ECG
Evaluation Cooperation Group

Integrating Gender into Project-level Evaluation

ECG Reference Document – Appendixes

June 2017

Acknowledgments

The methodological guidelines are the product of a collective work of an ECG Task Force composed of the following institutions and members, led by the African Development Bank (AfDB), and supported by Michael J. Bamberger, Expert Consultant.

Task Force members	
Independent Development Evaluation (IDEV) **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)
Independent Evaluation Department (IED) **Asian Development Bank (AsDB)**	Farzana Ahmed, Lead Evaluation Specialist Hyun Son, Principal Evaluation Specialist
Evaluation Department (EvD) **European Bank for Reconstruction and Development (EBRD)**	Shireen El-Wahab, Principal Evaluation Manager Beatriz Perez-Timermans, Principal, Evaluation Manager
Operations Evaluation (EV) **European Investment Bank (EIB)**	Emmanuel Pondard, Evaluation Specialist
Independent Evaluation Office (IEO) **Global Environment Facility (GEF)**	Geeta Batra, Deputy-Director Anna Viggh, Senior Evaluation Officer
Independent Evaluation Office (IEO) **IFAD**	Catrina Perch, Evaluation Specialist Mark Keating, Evaluation Officer
Independent Evaluation Office (IEO) **UN Women**	Shravanti Reddy, Evaluation Specialist Sabrina Evangelista, Evaluation Specialist
Independent Evaluation Group (IEG) **World Bank Group (WBG)**	Gisela Garcia, Evaluation Officer Elena Bardasi, Senior Economist

Table of contents

Appendix 1: The rationale for integrating gender into project and program evaluations

The following paragraphs (adapted from Bamberger, 2013) present a rationale that can be adapted by different agencies to stress the importance of incorporating a gender focus into agency evaluation programs. In this case, the examples are taken from country-level impacts but sector examples can also be included. For a more comprehensive review see the Overview of the 2012 World Development Report on *Gender and Development.*

Why is it important to incorporate gender into M&E?

In every society, there are rules governing appropriate behavior for men and women and girls and boys, in the home, the community, the labor market, schools, and in politics. Some of these rules are regulated by social customs, others by laws or the operation of the labor market. Sometimes the forms of control are subtle, while others may be enforced by legal sanctions or the threat of violence. While some sectors of society may believe these rules to be based on "natural" differences between men and women, the rules are, in fact, socially constructed and vary from one society to another and over time. However, despite differences across societies, in every country that has been studied, these rules place women at a disadvantage with respect to key dimensions of development.

The persistence of significant gender inequalities in all regions negates fundamental human rights and the expansion of human freedoms. In addition, gender inequalities are serious barriers to the achievement of development objectives (Box A 1-1).

In *Gender Equality and Development* (World Bank, 2012), it is argued that promoting gender equity can make a major contribution to development: first, by fully utilizing the capacities of both women and men; second, through improved development outcomes for the next generation; and third, by making institutions more representative. Gender equity will open the doors to more policy choices and institutions will become more representative.

Box A 1-1: Estimated Economic Costs of Gender Inequality: Some examples from Africa, the Middle East and Asia.

- In the Middle East and North Africa, if women's labor force participation had increased in the 1990s at the same rate as women's education, the average household income would have been 25 percent higher.
- Tanzania could increase growth by one percent by removing barriers to women entrepreneurs.
- If India increased its ratio of female to male workers by ten percent, gross domestic product would increase by eight percent.
- Total agricultural output in Sub-Saharan Africa could increase by six to 20 percent if women's access to agricultural inputs were equal to men's.
- Asia is losing between US$42 billion and US$47 billion per year due to women's limited access to employment opportunities.
- Asia is losing between US$16 billion and US$30 billion per year as a result of girls' limited access to education.

In light of this compelling evidence, many governments and international development agencies have prioritized gender equality as one of their top development objectives. Achieving gender equality requires integrating gender into all aspects of programming, budgeting, implementation, monitoring, and evaluation. Many organizations have operationalized their gender equality strategies through gender action plans (GAP) that stress the critical role of appropriate gender tools for data collection and analysis for monitoring and evaluation (M&E). While a GAP provides a useful framework for an integrated approach to gender equality, it is not essential and many agencies begin by building gender into their existing M&E systems. They may then develop a broader gender framework after gaining experience with gender M&E.

Source: M. Bamberger (2013), Engendering M&E PREM Series on Nuts and Bolts of M&E systems. Note No. 27. World Bank.

