can also be argued that a gender focus is essential to reflect agency objectives and policy goals on gender, equity and human rights along with commitments to UN declarations and other international agreements. The evaluation can both identify ways in which

**Figure 2.1 Examples of primary, secondary and tertiary gender outcomes from the case studies**

### Primary gender outcomes

Gender objectives included in the project results framework:

- Women farmers will receive annual income from groundnuts
- Women become involved in leadership positions in village development committees
- More women are hired by, and achieve career advancement in, the bus company

### Primary gender outcomes

Potential outcomes for the target population that are not included in the results framework. Sometimes they are mentioned in the project documents, often they are not.

- Women's time burden may be reduced through provision of time-saving infrastructure
- Men may be able to work less hours if women are contributing more to the household economy
- Women participants may become more actively involved in management of other village activities
- Women employed by the bus company may increase their control over household resources and decisions

### Tertiary gender outcomes

Potential outcomes that might affect a broader population not involved in the project. This could affect: (a) other households in the project areas, (b) neighboring communities, (c) local, regional or national populations and agencies.

- Women farmers in neighboring communities may increase/improve groundnut production
- Closer cooperation between women and men farmers may increase productivity for the household economy
- Local government agencies may promote the active participation of women in other village projects
- Bus routes may be designed to take into account women's "multi-chaining" travel needs and their security and comfort

women's potential contribution to project objectives is being reduced because of how the project is designed and implemented; by external factors (legal, economic and political systems); by showing that there are many important secondary and tertiary effects beyond direct project objectives that could increase project benefits.

The case studies in Part II illustrate the wide range of potential secondary and tertiary gender outcomes for a project promoting women's participation in groundnut production in Africa, women's participation in the leadership of a village development project in Central Asia and a transport project in Eurasia promoting women's access to employment and advancement

opportunities. For example, some of the potential secondary outcomes for the groundnut project include reducing women's overall time burden (through introducing labor saving devices and providing child-care facilities in project locations), increasing women's expenditure on household necessities and investment in housing, and promoting gender equality in the household and the community. Some of the potential negative secondary outcomes include: men demanding more of women's time to work on their farm, men may control the marketing of women's groundnuts and may retain some of the women's earning, or children may miss school to work on the farm. The case studies also identify indicators and data collection methods for assessing each potential secondary and tertiary outcome.

### 2.3 Strategies for deciding which evaluations should have a more in-depth gender-focus

IEOs have an evaluation portfolio indicating all the different evaluations to be conducted over a one to three year time period. Within this portfolio decisions must be made concerning which evaluations should incorporate a gender focus. There are several criteria to guide the decision as to where GRE should focus.

**A. Institutional intervention points.** All agencies have institutional intervention points in the policy and project cycles where gender could be prioritized (see Box 2.1) and consequently where it might be appropriate to conduct a GRE. These include: agency policy documents, agency-wide progress assessments, and gender-focused policy documents and project designs.

**B. Portfolio analysis and meta-analysis.** Portfolio analysis applies a standard set of rating scales or indicators, which can include gender

to all country programs or kinds of projects to permit a comparison across the agency. This analysis can be used to select a sample of projects that either perform well or badly on gender criteria, for inclusion in the GRE sample.

The IFAD study "What works for gender equality and women's empowerment – a review of practices and results" (2017) is an example of a meta-analysis (evaluation synthesis). The study reviewed a sample of 57 IFAD projects that addressed issues of gender equality and women's empowerment (GEWE) across IFAD projects in agriculture and rural development. The study identified 121 GEWE practices classified into four main groups: practices to improve access to resources, services and opportunities (39 percent); practices to strengthen women and men's awareness, consciousness and confidence (25 percent); practices that address political, legal and

### Box 2.1 Institutional intervention points for identifying priorities for GRE

#### A. Agency policy documents and vision statements

- Managing Director’s Global Policy Agenda
- Agenda 2030 and the SDG agenda
- Agency commitments (e.g. IDA gender commitments)
- Vision documents

#### B. Agency-wide progress assessments

- Mid-term strategy review
- Results framework
- Corporate scorecard

#### C. Gender-focused policies and projects

- Operational plan for gender equality and women’s empowerment
- Policy document on gender and development (for example gender in development strategy)
- Strategic gender initiatives
- Gender mainstreamed projects (that is there are already policies and project designs in place to incorporate gender issues and approaches).

institutional constraints (24 percent); and practices to reduce women’s time poverty (12 percent).

**C. Gender flags and checklists for assessing the treatment of gender.** Many agencies have developed checklists for assessing whether, and how well, gender issues have been addressed in country program frameworks, sector programs or projects<sup>1</sup>. The indicators either use a “Yes/No” format or rate how well the issue was addressed. Box 2.2 illustrates the questions included in the IEG Gender Flag Implementation Completion Report (ICR) evaluation template. Another example from the African Development Bank (AfDB) would be the assessment of gender issues during quality at entry reviews, and the checklists

to mainstream gender in governance and infrastructure interventions. Many other agencies have similar gender checklists. Checklists can either be used for self-assessment by operations staff or for external assessment by IEOs or consultants. It is also important to distinguish between checklists that have been used during project design and those that have been used to assess project implementation and outcomes.

Other possible gender checklist items (suggested by ECG colleagues) include:

- What were the project’s achievements in terms of promoting gender equality and women’s empowerment? This includes assessing whether there are changes in:

1 A number of organizations such as UN Women, USAID, DfID and CARE international have developed checklists for staff to assess how well gender issues have been addressed in the design and implementation of their programs. For example, DFID staff are asked questions such as: “Have we counted all women and girls?”, “Have both women and men been consulted?”, “Have we invested equally in women and men?”, “Do women and girls receive a fair share (of program resources?)”

### Box 2.2 The IEG Implementation Completion Report Review “Gender Flag”

The ICCR “Gender Flag” has two objectives: to document systematically the presence of gender dimensions in individual World Bank projects; and to create incentives to ICR authors to report on gender. The drop-down menu includes five main questions:

1. Is gender a relevant aspect of the project development objective?
2. Does the ICR include sex-disaggregated female or male-specific indicators?
3. Are there indicators that could have been sex-disaggregated and were not?
4. Does the ICR discuss specific gender issues?
5. Please comment on any other issues regarding gender features of the ICR.

Source: Elena Bardasi and Gisela García (2016) What works: Integrating gender into IEG evaluation work. Appendix 3. World Bank

women’s access to resources, assets and services; women’s influence in decision-making; workload distribution among household members; women’s health, skills, income and nutritional levels; and in gender relations within households, groups and communities in the project area; and so forth.

- What percentage of total project resources was invested in activities to promote gender equality and women’s empowerment and how does that compare with other projects?
- To what extent did the project define and monitor sex-disaggregated results to ensure that gender equality and women’s empowerment objectives were being met?
- Was the project implementation structure adequate to support effective

implementation of gender equality and women’s empowerment goals?

**D. Strategic considerations.** It is important to select the evaluations strategically to ensure maximum impact. Some factors to take into consideration include:

- For agencies unfamiliar with GRE, it will be important to seek “quick wins” by identifying evaluations that will be relatively economical and easy to implement, where data can be easily collected, and where it is possible to demonstrate a significant operational impact (value-added)<sup>2</sup>.
- The GRE should select a topic aligned with organizational mandates and donor priorities.
- The evaluation should be easy and economical to replicate.

<sup>2</sup> An example of a quick-win evaluation could be an evaluation of the effects of limited access to rural transport on women farmer’s post-harvest loss. For example, in many countries in Africa women farmers make a major contribution to small-farm agricultural output. They often suffer more than men from lack of transport services to get their product to market (it is easier for male farmers to hitch a lift on a passing truck), and consequently the post-harvest loss for women farmers is often much higher than for men. It is usually relatively simple and cheap to demonstrate this is an evaluation, and often the lost production is a large number. Consequently, the findings of the evaluation can have a big impact. For a first GRE it is also important that this finding is likely to be non-controversial and to have clear operational implications (funding a rural transport project). The evaluation could also be easily and cheaply replicated.

## 2.4 Defining key gender evaluation questions

### A. The importance of a clear definition of the evaluation questions.

Much of the current evaluation literature focuses on evaluation design, while less attention is paid to ensuring that the evaluation will address the questions of priority to stakeholders. There is no single “best” gender evaluation design, and the design must be selected that is best suited to answering the key gender questions. The GRE must be question-driven and not methods-driven (Stern *et al.*, 2012). It is important to ensure that the GRE is designed to respond to the concerns of stakeholders (“question-driven”) and to avoid designs where the kind of questions that can be asked are determined by the prior choice of the evaluation design (“methods-driven”).

### B. Proposed framework for selecting project level GRE questions.

The Africa Groundnuts case study (Chapter 6) illustrates the framework that this reference document recommends for standard GRE evaluation reports. Questions, adapted to the

specific characteristics of each project, should be identified to address each of the six sections of the report framework (Box 2.3).

### C. The importance of broad-based stakeholder consultations:

Who selects the evaluation questions and whose voice counts? A well-designed GRE begins with a stakeholder analysis. This identifies the stakeholders that the evaluation is intended to serve and elicits the gender-related questions that each would like to address. In cases where there are multiple stakeholders, the evaluation team must clarify which questions will be prioritized. Some evaluations use a matrix that lists stakeholders, clarifies their role in the evaluation, and identifies their evaluation questions.

The question “whose voice counts?” is particularly important for gender evaluations. There is often a tendency to prioritize the concerns of senior government officials and funding agencies;

#### Box 2.3 Recommended structure for standard GRE reports

1. **History and context:** What are the historical events or approaches to gender that influenced the design of this project? What are the major issues that the project should address?
2. **Relevance** of the project design for addressing important gender issues
3. **Efficiency:** How efficiently was the project organized to address the gender issues? Were there other approaches that could have achieved the same gender objectives in a more cost-effective way?
4. **Efficacy/effectiveness:** How successful was the project in achieving its short, medium and long-term gender objectives? Were there any serious unintended gender-related outcomes? Could they have been avoided or were they due to factors beyond the project’s control?
5. **Sustainability and resilience:** What evidence is there that the gender outcomes and impacts will be sustained over time? Have potential negative reactions to women’s empowerment (“push-back”) been taken into consideration? Did the project strengthen the ability of women and implementing agencies to identify and address gender-related shocks and stresses?
6. **Gender-related lessons learned?** Did the project include gender-related learning and dissemination mechanisms and were they used effectively? What lessons were learned concerning the selection, design, implementation and sustainability of gender-responsive projects?

yet non-governmental organizations (NGOs), academia and civil society, all of whom have important insights on gender issues might receive less attention. IEOs are sometimes open to the criticism that intended beneficiaries, including women and, even more so groups who may be excluded from the project, are not consulted. The “no-one left behind” focus of the SDGs reflects this concern.

New information technology such as smartphones, crowd-sourcing and social media analysis can make it easier to broaden the range

of stakeholders whose views can be heard (Bamberger, 2017). However, critics argue that information technology (including Big Data) can be used in a top-down way that disempowers women and other community groups.

**D. Generating evaluation questions through the theory of change.** A well-articulated gender-responsive theory of change (TOC) can be a powerful tool for generating evaluation questions, and can identify some of the complex processes of behavioral change, empowerment mechanisms and systems of social control

#### Box 2.4 Examples of widely-used gender indexes and checklists that can be used to generate questions for project GRE

- | **The Africa Gender Equality Index.** The index covers three dimensions: 1) equality in economic opportunities; 2) equality in human development; and 3) equality in law and institutions.
- | **The SDGs.** Sets of indicators have been proposed for each of the 17 SDGs. For many of the goals, more detailed sets of indicators have been proposed by specialized agencies (e.g. UN Women, UN Habitat) even though these are not included in the official list of SDG indicators.
- | **The Gender-related Development Index (GDI).** This adjusts the Human Development Index to take into consideration gender inequalities with respect to life expectancy, education and income.
- | **Gender Empowerment Measure (GEM)** compares relative female and male representation in political and economic power. While it is a useful starting point, it should be used with caution as it is only based on a small number of indicators.
- | **Social Watch’s Gender Equity Index (GEI)** seeks to address the limitations of GDI and GEM by incorporating more indicators on three dimensions: education, participation in the economy and empowerment (measured by the percentage of women in professional, technical, managerial and administrative jobs; and the number of seats women have in parliament and in decision-making ministerial posts).
- | **The World Economic Forum’s Gender Gap Index (GGI).** This combines a range of indicators covering economic participation, economic opportunity, political empowerment, educational attainment, and health and well-being.
- | **Regional indexes.** A large number of regional indexes have been developed that seek to capture the unique characteristics of different regions. For example, the African Women’ Progress Scoreboard (AWPS).
- | **Country-level gender indicators.** Following the launch of the MDGs in 2000, many countries have developed very extensive national data bases, although these vary in their coverage of gender indicators.
- | **Thematic indicators.** Many sector and thematic gender indexes have also been developed by sector agencies such as the United Nations Food and Agriculture Organization (FAO) and the United Nations Education, Scientific and Cultural Organization (UNESCO).

See Appendix 8 for links to the indexes.

that create barriers to women’s empowerment and access to program services. The UKAID/ ACTIONAID theory of change for programs to combat violence against women is a good example of how a gender-theory of change can be articulated and the questions that it helps identify<sup>3</sup>.

**E. Other sources of questions.** The following sources can help generate ideas for gender-responsive evaluation questions:

- Meetings with project agency staff
- Participatory consultations with target populations and civil society
- Review of documents from earlier and similar projects
- Literature reviews and systematic reviews

- Key informants and experts
- Sharing experiences with other IEOs
- If time permits, an initial exploratory or diagnostic study, where an observer spends some time living in or getting close to the community (Salmen, 2017)
- Gender checklists (see following section) can also generate ideas.

**F. Using gender checklists to generate ideas for evaluation questions.**

There are many gender checklists and indexes that can provide ideas for gender evaluation questions (see Box 2.4). Although many of these checklists and indexes were developed for use at the national level to provide a basis of comparison with other countries and regions, they can provide a useful source of ideas for developing questions for project

3 UKAID/ACTIONAID *A theory of change for tackling violence against women and girls*

**Box 2.5 Examples of standard gender checklist questions applied at the project level**

1. The World Bank IEG has a standard set of “gender flag” questions used in the evaluation of different kinds of Bank interventions (IEG, 2016).
2. The Asian Development Bank (AsDB) has checklists of outcome indicators for different sectors such as energy and transport (AsDB 2012, 2013).
3. AfDB includes gender in its checklist for assessing quality at entry.
4. The AsDB and Australian Aid toolkit on gender equality results and indicators includes indicators for 12 different sectors including: education, energy, environment, finance and private sector development, health, humanitarian and disaster preparedness, law and justice, public sector management, rural development, transport, urban development and water supply and sanitation (AsDB, 2013).
5. IFAD uses a six point gender-marker to assess projects at design, during implementation and at completion: gender blind, gender neutral, gender aware, partial gender mainstreaming, gender mainstreaming and gender transformative (IFAD, 2017).
6. The World Bank has used standard gender indicators in portfolio analysis to rate Bank country programs on the relevance of the Bank’s gender policy in the country program, the integration of gender in country policy and projects and the gender-responsive results (Bamberger, Vaessen and Raimondo, 2016 pp. 138-9).

### Box 2.6 A gender-responsive adaptation of the OECD/DAC evaluation criteria

- **Effectiveness:** A measure of the extent to which the intervention achieved its objectives, particularly in terms of the benefits achieved by women and men and without reference to the costs incurred to obtain them.
- **Efficiency:** Analysis of the degree to which gender equality results are achieved at a reasonable cost, whether the benefits have an equivalent cost for women and men, and whether they are allocated equitably.
- **Relevance:** A measure of the extent to which the intervention objectives are adjusted to attend to the different problems and needs of women and men. The criterion also focuses on whether the methodology adopted by the intervention helps women to perceive the limitations imposed on them and to overcome them.
- **Impact:** The contribution of the intervention to a broader policy on gender equality, to the sectoral objectives of equality, and to the advancement toward equality on a long-term basis.
- **Sustainability:** The proportion of the achievement in gender equality that are maintained after the funding period. This is linked to the inclusion of strategic gender needs in the intervention and the ownership by women and men.

Source: Espinosa (2013, p.170) cited in Bamberger, Vaessen and Raimondo (2016, P. 176)

evaluations. Box 2.5 gives examples of gender checklists that were designed for use at the project level.

**G. Developing a gender-responsive version of the OECD/DAC criteria.** Most IEOs incorporate the OECD/DAC evaluation indicators (relevance, effectiveness, efficiency, impact, sustainability) into their evaluation frameworks, although they

often add additional dimensions or perhaps drop one. Many agencies have adapted these indicators to make them gender responsive. Box 2.6 presents a gender-responsive version of the OECD/DAC indicators developed by Espinosa (2013). Many agencies have developed much more detailed gender-responsive versions of the indicators that are adapted to particular sectors or types of agency interventions or evaluations<sup>4</sup>. ■

<sup>4</sup> See, for example, the gender-responsive checklists developed by the World Bank IEG for conducting country program evaluations. The matrix also includes additional dimensions such as results framework, alignment, selectivity, risk to development outcome and institutional development (IEG, 2016).



# Chapter 3

## Designing gender-responsive evaluations at the project level

### 3.1 The gender-responsive approach to evaluation

While many gender evaluators have a preference for particular methods (for example, some prefer qualitative, others prefer quantitative while others always use focus groups), GRE can use all of the standard evaluation tools and techniques. The unique feature of GRE is that, whatever evaluation methods are used, it always incorporates a particular gender-responsive lens:

- Assessing project impacts on women's social and economic empowerment
- Assessing differential impacts on women and men and on the relationships between the sexes
- Assessing outcomes and not just outputs
- Incorporating participatory, consultative approaches that give voice to all actors and particularly vulnerable populations
- Focusing on social inclusion to ensure no groups are excluded or are worse off as a result of the project
- Incorporating issues such as time-use, access to and control of resources, and

gender-based violence not included in conventional impact evaluations

- Mostly using mixed methods to combine the ability to generalize from quantitative methods with the ability to capture lived-experiences and an in-depth understanding of the attitudes and behavior of individuals and groups that qualitative methods can provide
- Focusing on understanding project implementation processes, behavioral change and systems of social control.

Most GRE are based on a gender analysis framework that helps identify the key issues and questions that should be covered. There are many different gender analysis frameworks (Davis and Guevara, 2016; Overholt *et al.*, 1985; UNEG, 2013; UN Women, 2015a; Brisolara *et al.*, 2014; Hesse-Biber, 2012), and Box 3.1 illustrates some of the main elements that are often included. Gender analysis tends to be more operationally and quantitatively focused than feminist evaluation, and at least until recently has been more widely used by development agencies. It is useful to think of a continuum with basic gender analysis at one end (mainly

### Box 3.1 Elements of a gender analysis framework

- A. Context and situation analysis: identifying contextual constraints (legal, political, economic, cultural, attitudes, and so forth) and opportunities. Analysis of historical factors affecting attitudes to the project
- B. Identifying duty-bearers and assessing their capacity to reach out equally to boys and girls, women and men and to promote gender equality
- C. Articulating a gender theory of change:
  - Recognizing that women and men are not homogenous groups. Understanding interactions among sex, age, class, disability, ethnicity and other dimensions. Where data and resources permit, the possibility should be considered of using social exclusion analysis to document these interactions (World Health Organization (WHO) undated)
  - Identifying primary, secondary and tertiary outcomes (see Chapter 2 and 8). This involves going beyond the gender objectives identified in the project design.
- D. Analysis of social control mechanisms, power relations and access to and control of resources at the household, community and other levels.
- E. Operationalizing the TOC through a log-frame and/or results framework
- F. Defining the basic gender outcome indicators
- G. Comparing women and men on key gender-analysis dimensions that go beyond conventional evaluation indicators:
  - Time-use and time poverty
  - Access to and control of productive resources
  - Access to labor markets
  - Access to health, education and other public services
  - Also the extent and forms of violence against women
- H. Defining a framework for assessing gender-empowerment (See Chapter 7)
- I. Assessing the gender indicators in terms of their validity and reliability and the feasibility of data collection
- J. Incorporating gender objectives into the results framework
- K. Developing a gender-responsive M&E system

Source: Adapted from: UNEG (2014) Guidance on integrating human rights and gender equality in evaluation; Overholt *et al.* (1985) Gender roles in development project [applying the Harvard Gender analysis Framework]; and Bamberger (2013). "Engendering Monitoring and Evaluation".

sex-disaggregation of key indicators) and feminist evaluation at the other. In between can be located more intensive gender analysis that incorporates some of the indicators (such as time-use) that are listed in Box 3.1.