Appendix 2: The special challenges of gender evaluation within the IEO context

[This is an expanded version of the discussion in Part I - Chapter One]

The challenges and opportunities resulting from the IEO mandate

The Reference Document recognizes that IEOs operate within a mandate that defines the scope of their evaluations, the time-frames within which they operate and the evaluation methodologies that they use. Given these parameters, the IEO approach to evaluation in general, and GRE in particular, is different from the evaluation scenarios and methodologies discussed in most evaluation textbooks. 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 and codes of conduct defined by their respective agencies.

This mandate provides both unique opportunities and challenges for conducting GREs. Opportunities arise because IEOs report directly to the Board of Directors, meaning there are institutionally defined mechanisms for the dissemination and use of the evaluation, and procedures to ensure the objectivity and independence of the evaluations. Challenges arise because IEO evaluations are conducted ex-post after projects have closed, so typically they cannot influence the kinds of baseline or implementation data collected on the projects being evaluated. Consequently, it is not always possible to apply most of the pre-test:post-test experimental and quasi-experimental evaluation designs advocated by evaluation textbooks. These challenges are particularly important for GRE where the processes of women's empowerment are regulated through social mechanisms and processes of behavioral change, which ideally should be observed over time rather than assessed through recall at the end of the process. However, this reference document refers to some of the new information technologies (smart phones and big data) that offer the possibility to reconstruct baseline and longitudinal data from existing data sets, such as twitter and other social media, satellite time series, phone records and electronic financial transaction data such as ATMs. These sources open-up the possibility to broaden the range of evaluation methodologies.

Special challenges facing gender evaluation

In addition to challenges that all development evaluations face, GREs face a number of their own special challenges:

a. *Relevance*: Many agency staff are not convinced that gender issues are relevant in all sectors. For example, some staff working in infrastructure, finance or trade may argue that their sectors are "gender neutral" and that both men and women have the same needs and will benefit (or be affected negatively) equally.
b. *Cost and time*: GRE frequently involve additional costs as more data has to be collected, and data may be more expensive or time-consuming to collect. Given the demands on evaluation resources, these considerations can be a serious constraint.
c. *Data is not available*: Where an agency has not collect gender data, this lack of data is given as the reason for not including gender in an evaluation, creating a 'chicken-and-egg' situation. As gender issues cannot be analyzed, it is sometimes argued that we do not know if gender issues are relevant – so data is not collected. GRE data can be more difficult to collect, so it may be hard to make the case for its collection without evidence as to its relevance.
d. *GRE requires the use of new and unfamiliar methodologies*: Some researchers who have established their professional reputations by conducting certain kinds of evaluation (e.g.,

collection and analysis of conventional quantitative data) may be reluctant to incorporate new kinds of data and methods with which they are not familiar. For some researchers, GRE can also seem threatening as it may question the validity of some of the traditional evaluation methods[i]. Some staff, particularly those trained in economics and quantitative methods, may not consider the more qualitatively oriented GRE as being "professional" evaluations.

e. *There may be concerns about the perceived political ideology of feminist researchers*. Although this will frequently not be the case, in some countries feminists are perceived as having a political agenda that may be disruptive or that may divert evaluations away from what are perceived to be their intended purpose for the agency.

Appendix 3: Overview of GRE Evaluation Designs

1. Evaluation designs used for standard GREs

The following are the most common evaluation designs used for standard GREs:

A. Descriptive post-project field visits. This is probably the most widely used GRE. Desk research and personal or phone interviews with key informants are combined with country visits by IEO staff and/or consultants. Rapid project visits and individual or group interviews with community groups, are combined with meetings with implementing agencies and local government officials, civil society and other key informant interviews. Focus group discussions (FGDs) with community groups, civil society or implementing agencies may also be included.

B. Quasi-experimental designs. When time and resources permit a survey may be conducted. This may use a pre-test: post-test comparison with a sample of the project population but where possible a comparison group should also be included. Given that IEO evaluations are conducted retrospectively, baseline (pre-test) data will normally use technique for reconstructing baseline data.

C. Theory based evaluation. A theory based evaluation, such as a theory of change (TOC), can provide a useful framework for the design and implementation of an evaluation. Frequently, the standard GRE will include a relatively simple form of TOC while special, in-depth evaluations may use a more elaborate form. Consequently, not all of the elements described here will be incorporated into every standard GRE.