GRE face several special challenges in addition to those faced by all evaluations (Appendix 2). These relate to: lack of understanding of the basic concepts about

gender equality and women's empowerment, and the relevance of gender in some "hard" sectors, the additional costs and time of using GRE, data availability, and the use of unfamiliar methodologies. These factors must be taken into consideration when developing strategies to integrate gender into IEO evaluations and when marketing the benefits of GRE both within and outside the IEO.

## 3.2 Designs to collect essential data for gender-responsive evaluations

Table 3.1 identifies eight evaluation designs that can be used to collect all the essential data required for most standard GRE (see Appendix 3 for a description of the designs):

- Desk reviews of project documents, secondary sources and systematic reviews
- Gender-responsive theories of change
- Sex-disaggregation of key indicators
- Key informant interviews
- Focus groups
- Case studies
- Project site visits
- Beneficiary surveys can complement site visits to validate information provided by project officers and partner agencies; as well as qualitative information about gender benefits such as time savings, increased visits to health facilities, or increased income for women.

Table 3.1 also presents the advantages and limitations of each design for GRE and offers some tips on how to avoid common pitfalls and strengthen the methodological rigor of each approach. The table also recommends that a mixed-methods approach be incorporated into all of the designs. This is particularly important for GRE as many of the gender indicators on questions such as, for example, empowerment, behavioral change and domestic violence are difficult to measure, so triangulation of different quantitative

and qualitative measures can significantly strengthen the validity of the data.

It should also be noted that many of the limitations identified in the following section result from not following best practice in, for example, the selection of focus group participants or the analysis of focus groups, in-depth interviews and case studies. Many of these could be addressed by allowing more time and resources for data collection and analysis, and by ensuring rigorous supervision. All the designs are easier to use for projects that include gender objectives than for those that do not, as in the former case much of the required data is already being collected, whereas in the latter special data collection activities will be required. Also, when a project has no defined gender objectives it will often be difficult to make the case as to why gender analysis should be conducted (see discussion above). Some of the gender-responsive limitations of the different designs include:

- Many projects with gender objectives have a relatively narrow definition of gender outcomes so the case will often have to be made as to why data should be collected on secondary and tertiary gender outcomes that are not identified in the project design.
- Many of the methods include potential bias. In some cases, this is because data and opinions are only collected from project staff and government agencies, and many types of informants such as project beneficiaries, vulnerable groups and perhaps civil society are not interviewed. In other cases insufficient time and care may be taken in the

selection of subjects so that some groups (for example women with young children) may be excluded from focus groups while others may participate with a hidden agenda unknown to the evaluator. For example, the evaluator may be unaware that some government officials may attend to discourage participants from criticizing the project.

- In other cases there may be bias in terms of analysis and reporting. A common weakness of case studies is that quotes that may not be representative are selected to prove a point. The same can be true of focus groups. Another common reporting weakness is to use terms like “most respondents felt that ...”, or “many people said that ...” when in both cases there may only have been a few people.
- Another form of bias is that evaluators only get to visit a few successful projects that have been selected by the implementing agency.
- Field work is often conducted under budget and time pressure so that insufficient care is given to design and implementation of the different evaluation methods. This is particularly true for many qualitative methods (key informants, focus groups, project visits, case studies) where the design, implementation and analysis procedures are not as clearly defined as for quantitative methods such as surveys.
- Not all the problems relate to qualitative methods, and a common issue with quantitative surveys is attempting to reduce difficult to measure information such as political participation, behavioral change or sexual harassment with simple quantitative indicators (“how much control do you have over major household decisions?”, how much influence do women have over community decisions on project design?”).

### 3.3 Designs for more in-depth gender-responsive evaluations

Table 3.2 identifies seven designs, many with sub designs, that can be used for more in-depth GRE (see Appendix 3). These include:

- Theory-based methods including: contribution analysis (Mayne, 2011), outcome harvesting (Wilson-Grau and Britt, 2012) and realist evaluation (Pawson, 2013).
- Broader applications of qualitative methods including: diagnostic studies, key informant panel studies, participatory group consultation methods, story-telling and sense-making
- Case-based methods including: exploratory and descriptive cases, and qualitative comparative analysis (QCA) (Byrne and Ragin, 2009)
- Experimental and quasi-experimental designs, including: post-project comparison designs, natural experiments and pipeline designs, and reconstructing baseline data (Shadish, Cook and

Campbell, 2002; Bamberger, Rugh and Mabry, 2012, chapter 11).

- Systems and complexity approaches including: systems mapping, social network analysis, systems dynamics and critical systems heuristics (Williams and Hummelbrunner, 2011)
- Concept mapping (Kane and Trochim, 2007)
- New information technology (ICT and big data) (Meier, 2015; Bamberger, 2017)

In addition to some of the problems identified previously with the essential designs, some of the limitations can include:

- Greater theoretical complexity and the need for a higher level of research experience
- Different kinds of selection and analysis bias
- Many of the practical limitations result from the fact that the studies are often conducted under budget and time constraints, and consequently do not follow all the methodological guidelines concerning, for example, sample selection and data analysis.

Table 3.3 gives examples of how the different methodologies have been applied in GRE.

### 3.4 Evaluability assessment

Once the evaluation methodology has been defined, the data collection methods selected, and indicators defined, it is important to conduct an evaluability assessment. This ensures the proposed design can be implemented within the time and budget constraints, and within the current organizational framework, and that it will be possible to answer the key evaluation questions within these constraints. It is also important to ensure that the evaluation can be implemented within the defined time-frame. One of the common weaknesses of many evaluations is that for administrative and perhaps political considerations, evaluations are often conducted when it is still too early to assess the achievement of the desired outcomes and impacts. For example, when a program to enroll more low-income girls into secondary school has only been operating for perhaps two years, it will not be possible to assess the impacts on poor girls' access to labor markets or raising the

age to marry. However, these will often be project objectives and their assessment will be included in the evaluation terms of reference.

One useful tool to assess the viability of measuring different outcomes is trajectory analysis (Woolcock, 2009). This projects the time-trajectory over which different kinds of outcome are likely to be achieved. Sometimes high and low trajectories will be projected based on different sets of assumptions. The estimated trajectories will often make it clear that, even on the most optimistic assumptions, certain outcomes will not be achieved by the time the evaluation ends. This will show that either the evaluation should be delayed, or that it may only be possible to assess project outputs but not outcomes. The trajectory analysis will often be derived from the project theory of change, which is why it is important to include a time-line in the TOC. ■

**Table 3.1 Approaches used in basic GRE designs: strengths, limitations and tips**

*Note: It is recommended that a mixed-methods approach be integrated into all of the GRE designs. Much of the gender-specific data is difficult to collect and interpret, and the different perspectives and interpretations provided by triangulation are particularly valuable.*

Approach	Method	Strengths	Limitations	Tips
<b>A Desk review of project documents, secondary sources and systematic reviews</b>	<ul style="list-style-type: none"> <li>a. Review of gender objectives and indicators in project design, implementation, and M&amp;E.</li> <li>b. How concepts such as empowerment are used</li> <li>c. Review of reports by other agencies</li> <li>d. Findings and lessons from previous projects</li> </ul>	<ul style="list-style-type: none"> <li>• Ensures the evaluation is addressing project objectives</li> <li>• Detailed information available for projects with a gender objective</li> <li>• Systematic reviews ground the evaluation in what kinds of outcomes can realistically be expected.</li> </ul>	<ul style="list-style-type: none"> <li>• For projects without gender objectives, very little information is available</li> <li>• Gender objectives often only address a narrow range of direct outcomes ignoring secondary and tertiary outcomes</li> <li>• Many sources only focus on women</li> </ul>	<ul style="list-style-type: none"> <li>• Check if project teams have any reports not included in the project files.</li> <li>• Apply the CREATIVE indicators assessment checklist (Box 4.1)</li> <li>• Apply the evaluation design matrix (Appendix 5)</li> </ul>
<b>B Theory of change (TOC)</b>	<ul style="list-style-type: none"> <li>a. A TOC is often not developed during project design, and must be reconstructed by the IEO</li> <li>b. Broaden TOC to model behavioral change, emergence and factors limiting intended gender outcomes (social control)</li> </ul>	<ul style="list-style-type: none"> <li>• Provides a framework for structuring the evaluation (identifying outcomes, key assumptions, intended processes of change)</li> <li>• Can also identify a counterfactual (rival hypothesis) to assess the contribution of observed outcomes</li> </ul>	<ul style="list-style-type: none"> <li>• TOC should be developed with stakeholders but this is time-consuming and difficult to arrange</li> <li>• TOC developed by consultants may not reflect stakeholder perspectives</li> <li>• Rival hypotheses often not defined</li> </ul>	<ul style="list-style-type: none"> <li>• Include a time-line over which sustainable gender transformation should be measured</li> <li>• Define the steps in the processes of empowerment and transformation</li> <li>• Ensure flexibility to model to emergence and backlash resulting from women's empowerment</li> <li>• Anticipate unintended outcomes</li> </ul>
<b>C Sex-disaggregation of key indicators</b>	<ul style="list-style-type: none"> <li>a. Collect available sex-disaggregated data and identify sources when sex-disaggregated data was collected but has not been analyzed</li> </ul>	<ul style="list-style-type: none"> <li>• Quick way to identify sex differences in access to project services or outcomes and identify areas for further investigation</li> <li>• Useful for demonstrating gender differences to operations staff</li> </ul>	<ul style="list-style-type: none"> <li>• Disaggregated data may not be available or expensive to extract</li> <li>• Data on sex may not be accurate or complete.</li> <li>• Meeting attendance data may over-estimate women's level of involvement</li> </ul>	<ul style="list-style-type: none"> <li>• Sex-disaggregated data may not be complete or reliable (See Table 4.1)</li> </ul>
<b>D Focus group discussion (FGD)</b>	<ul style="list-style-type: none"> <li>a. Groups of six to ten people are interviewed together. Usually members share some characteristics (age, sex, participation in the project), but groups can include different perspectives.</li> <li>b. Information is obtained from each participant on every topic.</li> </ul>	<ul style="list-style-type: none"> <li>• Economical way to cover all sectors of the target population</li> <li>• Group interactions create synergy and elicit information that may not surface in individual interviews</li> <li>• Participants may be willing to speak more freely in a group setting</li> </ul>	<ul style="list-style-type: none"> <li>• Due to time pressure member selection may not be carefully controlled</li> <li>• Groups may be dominated by a few influential people or by someone nominated by government</li> <li>• The facilitator may influence the discussion, encouraging people to respond in a particular way</li> </ul>	<ul style="list-style-type: none"> <li>• Important to avoid selection bias or the group being dominated by a few influential participants (see Table 4.1)</li> </ul>

Approach	Method	Strengths	Limitations	Tips	
<b>E</b>	<b>Case studies</b>	a. In-depth study of certain individuals or groups to illustrate or explain survey findings  b. Focus on process, personal experience and behavior.	<ul style="list-style-type: none"> <li>Well-presented cases create greater impact than statistics</li> <li>Help understand different ways people respond to the project</li> </ul>	<ul style="list-style-type: none"> <li>Cases may be “cherry-picked” to find quotes or examples that may not be representative</li> <li>Cases are often selected in an ad-hoc way and do not cover all sectors of the population</li> </ul>	<ul style="list-style-type: none"> <li>Ensure selection, design and analysis are coordinated with other parts of the evaluation.</li> <li>Ensure cases are comparable with other parts of the evaluation</li> <li>For GRE include both women and men from the extended household and, where appropriate community organizations such as the church.</li> </ul>
<b>F</b>	<b>Site visits</b>	a. Short (several hours to one to two day) visits to project locations to see the project in action and to meet with staff, partners and beneficiaries	<ul style="list-style-type: none"> <li>Useful to validate information obtained from project staff and government officials</li> <li>Better understanding of the project reality than written reports which may be incomplete or biased</li> </ul>	<ul style="list-style-type: none"> <li>Implementing agencies may only arrange visits to the best projects</li> <li>Difficult to meet with critics or families with complaints</li> <li>Difficult to meet with women</li> </ul>	<ul style="list-style-type: none"> <li>Avoid, or at least be aware of bias in how communities to be visited are selected by the local agency.</li> <li>Avoid only meeting beneficiaries and project agencies. Try to meet non-project informants and who can give a different perspective (See Table 4.1).</li> <li>NGOs may not be objective informants if they are contracted to implement parts of the project.</li> </ul>
<b>G</b>	<b>Beneficiary and other household surveys</b>	a. Survey covering beneficiaries (and non-beneficiaries) with questions on attitudes and experiences with project  b. Can use structured quantitative surveys or more open qualitative interviews  c. Ideally should use mixed methods	<ul style="list-style-type: none"> <li>Representative sample of the project population</li> <li>Can collect better information than relying on project records and secondary data</li> </ul>	<ul style="list-style-type: none"> <li>Time consuming and expensive</li> <li>Often does not include a comparison group</li> <li>Difficult to collect many kinds of gender information from QUANT surveys</li> </ul>	<ul style="list-style-type: none"> <li>It is often possible to conduct an economical and rapid survey using student teachers, nurses or university students. This is useful to collect basic information on, for example, who knows about the project, who does and does not participate. Women must be interviewed in situations where they can speak openly.</li> </ul>

Table 3.2 Tools for more in-depth GRE

Approach	Method	Strengths	Limitations
<b>A. Theory-based methods</b>			
• Contribution analysis	a. Develop “program story” and collect evidence to support and challenge	<ul style="list-style-type: none"> <li>• Can estimate contribution of project to outcomes when attribution analysis not possible</li> </ul>	<ul style="list-style-type: none"> <li>• Often rival hypotheses are not identified and tested</li> </ul>
• Outcome harvesting	a. At end of project beneficiaries and other stakeholders asked to identify most important outcomes during project	<ul style="list-style-type: none"> <li>• Participatory approach capturing beneficiary perspectives</li> <li>• Larger numbers of outcomes captured providing multiple perspectives</li> </ul>	<ul style="list-style-type: none"> <li>• May only capture positive outcomes</li> </ul>
• Realist evaluation	a. Asks what works, for whom, in what ways, to what extent, in what contexts and how? b. Focus on reasoning of the actors involved. How do they respond to interventions?	<ul style="list-style-type: none"> <li>• Provides much broader understanding than just asking “did it work?”</li> <li>• Tries to explain processes of reasoning and behavioral change</li> <li>• Looks upon actors as proactive and not just reactive</li> </ul>	
<b>B. Broader applications of qualitative methods</b>			
• Initial diagnostic studies	a. Spending time in the community or project area prior to project launch	<ul style="list-style-type: none"> <li>• Understanding the beneficiary perspective and cultural dynamics</li> <li>• Detecting issues the project design may overlook</li> </ul>	<ul style="list-style-type: none"> <li>• Difficult to commission studies before official project launch</li> <li>• Needs ethnographic training</li> </ul>
• Key informant panel studies	a. Researcher develops friendship with different types of individuals who are visited periodically to obtain updates on what the community thinks about the project, what they are hearing, and what is happening	<ul style="list-style-type: none"> <li>• Independent feedback to avoid only getting information from project staff</li> <li>• Identifies unintended outcomes</li> <li>• Identifies vulnerable groups and those excluded</li> </ul>	<ul style="list-style-type: none"> <li>• Risk of bias if researcher mainly develops contacts with particular kinds of people and does not develop relationships with others</li> </ul>
• Participatory group consultation methods	a. PRA, Most Significant Change and other group consultation techniques used to develop social maps, historical timelines, power analysis and perceptions of causality	<ul style="list-style-type: none"> <li>• Visual and mapping methods work well with groups with low literacy</li> <li>• Participatory methods give voice to vulnerable groups, including women</li> </ul>	<ul style="list-style-type: none"> <li>• Can be manipulated by researcher (intentionally or unintentionally)</li> <li>• Often used to get quick community feedback without due attention to the methodology</li> </ul>
• Story-telling and sense-making	a. Individuals narrate short stories about events in the community. May focus on project or be open. b. Sense-making software used to analyze the stories	<ul style="list-style-type: none"> <li>• Gives voice to beneficiaries and counterbalance to funders evaluation criteria</li> <li>• Broadens the focus</li> </ul>	<ul style="list-style-type: none"> <li>• Many stories are brief and superficial</li> <li>• Only positive stories</li> <li>• Need to control for bias introduced by facilitator</li> </ul>
<b>C. Case-based methods</b>			
• Exploratory and descriptive	a. Describe projects, processes and participant attitudes b. Illustrate findings of QUANT studies	<ul style="list-style-type: none"> <li>• Puts flesh on the survey statistics</li> <li>• Compares lived experience of women and men</li> <li>• Explains context</li> </ul>	<ul style="list-style-type: none"> <li>• Cases often used for advocacy rather than objective reporting</li> <li>• Cases often not representative</li> <li>• Analysis often superficial</li> </ul>
• Analytical (QCA)	a. Matrix created with attributes of subjects and outcome b. Identifies configuration of attributes needed to achieve intended outcome c. Uses mixed methods	<ul style="list-style-type: none"> <li>• Used with small samples</li> <li>• Addresses complexity</li> <li>• Permits attribution analysis</li> <li>• Can combine with other kinds of evaluation</li> </ul>	<ul style="list-style-type: none"> <li>• Only permits small number of attributes in analysis</li> <li>• Attributes must be dichotomous (Yes/No)</li> </ul>

Approach	Method	Strengths	Limitations
<b>D. Experimental and quasi-experimental designs</b>			
• Post-project comparison designs	a. Sample survey comparing project and comparison group b. Can use propensity score matching to strengthen estimates	• Provides estimate of project impact • Can combine with mixed methods to strengthen estimates	• Risk of selection bias as there is no pre-test data
• Natural experiments and pipeline designs	a. Delays due to natural or administrative factors used to compare projects with areas where project has not yet started	• Can provide estimate of project impact which may have bias, but which can be strengthened through mixed methods	• Selection bias • Needs agile evaluation team to be able to detect areas where delays are occurring
• Reconstructing baseline data	a. Baseline can be “reconstructed” using secondary data, key informants, recall and PRA	• Valuable tool to strengthen evaluation designs	• Sources have potential bias • Many data sources do not have reliable gender data
<b>E. Systems and complexity approaches</b>			
• Systems mapping	a. Visual representation of system within which project operates b. Identifies linkages among components and external factors	• Visualizes interactions among all elements of systems • Helps describe systems of social control that limit gender outcomes	• Difficult to quantify and analyze processes of change
• Social network analysis	a. Analysis and mapping of processes of communication, influence and power within an organization or community	• Can compare women and men’s communication networks, positions in power structures and social capital	• Requires fairly large sample and sophisticated data analysis
• System dynamics	a. A map representing stocks and flows among project/systems components. Estimates how complex systems respond to project interventions	• Helps assess the effectiveness of different project interventions on different parts of a system • Can be used to identify factors limiting the effectiveness of project interventions	• May require more complex measurement and analysis
• Critical systems heuristics	a. Analysis of the factors that determine what issues lie within the boundaries of the evaluation b. Studies how values affect the scope and focus of an evaluation	• Helps understand how feminist values can be incorporated into an evaluation	• May appear very abstract and theoretical to clients • Many clients believe evaluations should be value free, so difficult to accept premises of this approach
<b>F. Concept mapping</b>	a. Experts or stakeholders identify key outcome indicators, which are converted into rating scales. b. Scales can assess changes over project life or compare project and comparison groups at end of project	• Helps develop broad-based indicators of project performance • When used on-line can provide economical way to measure project outcomes • Can involve wide range of gender specialists in indicator development	• Requires high level of specialist input and can be difficult to coordinate
<b>G. New information technology</b>	a. ICT devices used to generate feedback on project performance b. Social media analytics provides feedback on attitudes to projects c. Big data can generate wide range of data on contextual factors not previously available d. Can provide multiple indicators on issues such as poverty, conflict and forced migration	• Phones and social media give voice to women • Rapid and economical ways to collect gender-related information • Can involve difficult to reach groups • Data analytics can provide more powerful prediction tools and increase statistical power of the analysis • Real time social media analysis (e.g. twitter) can identify potential gender conflicts and problems	• Many IEOs have limited experience with big data and data analytics • Many evaluators are suspicious of big data • Need for bridge-building between evaluators and big data analysts

Table 3.3 Examples of the application of different approaches to GRE

Design	Variations	Examples/references
1 <b>Experimental and quasi-experimental</b>	RCTs, quasi-experimental designs, natural experiments	<ul style="list-style-type: none"> <li>Using RCT to evaluate the impacts of training of cross-border guards in Rwanda to reduce violence against women and improve socio-economic outcomes for women (Source: World Bank Gender Innovation Lab).</li> <li>Many of the RCTs conducted by the Poverty Action Lab assess the impact of development interventions on women (<a href="http://www.povertyactionlab.org">www.povertyactionlab.org</a>)</li> </ul>
2 <b>Statistical</b>	Econometrics, public expenditure incidence analysis, public expenditure tracking.	<ul style="list-style-type: none"> <li>Public expenditure incidence analysis used to assess what proportion of public expenditures in sectors such as health and education go to low income families including female-headed (Source: Davoodi <i>et al.</i>, 2003)</li> </ul>
3 <b>Theory-based</b>	Theory of change, process tracing, contribution analysis, realist evaluation	<ul style="list-style-type: none"> <li>Using theory of change and contribution analysis to assess the effectiveness of a 10 year OXFAM program to reduce violence against women in El Salvador (Source: Davis and Guevara, 2016)</li> </ul>
4 <b>Case-based</b>	Naturalistic, grounded theory, ethnography, process tracing, QCA, within-case analysis, simulations, network analysis	<ul style="list-style-type: none"> <li>QCA used to assess the effectiveness of UN Women interventions at the national level on women's economic empowerment. The country was used as the unit of analysis (Source: UN Women, 2016)</li> </ul>
5 <b>Participatory and qualitative</b>	Empowerment evaluation, feminist evaluation, PRA, Most Significant Change, Outcome Harvesting, Outcome Mapping	<ul style="list-style-type: none"> <li>Village women design a survey instrument to identify family needs in poor communities in India and then interpret and disseminate the findings. (Source: World Bank Social Observatory, India)</li> </ul>
6 <b>Review and synthesis</b>	Meta-analysis, narrative synthesis, realist synthesis	<ul style="list-style-type: none"> <li>Using a systematic review, covering all of the published literature, to assess the impacts of micro-credit on women's economic empowerment. (Source: Vaessen, Rivas and Leeuw, 2016)</li> </ul>
7 <b>New information technology</b>	Twitter and social media analysis, satellite images, ATM transactions, phone records, analysis audio and video images	<ul style="list-style-type: none"> <li>Tracking trends in gender-based hostility in factories in Indonesia (UN Global Pulse)</li> <li>Tracking effectiveness of on-line messaging to promote girls' empowerment</li> </ul>

Sources: Adapted from Stern *et al.* (2012), Bamberger, Vaessen and Raimondo (2016)



# Chapter 4

## Tools for collecting data for gender-responsive evaluations

### 4.1 Defining evaluation questions and constructing a design matrix

Ways to identify the key gender-related evaluation questions have been presented above. Once the questions have been identified, the next step is to construct a design matrix that:

- Lists the evaluation questions
- Lists the evaluation designs
- Lists the indicators required to measure each question
- Identifies the data sources for each question
- Identifies potential issues affecting data collection and validity

Appendix 5 gives an example of how the design matrix could be applied to two of the questions identified in a village development project: (1) did the project address women’s needs? and (2) how did the project affects women’s empowerment? The indicators included in the table are illustrative and there are other designs and indicators that could be used. The design matrix has several advantages. First, it ensures that all of the indicators required to measure each key evaluation question are organized in one

place. Second, it requires the evaluation team to ensure the data sources are identified to measure each indicator. Third, it requires the evaluator to check the feasibility of collecting all the data. In practice this is important as due to time pressures during the planning mission, the evaluation team will often rely on assurances from a particular agency that the required data is easily available, that it provides the required information and that the information is complete, reliable and valid. It is often discovered at a later point that there are problems with the data: it may not be easily available, it may be incomplete, it may not be providing the required information, or the data

#### Box 4.1 Even basic sex-disaggregated household data may be difficult to collect if only male “household heads” are interviewed

The author was involved in pilot-testing a household survey in Mombasa. The husband was being interviewed and was asked to list all of the household members. He only listed male children. However, we had observed some little girls peeping through the doorway, so we asked if he also had some daughters. “Oh, do you want me to mention my daughters as well” he asked, seeming quite surprised.

does not cover all the target population (see Box 4.1).

The CREATIVE Indicator Assessment Checklist in Box 4.2 is a useful tool for the assessment of

each proposed indicator in the data set. The checklist can be used as part of the Evaluability Assessment that should be conducted to assess the overall validity of the proposed evaluation design.

## 4.2 Methods of data collection

This section describes the data collection methods most commonly used for each of the eight basic GRE designs discussed in Chapter 3. This discussion recommended that a mixed methods approach should be incorporated into all the data collection activities. In addition to strengthening validity through triangulation of different data collection methods, mixed methods are particularly important for GRE to:

- Study processes of empowerment and behavioral change that are difficult to capture with a single data collection method;
- Strengthen generalizability of in-depth qualitative analysis (i.e. to ensure the sample of respondents is representative of the total sample population);

### Box 4.2 The CREATIVE Indicator Assessment Checklist

A good set of indicators should have the following characteristics:

- **Consistent:** the same responses are obtained by different data collectors
- **Relevant:** the indicators relate directly to the topic studied. This is particularly important when using proxy indicators.
- **Economical:** low cost for collection and analysis
- **Accessible:** the data is easy to collect and to organize in digital format. Also important to confirm that the owner of the data is willing to make it available
- **Time-bound:** The indicator relates to a clearly defined time-frame. Ideally the indicator will also provide comparable information at different points in time
- **Inter-agency comparability:** different agencies will generate comparable information
- **Valid – high construct validity:** The indicator actually describes and measures the intended construct. Sometimes several indicators will be required to measure a complex construct (e.g. vulnerability, empowerment), in which case it will be necessary to assess the validity of the set of indicators
- **Ethical:** the collection, dissemination and use of information follows accepted codes of ethics

Source: the author

- Identify and give voice to vulnerable groups.

For each approach the table summarizes the data collection methods and provides tips on how to avoid common weaknesses and strengthen the validity of each method. Tips include:

- Ensure the mixed methods designs use an integrated approach so that quantitative and qualitative samples are comparable. Often the quantitative and qualitative parts of a study are conducted independently and with little coordination;
- Many projects that do have gender objectives use a narrow definition of gender outcomes and it is important to identify potentially broader secondary and tertiary outcomes that could be considered for inclusion in the evaluation;
- Ensure that participatory approaches are used to give voice to all sectors of the population, including vulnerable groups. Also ensure women are interviewed in a context where they can speak freely;
- It is important to measure women's level of participation in project and community activities and not just record whether they were present;
- Be aware that even the most basic sex-disaggregated data may not be accurate or complete (Box 4.1). Some agencies may have incentives to overstate the number of women, or the sex-ratio may not be recorded accurately;
- Avoid biases in the selection of informants, communities, and focus group subjects. Local agencies may have incentives to only expose researchers to people or communities favorable to the project;
- Researchers should try to develop relations of confidence with informants from different groups to avoid relying exclusively on information from project staff or community leaders. Similarly, it is important to include women informants who can speak freely and not rely only on male family or community members;
- Gender and empowerment analysis should not just focus on the household but should also try to include the extended household and religious or other groups in the social control network<sup>1</sup> that can constrain the processes of empowerment the project seeks to promote;
- Ensure that data, for example from case studies or focus groups, is selected and presented objectively. Often report writers "cherry-pick" the most dramatic responses or those that support their own opinions;
- Field studies are often conducted under time and resource pressures so that insufficient time is allowed for careful preparation and selection of respondents.

<sup>1</sup> A social control network refers to all of the individuals, groups or institutions that can influence, positively or negatively, a woman's ability to benefit from the opportunities offered by the project to strengthen her economic and social empowerment. These may include members of the extended household, neighbors and friends, social and religious organizations, schools and places of employment. It can also include formal institutions such as the police, public service providers and legal and administrative agencies.

### 4.3 Developing and using gender responsive checklists, indexes and indicators

Some of the available gender checklists, indexes and indicators and how they help identify, formulate and measure gender questions were reviewed above. This section discusses some of the issues involved in the development of these important measures.

#### A. Choosing between standard sets of indicators and project-specific indicators.

There is a trade-off between using standard and customized indicators. While standard indicators permit comparisons across projects, sectors or countries, customized indicators have the flexibility to adapt to the specific cultural and political context of a project as well as to any unique project design features. Often the two types of indicators will be combined.

#### B. Indicators to measure gender equality.

Gender equality indicators are often used to assess how far a country, agency or project is from achieving complete gender equality of access to services and benefits, participation in projects, or involvement in decision-making. Box 2.4 lists some of the widely-used indicators for measuring gender equality. Most of these indicators are used at the country level (for example to compare men and women's representation in parliament, use of cellphones, school enrolment rates). However, in countries where disaggregated data is available, it may be possible to use the indicators at the regional or

local level.

There are also indicators that can be used at the local level. IEG (2016) presents a useful list of indicators classified into women's economic empowerment and women's political empowerment. They also explain that gender equality indicators can be used as a metric to assess how far a project or department is from achieving gender equality. Other indicators can be used to: compare men and women's time-use or access to and control of productive resources (e.g. Harvard Gender Analysis Framework) (see Overholt *et al.*, 1985, and Rao *et al.*, 1991), and to assess internal organizational change with respect to gender dynamics (UNDP's Gender Mainstreaming Scorecard<sup>2</sup>, DFID's Gender Sensitive Indicators<sup>3</sup> and the ADB/AUSAID Toolkit on Gender Equality Results and Indicators<sup>4</sup>).

#### C. Using indicators to identify beneficiaries, affected populations and groups who are excluded.

A key challenge for many evaluations is to identify sectors of the target population that are excluded from the project or from access to project benefits. Sometimes this involves a dichotomous included/excluded classification but for multi-component projects there may be a scale to assess the relative level of inclusion or exclusion. One example is a scale used to identify groups falling below the poverty line on the basis of a set of indicators. For example, the Grameen Foundation Progress Out of Poverty Index<sup>5</sup>

2 Some of the gender sensitive organizational indicators used by UNDP include: corporate commitments, implementation mechanisms, internal capacities, gender mainstreaming in the project cycle, accountability mechanisms and organizational culture

3 Some of DFID's gender sensitive indicators include: the impact/effectiveness of activities designed to promote access for women and men; the impact/effectiveness of targeted activities; the impact/effectiveness of activities designed to develop gender-awareness and skills amongst policy-makers, managers and implementation staff; the impact/effectiveness to promote gender equality among staff (DFID. The Gender Manual).

4 <https://www.adb.org/sites/default/files/institutional-document/34063/files/tool-kit-gender-equality-results-indicators.pdf>

5 <http://www.progressoutofpoverty.org/about-ppi>

estimates community poverty levels based on ten locally adapted indicators.

**D. Adapting indicators to local conditions and local social norms.** Many indicators, particularly qualitative indicators, must often be adapted to local conditions and social norms. Indicators used to measure poverty, empowerment or forms of sexual harassment are examples where there is a need to adapt the indicators.

Many indicators of women's empowerment are defined by international researchers and may not be directly appropriate for local conditions. Hashemi and Schuler (1996) found that many studies on the impacts of microcredit on women's empowerment in Bangladesh used indicators such as women's control over household budgets that local women considered inappropriate as an indicator as they had no expectation (at least in the immediate future) of completely controlling the household budget. When consulted they had much more modest goals such as being allowed to leave the compound and visit the village, or to attend meetings to learn about the creation of a village bank. Developing these culturally appropriate empowerment indicators was very time consuming as trust had to be established before beginning to discuss specific empowerment indicators.

**E. Gender-responsive performance indicators. Most agencies have defined a set of indicators to assess project performance.** These usually build on the OECD/DAC dimensions of: relevance, efficacy, efficiency, effectiveness, sustainability and (sometimes) impact. Many agencies have expanded the list to include, for example: the adequacy of monitoring and evaluation (M&E) systems and developing learning mechanisms, addressing equity, borrower performance (government and implementing agency), and Bank performance. The assessment of gender performance is usually based on a sub-set of the

OECD/DAC indicators that are adapted to address gender (Box 2.7).

For most IEOs the standard, easy-to-use indicators do not provide an in-depth set of gender-responsive indicators. One of the potential challenges with a standard set of indicators is that they may reduce staff incentives to innovate and customize indicators for the specific sector and country context. It is likely that, as gender equality becomes internalized as a development objective, gender indicators will begin to draw more heavily on the wide range of gender-responsive indicators available at national and local levels. Some of the dimensions that should be used at the project level include:

- Gender equality in access to services, education, access to communications (particularly cellphones and internet) and to political participation;
- Gender-based violence and sexual harassment;
- Time-use;
- Access to, and control of economic resources;
- Participation in project selection, design and implementation;
- Leadership in community organizations;
- Empowerment indexes;
- Participant perceptions of changes that have occurred and constraints to achieving these changes;
- Women's mobility (to travel within and outside the community).

**F. Indicator assessment checklist.** It is important to assess indicators to ensure they provide valid measures, that they are methodologically sound, easy to measure and

economical to collect. Box 4.2 presents the CREATIVE indicator assessment checklist as one possible option. There are several similar indicator assessment checklists<sup>6</sup>.

## 4.4 Integrating data collection for gender evaluations into ongoing surveys

It is sometimes possible to reduce the cost and time of data collection by coordinating with ongoing or planned surveys conducted by other agencies. These piggy-backing approaches can include:

- Negotiating the inclusion of additional gender-related questions into an ongoing survey;
- Negotiating with statistical offices and other survey agencies to include a set of standard gender questions into all surveys;
- Incorporating gender modules to be applied to a sub-sample of respondents (e.g. wives and partners, male and female high-school and college students, the elderly) in the main survey. Usually the

module will be applied to individuals in the sampled households, but it is also possible to combine several respondents and use focus groups;

- Gender-focused analysis of completed or ongoing evaluations;
- Using survey analysis to develop a typology that can be used to select a sample for gender-focused case studies or in-depth interviews.

With all of these approaches it is important to ensure that the main survey sample covers all of the desired target population, that it is collecting the right kinds of information and that it creates a relationship with respondents that will be conducive for the administration of the follow-up gender module. ■

6 Other indicator assessment checklists include SMART (Specific, Measurable, Attainable, Relevant and Time-bound). Useful references: Save the Children, <https://sites.google.com/site/savethechildrendme/Home/smart-indicators>; and The Road to Results: Designing and Conducting Effective Development Evaluations (Morra-Imas and Rist, 2009).

Table 4.1 Data collection methods used in basic GRE designs: tips and examples

Approach	Data collection methods	Tips
<b>A Mixed methods:</b> Recommended for use in all GRE	<ol style="list-style-type: none"> <li>Identify all of the main QUANT and QUAL methods to be used in the evaluation and the information to be collected by each</li> <li>Ensure that at least two independent measures will be used for each key indicator (usually one QUANT and one QUAL)</li> <li>Triangulate findings to check for consistency and to obtain broader understanding</li> </ol>	<ul style="list-style-type: none"> <li>Ensure that QUANT and QUAL data collection are integrated in the evaluation design and that both are generated from the sample sampling frame. Often the QUANT and QUAL teams work independently so that it is difficult to know if the two sets of data can be compared</li> <li>Triangulation is a critical tool and should be used systematically. Particularly useful for GRE to compare what women say about their behavior and what is actually observed</li> </ul>
<b>A Desk reviews</b>	<ol style="list-style-type: none"> <li>Review of all project documents and identification of (i) gender objectives, (ii) gender implementation strategies, (iii) how gender is treated in the TOC, (iv) discussion of gender issues, (v) identification of potential secondary and tertiary gender outcomes</li> <li>Review of secondary data, reports and surveys. Identify surveys or reports covering project areas that might provide sex-disaggregated data. Check if there are any surveys including sex-disaggregated data that have not been published</li> </ol>	<ul style="list-style-type: none"> <li>Check if project teams have any studies or reports not included in the project file</li> </ul>
<b>B Theory of change (TOC)</b>	<ol style="list-style-type: none"> <li>Update any TOC created during project design, or if this does not exist, reconstruct the TOC retrospectively</li> <li>Ensure that conventional TOC are expanded to include gender-related dimensions</li> <li>Ensure that key assumptions about the processes of gender transformation and empowerment are identified and tested</li> <li>Use appropriate mixed method tools to collect and update data</li> </ol>	<ul style="list-style-type: none"> <li>Ensure the TOC includes a time-line over which changes should occur. This should indicate the time trajectory over which sustainable gender transformation should be measured</li> <li>Ensure the steps in the processes of empowerment and transformation are clearly articulated</li> <li>Ensure sufficient flexibility in the TOC to adapt to emergence and backlash resulting from women's empowerment</li> <li>Anticipate potential causes and manifestations of unintended outcomes</li> </ul>
<b>C Sex-disaggregation of key indicators</b>	<ol style="list-style-type: none"> <li>Often sex-disaggregated data has been collected but not published, so check sources to see if disaggregated data is available and if it is feasible to analyze</li> </ol>	<ul style="list-style-type: none"> <li>Check the reliability of sex-disaggregated project data, as staff may over-estimate the number of women if they are pressured by donors to promote women's participation</li> <li>Try to include measures of women's level of participation as sometimes they may attend meetings but not speak or vote</li> <li>Try to obtain estimates of the types and numbers of women (and men) excluded from the project</li> </ul>
<b>D Key informant interviews</b>	<ol style="list-style-type: none"> <li>Structured or semi-structured interviews are conducted with people knowledgeable about the project</li> <li>The sample should cover all types of informants and should avoid bias (for example, not including critics of the project)</li> </ol>	<ul style="list-style-type: none"> <li>Try to develop relations of confidence with a few informants not involved with the project who can provide an independent (or at least different) perspective on project activities and what people are saying</li> <li>Ensure informants include a broad spectrum of women and not just women leaders and those in positions of power</li> </ul>
<b>E Focus group discussions (FGD)</b>	<ol style="list-style-type: none"> <li>Groups of 6-10 people are assembled and interviewed together. Usually all members share some characteristics (age, sex, economic level, participation in the project or not), but sometimes groups may be selected to include different perspectives</li> <li>Information must be obtained from each participant on each topic</li> <li>Ideally the FGD is conducted by one facilitator and one person observing and taking notes</li> </ol>	<ul style="list-style-type: none"> <li>Try to avoid selection bias. Provide precise guidance on the different groups who should be invited to the FGDs. Select a person of confidence (ideally from the community), who can report back on which groups did and did not attend and can also advise on whether any people were attending to check on the group or to promote the government or a political party perspective (often the evaluator is not aware of this)</li> <li>Avoid the discussion being dominated by one or two people who are often better educated or have influence</li> <li>Avoid bias being introduced (unintentionally) by the facilitator</li> </ul>

Approach	Data collection methods	Tips	
		<ul style="list-style-type: none"> <li>• Ensure that all participants give their views on each question</li> <li>• If certain groups do not attend (for example women with small children, younger women, or women from a particular ethnic group or who have particular kind of employment), try to arrange time for quick follow-up individual visits to some of these people. People who do not attend are often those who it is most important to talk with as they are excluded from the project</li> </ul>	
<b>F</b>	<b>Case studies</b>		
	<ul style="list-style-type: none"> <li>• <b>Descriptive case studies</b></li> </ul>	<ol style="list-style-type: none"> <li>a. Develop typology of respondents (types of farm, school enrolment, access to health services and so forth) and select cases to illustrate each category</li> <li>b. Prepare checklist of questions to be addressed</li> <li>c. Combine some of following methods:               <ul style="list-style-type: none"> <li>• Unstructured/semi-structured interviews</li> <li>• Observation checklist</li> <li>• Observation of subject in different settings: home, project, community, and so forth</li> </ul> </li> <li>d. Consider using audio/video recording</li> <li>e. Request respondents to keep diaries of, for example: time use, travel, expenditures</li> </ol>	<ul style="list-style-type: none"> <li>• Ensure the selection, questions asked and analysis of the case studies are fully coordinated with the other parts of the evaluation</li> <li>• Ensure there are clear criteria for the selection of cases and that the sample is broadly representative of the survey sample (if one is used)</li> <li>• For gender-focused case studies, make sure to interview both women and men from the extended family and, where appropriate community organizations such as the church. The empowerment goals of a project are often constrained by mechanisms of social control, often from people who do not live in the household</li> </ul>
	<ul style="list-style-type: none"> <li>• <b>Qualitative, comparative case studies (QCA)</b></li> </ul>	<ol style="list-style-type: none"> <li>a. Select cases</li> <li>b. Prepare matrix with attributes of subject and outcome -usually in binary form (1 and 0)</li> <li>c. Complete matrix for each subject</li> <li>d. Analysis to identify configurations associated with positive outcome</li> </ol>	<ul style="list-style-type: none"> <li>• Ensure that the attributes of the household or organization (school, village bank, and so forth) also include relevant QUAL indicators (for example the degree of control that women have over household or community decisions) and not just variables like age and education</li> </ul>
<b>G</b>	<b>Site visits</b>		
	<ol style="list-style-type: none"> <li>a. Meetings with government, project staff and key informants</li> <li>b. Meetings with different groups in the project</li> <li>c. Observations of the conditions of project infrastructure and service delivery</li> <li>d. Meetings with NGOs and other organizations involved in the project</li> <li>e. Meetings with both beneficiaries and members of the target population not involved in the project</li> </ol>	<ul style="list-style-type: none"> <li>• Try to avoid, or at least be aware of bias in how communities to be visited are selected. Agencies like to showcase their best projects and avoid visits to problem projects. Coordinate ahead of time to clarify the visit criteria</li> <li>• Also try to avoid bias with respect to the people and groups visited</li> <li>• Find ways to meet with women or men who are not participating in the project. This is critical in male-dominated cultures where some men may prohibit their spouse from participating and may even use violence if she does not “obey”. Often there are informants like the district nurse who knows most families and can provide this information. The local police chief may be a useful informant if the project has generated domestic violence</li> <li>• Remember that NGOs and civil society may not be objective informants if they are contracted to implement parts of the project</li> </ul>	
<b>H</b>	<b>Household surveys</b>		
	<ol style="list-style-type: none"> <li>a. These may be used as part of a more in-depth evaluation, or they may involve a rapid survey to provide background data for a qualitatively focused evaluation</li> <li>b. The survey may also be used to develop a typology (for example, types of women’s employment, or women using different modes of transport to get to work) that can be used to select case studies</li> </ol>	<ul style="list-style-type: none"> <li>• Even if resources do not permit a formal survey, it is often possible to conduct an economical and rapid survey using student teachers, nurses or university students. This can be a useful way to collect basic information on, for example, who is familiar with the project, who participates and who does not. A very short survey can often be conducted very rapidly and cheaply, while a longer survey can become very time consuming and expensive</li> <li>• When interviewing women, it is essential that they are interviewed in a situation where they can speak openly</li> </ul>	



# Chapter 5

## Managing gender evaluations

### 5.1 The role of the evaluation manager in gender-responsive evaluations

In many IEOs, the manager or team leader of an evaluation with a gender component is often not a GRE specialist and may not even be a gender advocate. However, s/he will play an important role in ensuring the successful design, implementation, analysis and use of the GREs. GREs may require an advocate as well as a strategist who can ensure the

design of the GRE is aligned with project and organizational objectives. One of the important tasks will be to draw on the gender expertise within the IEO as well as the agency's gender office (or gender focal point) to become as conversant as possible with gender issues and the agency's experience, approaches and policies relating to gender.

### 5.2 The role of the manager at each stage of a gender-responsive evaluation

**Promoting GRE.** The manager must be an advocate, helping to convince the IEO of the importance of gender and the value-added that a GRE focus brings to any evaluation. The manager may also have a role mobilizing funds to cover the additional costs of GRE from IEO, other parts of the agency and from outside funding sources (e.g. gender trust funds). The manager also helps select strategically the first GREs to achieve some “quick wins” that demonstrate the value-added of a gender focus in the evaluations. It is important to select the first evaluations in countries or sectors where the findings will appeal to technically oriented sector specialists and economists, and where the approach could be easily replicated in other

project evaluations. Alliance building is another important responsibility to build up a network of allies and technical resources within and outside the agency.

**Planning and team building.** GRE will often require building new areas of expertise into the evaluation team. This can present challenges as established team members may not be familiar with some of the new GRE methodologies and research approaches. In some cases, this can be disruptive if, for example, established quantitative researchers feel that some of the new qualitative methods are less rigorous and “professional”. These considerations may require more time for team-building, allowing team members to understand

each other's approaches and to build trust.

When using mixed methods approaches the manager must ensure that team members from disciplines with which the evaluation team has not previously worked (for example, feminist research and ethnography) are brought in from the start of the planning. This is important because there is a tendency for GRE to be planned by the core team and to only bring in gender specialists and other new disciplines at the last moment when it is too late for them to contribute to the overall evaluation design.

### **Defining the evaluation objectives and key questions.**

As discussed earlier, many organizations have a narrow definition of gender objectives, and many important secondary and tertiary outcomes are not taken into consideration. The manager may have to negotiate with operations staff and other stakeholders to broaden the range of issues included in the GRE. It is also important to understand the gender-related questions of concern to different stakeholders. This may be more difficult than for conventional sector evaluations as many operations staff may not fully understand some of the gender-issues that may arise. For example, the goal of ensuring that women farmers have equal access to all stages of the value-chain may raise many issues that had not appeared when mainly working with male farmers (for example, the provision of basic child-care facilities for mothers who must bring a child with them to work, or issues of sexual harassment).

Another challenge is to ensure the most appropriate evaluation methods are selected to address the stakeholder questions. Many gender specialists, like other evaluators, often have their preferred evaluation methods (e.g. participatory consultations, case studies, focus groups, time-

use analysis), and it is the responsibility of the manager to ensure that the evaluation is demand driven (selecting methods to answer the key questions) and not methods driven (only focusing on questions that can be answered with the preferred evaluation tools).

**Deciding if a GRE is required and the appropriate level of intensity.** At least two sets of factors will determine whether GRE should be incorporated into the evaluation. First, are there important gender issues inherent in the project? Second, if gender can only be addressed in a proportion of evaluations, how does this evaluation compare with others in the portfolio in terms of its gender priority? If gender is to be addressed, then the manager should work with IEO management to agree on the level of intensity of treatment of gender.

**Identifying priority stakeholders and their information needs.** In order to design the evaluation, it is essential to have a clear understanding of the gender-related questions to be addressed, which requires understanding who are the key stakeholders whose information needs should be prioritized. For multi-component programs there could be many stakeholders, so information requests must be prioritized. A gender focus also requires broadening the number and types of stakeholders, and this may require negotiation. Sometimes a government agency may assume that gender issues are adequately represented by one or a few gender focal points from government agencies, while the evaluation team may feel that a much broader range of women's groups, including from outside government, should be represented.

**Defining scope, boundaries and timeline of the evaluation.** Another task for the manager is to define the scope of the evaluation. This relates to the time period covered, how broadly

### Box 5.1 Key questions on the scope of the evaluation

- A. Time period:** What period of time will be covered by the evaluation? Will it start at project launch or will it explore the historical context? Will it end soon after project completion, or will it project into the future to assess sustainability?
- B. Narrow or broad definition of outcomes:** Will the evaluation only focus on direct outcomes defined as project gender objectives or will it also assess broader secondary and tertiary gender outcomes?
- C. Geographical coverage:** Will the evaluation only focus on the project communities or areas, or will it also look at broader geographical areas?
- D. Effects on whom?** Will the evaluation only look at effects on project women, on their immediate family or on the broader social network?

outcomes are defined, geographic coverage, and effects on whom? (Box 5.1). Finally, it is important to define the range and complexity of outcome indicators. Will the evaluation only measure quantitative outcomes (income, school enrolment) or will it also assess more complex qualitative dimensions such as empowerment? These decisions will have a major effect on the cost, duration and complexity of the evaluation.

**Reconciling resources and timelines with stakeholder priorities.** The manager must ensure the evaluation is not committed to producing gender-related findings that go beyond the available resources, that cannot be answered within the evaluation timeline or that require access to kinds of information that will not be collected. These questions are more difficult to decide for GRE as there is often less experience in the time and resources that will be required to address these questions.

**Evaluability assessment.** Once the evaluation design has been developed, it is good practice to conduct an Evaluability Assessment to ensure the evaluation is methodologically sound, is addressing all the key stakeholder questions, and that the questions can be addressed with

available resources and within the evaluation time-frame. These questions are particularly important for GRE as often they are introducing relatively new and less tested methodologies. There may also be additional organizational challenges as some of the required gender data may not be part of the standard project data collection system, and because in some cases there may be resistance or lack of support in collecting some of the data. The manager would take the lead role in coordinating the evaluability assessment.

**Promoting dissemination and use of the evaluation and ensuring stakeholder buy-in.** There is extensive evidence that GRE, like other forms of evaluation are under-utilized, and a key role for the manager is to promote utilization and develop an effective dissemination strategy to reach all stakeholders, including importantly intended program beneficiaries and women’s organizations. Appendix 8 presents a six-step strategy for promoting utilization and dissemination of GRE evaluations. This also requires that evaluation findings are built into the organizational learning strategy, and that gender indicators and GRE findings are built into key agency reports. ■



## References

- Asian Development Bank (2012) Gender toolkit: Energy. Going beyond the meter.
- Asian Development Bank (2013) Gender toolkit: Transport. Maximizing the benefits of improved mobility for all.
- Asian Development Bank and Australian Aid (2013b) Toolkit on gender equality results and indicators. Manila.
- Bamberger, M; Raftree, L and Olazabal, V (2016) The role of new information and communication technologies in equity-focused evaluation: opportunities and challenges. *Evaluation* 2016 Vol 22(2) 228-244
- Bamberger, M (2013) Engendering M&E. PREM Notes. World Bank
- Bamberger, M & Segone, M (2011) How to design and manage equity focused evaluations. New York: UNICEF
- Bamberger, M; Segone, M & Reddy, S (2014) National evaluation policies for sustainable and equitable development: How to integrate gender equality and social equity in national evaluation policies and systems. EvalPartners, UN Women and IOCE. [http://c.ymcdn.com/sites/www.afrea.org/resource/resmgr/books/NationalEvaluationPolicies\\_w.pdf](http://c.ymcdn.com/sites/www.afrea.org/resource/resmgr/books/NationalEvaluationPolicies_w.pdf)
- Bamberger, M; Segone, M. and Tateossian, F (2016) Evaluating the sustainable development goals within a “no-one left behind” lens through equity-focused and gender-responsive evaluations. Eval Gender, UN Women, UNICEF, UNEG. Available at: [https://www.academia.edu/29216199/Evaluating\\_the\\_Sustainable\\_Development\\_Goals\\_with\\_a\\_No\\_one\\_left\\_behind\\_lens\\_through\\_equity-focused\\_and\\_gender-responsive\\_evaluations](https://www.academia.edu/29216199/Evaluating_the_Sustainable_Development_Goals_with_a_No_one_left_behind_lens_through_equity-focused_and_gender-responsive_evaluations)
- Bamberger, M; Vaessen, J & Raimondo, E (2016) Dealing with complexity in development evaluation: a practical approach. Thousand Oaks, CA: Sage
- Bamberger, M (2017) Integrating big data into the monitoring and evaluation of development programs. UN Global Pulse and the Rockefeller Foundation. New York.
- Bardasi, E and Garcia, G (2016) Integrating gender into IEG evaluation work. Independent Evaluation Group. World Bank. Washington: DC
- Brisolara, S., Seigart, D and Sengupta, S (eds) (2014) Feminist evaluation and research: Theory and practice. New York: Guilford Press.
- BRIDGE (2007) Gender and indicators, Cutting Edge Pack.
- Byrne, D and Ragin, C. (2009) The Sage Handbook of Case-Based Methods. Thousand Oaks, CA: Sage
- Byrne, D (2009) “Complex realist and configurational approaches to cases: A radical synthesis.” In Byrne and Ragin (op.cit)
- Centre for the Study of Social Exclusion <http://sticerd.lse.ac.uk/case/>
- Cornwall, A (2008) Unpacking “participation”: models, meanings and practices. *Community Development Journal* (2008) 43(3):269-283
- Cornwall, A (2003) Whose voices? Whose choices? Reflections on gender and participatory development. *World Development* Vol 31, Issue 8 August 2003 PP 1325-1342
- Cousins and Whitmore (1998) Framing participatory evaluation. *New Directions for Evaluation*
- Davis, A & Guevara, M (2016) “A case study in complexity: Evaluating a long-term effort to prevent gender-based violence in El Salvador”. In Bamberger, Vaessen and Raimondo (editors) op.cit.

- Davoodi, Tiengson and Asawanuchit (2003) How useful are benefit incidence analyses of public education and health spending? IMF
- DeVault, M; and Gross, G (2012) "Feminist qualitative interviewing: Experience, talk and knowledge" in Hesse-Biber (ed) , 2012
- DFID (2008) The gender manual: A practical guide. London
- Espinosa, J (2013) "Moving towards gender-responsive evaluation? Practices and challenges in international development evaluation." *Evaluation*, 19, 171-182
- Funnel, S and Rogers, P (2011) Purposeful program theory: Effective use of theories of change and logic models. CA: Jossey-Boss
- Goetz, A.M and Gupta, R.S. (1996) "Who takes the credit? Gender, power and control over land use in rural credit programs in Bangladesh", *World Development*, Vol 24, No. 1 pp. 45-63
- Grieco, M; Apt, N and Turner, J (1996) At Christmas and on rainy days: Transport, travel and the female traders of Accra. UK: Avebury.
- Hashemi, S; Schuler, S Riley, A (1996) "Rural credit programs and women's empowerment in Bangladesh." *World Development*. 24 (4) 635-53
- Hesse-Biber, S (ed) (2012) The handbook of feminist research: Theory and Praxis. Ca: Sage
- IFAD 2017 What works for gender equality and women's empowerment: a review of practices and results. An evaluation synthesis
- Kane, M and Trochim, W (2007) Concept mapping for planning and evaluation. CA: Sage
- Karkara N, "Advocating for evaluation: A toolkit to develop advocacy strategies to strengthen an enabling environment for evaluation;" Segone M, Catsambas T, Rugh J (Eds), EvalPartners in cooperation with UN Women and IOCE (2013). Available online at: <http://www.mymande.org/evalpartners/advocacytoolkit>
- Kumar, S (2013) Understanding equity and growth from a development perspective. Background paper. New York: Rockefeller Foundation
- Kumar, S (2002) Methods for community participation: A complete guide for practitioners. London: ITDG
- Kurtz, C. F., & Snowden, D. J. (2003) The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Systems Journal*, 42(3), pp 462-483
- Krugman, J; Hanmer, L; Twigg, S; Hasan, T; McCleary-Sills & Santamaria, J (2014) Voice and Agency: Empowering women and girls for shared prosperity. Washington, D.C: World Bank
- Le Blanc, D. (2015) *Towards integration at last? The sustainable development goals as a network of targets*. DESA working paper No. 141 March 2015
- Levin, H and McEwan, P (2011) Cost-effectiveness Analysis Methods and applications. Second Edition. Sage Publications
- Lykes, M and Herschberg, R (2012) "Participatory action research and feminism: Social inequalities and transformative praxis." In Hesse-Biber (ed) 2012
- MacKay, K (2007) How to build M&E systems to better support government. Independent Evaluation Group. World Bank
- Mayne, J (2012) "Contribution analysis: Addressing cause and effect." In Forss, Marra and Schwartz (2012) Evaluating the complex: Attribution, contribution and beyond. Transaction publications
- Meier, P (2015) Digital humanitarians: How BIG DATA is changing the face of humanitarian response. New York: CRC Press
- Morra-Imas, L and Rist, R (2009) The road to results: Designing and conducting effective development evaluations. Washington, DC: World Bank
- OECD-DAC (2016) Handbook on the OECD-DAC gender equality policy marker
- Overholt, C; Anderson, M; Cloud, K & Austin, J (1985) Gender roles in development projects. A case book. Connecticut: Kumarian

- Oxfam America (2015) Women's economic empowerment and domestic violence: Links and lessons for practitioners working with intersectional approaches. [https://www.oxfamamerica.org/static/media/files/Womens\\_Empowerment\\_and\\_Domestic\\_Violence\\_-\\_Boris\\_Hughes\\_hX7LscW.pdf](https://www.oxfamamerica.org/static/media/files/Womens_Empowerment_and_Domestic_Violence_-_Boris_Hughes_hX7LscW.pdf)
- Patton, M (2008) Utilization-focused evaluation. California: Sage
- Patton, M (2011) Developmental evaluation: applying complexity concepts to enhance innovation and use. New York: Guilford.
- Pawson, R (2013) The science of evaluation: A realist manifesto. CA: Sage
- Pillow, W and Mayo, C (2012) Feminist ethnography: Histories, challenges and possibilities. In Hesse-Biber (ed) 2012 op.cit
- Poverty Action Lab [www.povertyactionlab.org](http://www.povertyactionlab.org)
- Raimondo, E; Vaessen, J & Bamberger, M 2016 "Towards more complexity responsive evaluations: Overview and challenges" in Bamberger, Vaessen and Raimondo, (Eds) op.cit Chapter 2
- Rao, A., Anderson, M and Overholt, C (eds) (1991) Gender analysis in development planning. A case book. Connecticut: Kumarian
- Salmen, L (1987) Listen to the people: Participant observer evaluation of development programs. New York: Oxford University Press
- Shadish, W; Cook, T and Campbell, D (2002) Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton-Mifflin
- Stern, E, N. Stame, J. Mayne, K. Forss, R. Davies & B. Befani (2012) Broadening the range of designs and methods for impact evaluation. Working Paper No 38. London, UK: Department of International Development
- UN Development Group (2015) Guidance Note: On supporting the tailoring of the 2030 agenda for sustainable development to nation contexts. Version 1.0 September 14 2015
- UN ESCAP (2007) Economic and social survey of Asia and the Pacific
- UN Evaluation Group (2013) Integrating human rights and gender equality in evaluation. Guidance Document
- UN Evaluation Group (2014) Integrating human rights and gender equality in evaluations
- UN (2015a) Transforming our world: The 2030 agenda for sustainable development
- UN Global Pulse (undated) Projects. [www.unglobalpulse.org](http://www.unglobalpulse.org)
- UN Women (2015a) Monitoring gender equality and the empowerment of women and girls in the 2030 Agenda for Sustainable Development: Opportunities and challenges. Position paper August 2015. New York: UN Women
- UN Women (2014) An empowered future: corporate evaluation of UN Women's contribution to women's economic empowerment. Independent Evaluation Office. New York: UN Women
- UN Women Women and the sustainable development goals. <http://www.unwomen.org/en/news/in-focus/women-and-the-sdgs>
- UN Women, Independent Evaluation Office (2015) How to manage gender-responsive evaluation. Evaluation Handbook. New York: UN Women
- USAID (2013) Complexity-aware monitoring
- UKAID/ACTIONAID (undated) A theory of change for tackling violence against women and girls
- Vaessen, J., E. Raimondo and M. Bamberger (2016) "Impact evaluation approaches and complexity" in Bamberger, Vaessen and Raimondo (eds) op.cit chapter 4
- Vaessen, J., A. Rivas and F. Leeuw (2016) Microcredit and women's empowerment: complexity in systematic review. In Bamberger, Vaessen and Raimondo (op.cit)
- White, H. (2009) Theory-based Impact Evaluation: Principles and Practice," *Journal of Development Effectiveness*, 1 (3), 271-284
- Williams, B., & Hummelbrunner, R. (2011) Systems concepts in action: A practitioner's toolkit. Palo Alto, CA: Stanford University Press

- Wilson-Grau, R., & Britt, H. (2012) Outcome harvesting. Cairo, Egypt: The Ford Foundation's Middle East and North Africa Office. Available at <http://www.outcomemapping.ca/resource/resource.php?id=374>
- Woolcock, M (2009) "Towards a plurality of methods in project evaluation: A contextualized approach to understanding impact trajectories and efficacy." *Journal of Development Effectiveness* 1(1): 1-14
- World Bank (2012) Gender equality and development. World Development Report. Washington, D.C: World Bank
- World Bank (2010) Evaluation of World Bank support 2002-2010: Gender and Development. Washington. DC: <https://www.oecd.org/derec/worldbankgroup/46993392.pdf>
- World Bank (2012) Mainstreaming gender in road transport. Chapter 2 "Gender trip patterns and mobility constraints. Washington, DC
- World Bank IEG (2016) Integrating gender into IEG evaluation work
- World Bank (2016) The Africa Gender Innovation Lab: Generating evidence to accelerate poverty reduction through gender equality. Available at: <http://www.worldbank.org/en/results/2016/05/09/the-africa-gender-innovation-lab-generating-evidence-to-accelerate-poverty-reduction-through-gender-equality>
- World Health Organization (undated) Social determinants of health: social exclusion. [http://www.who.int/social\\_determinants/themes/social\\_exclusion/en/](http://www.who.int/social_determinants/themes/social_exclusion/en/)
- Yin, R (2012) Applications of case study research. Thousand Oaks, CA: Sage
- Yin, R (ed) (2008) The case study anthology. CA: Sage
- Yin, R (2003) Case study research: Design and methods (third edition). CA: Sage



**ECG**  
EVALUATION COOPERATION GROUP

[www.ecgnet.org](http://www.ecgnet.org)





# ECG

EVALUATION COOPERATION GROUP



[www.ecgnet.org](http://www.ecgnet.org)