A TOC can be particularly useful for GRE as it can help identify the many subtle and difficult-to-measure factors that can constrain successful outcomes of gender interventions. Many of these factors are not normally addressed in conventional evaluations, so the TOC can also serve as a checklist to ensure that important questions and indicators are not overlooked. Ideally, a TOC for a GRE should include the following elements:

a. *Problem diagnostic*: Description of the problem being addressed and some of its causes. Where possible this should include a historical analysis to identify how past experiences with interventions in this area may affect attitudes to the present project.
b. *Project Intervention*: Identification of the gender-responsive components and interventions.
c. *Implementation processes*: The processes through which the different components/services will be delivered.
d. *Outputs*: The gender-responsive outputs that each component is expected to produce.
e. *Outcomes*: The intended gender-responsive outcomes and the combination of outputs that are expected to contribute to each outcome. There is rarely a one-to-one relationship between a single output and a particular outcome, so it is important to identify the different outputs and external factors that can contribute to each one.
f. *Impacts or goals*: Some agencies break-down impacts into short, medium and long-term while others distinguish between impacts (that are clearly linked to the project) and broader development goals to which project impacts are only one contributing factor.

In addition to the above elements, that are usually presented in a linear figure with each element directly linked to the next level up, there are other important elements that should be addressed in the GRE.

- Contextual factors (e.g. economic, legal, organizational, political, cultural, climatic and historical factors that can affect how the project is designed, implemented and the gender-responsive outputs, outcomes it produces. This should be linked to factors affecting particular outcome and

impacts. These factors are closely linked to the challenges of addressing complexity in the evaluation design (Bamberger, Vaessen and Raimondo, 2016; Funnell and Rogers, 2011; Patton 2011).

- The TOC must be *falsifiable.* A theory must be testable (otherwise it is just a statement of belief or hope), which means that it must be possible to state that it has not worked or that there is no credible evidence to show that it did work. This requires more rigor in the articulation of the TOC than is usually the case. Requirements of a robust TOC include:
    - A timeline over which outputs, outcomes and impacts are to be achieved. A criticism of many TOC approaches is that they do not include this timeline, so that if expected outcomes have not been achieved defenders of the project can always argue that "our theory of change is valid, it is just that more time is needed to produce results." Ideally, there should be a relatively long timeline for GRE as many outcomes, particularly those involving behavioral or organizational change, evolve slowly over time. Also, the processes of transformative change are usually not linear, and there may be backlash and resistance which is recognized by feminist researchers in the saying "two steps forward – one step back".
    - Clearly defined and measurable indicators.
    - A set of alternative (rival) hypotheses that could explain how the expected outcomes could have been achieved. So even if the expected outcomes are achieved, it is not possible to conclude that these changes were due to the project intervention unless the rival hypotheses have been tested and found not to be credible.

g. *Emergence*: A criticism of most TOCs is that they are *static* and implicitly assume that the environment in which the program operates will not change over the life-time of the project. However, this static assumption is rarely true. Governments change, the local and national economy change, and other complementary or competing programs are launched. Equally important is the fact that the nature of the project and how it operates will also change. *Realist* evaluation (Pawson, 2013) has shown that projects change in response to interactions with affected communities and other stakeholders. Services and delivery styles that beneficiaries like will continue and be strengthened, while those that people do not like will often change or will die out if no-one uses them. Furthermore, programs will often evolve in new directions so that there will often be important unanticipated (positive and negative) outcomes. The TOC must have the flexible to adapt to the emergent environment.

   Emergence presents many challenges for the evaluator. In addition to the methodological challenges, there is the fact that many projects will be held accountable on the basis of the (almost always) relatively static results framework. While this framework can revise the original numerical targets, most results frameworks do not have the flexibility to incorporate completely new and unplanned outcomes or impacts, and implementing agencies are frequently not assessed on their flexibility to adapt to changing circumstances.

D. Objectives-based evaluation (results-based management). Many agencies incorporate into their project designs a results framework that defines a set of output, and outcome indicators that define the intended results. Usually these include baseline measures, initial intended targets to be achieved over a given period of time and the actual values achieved[ii]. The results framework is a good starting point for the evaluation, both because it provides a precise and comprehensive definition of intended results, and because projects are required to collect information on all of these factors.

However, the practical limitation for GRE is that most projects have only a few objectives that are disaggregated by sex, and even fewer specific gender objectives. So, while the results framework

provides a useful starting point, it will almost always have to be complemented by other sources of information. In some case it may be possible to reconstruct approximate gender objectives by, for example, asking staff how important the inclusion of women in committees and leadership positions or in access to training and project benefits was in the project design? Importance can be rated as "very important", "quite important", "not important" and "not considered". Many IEOs include these ratings in the gender flags (see Chapter 2, Section 2.4) that are used to rate project design and implementation in terms of how well gender issues are addressed. Women's actual participation or access to benefits in different activities can then be assessed in the evaluation. Of course, in cases where the project had gender-specific objectives, a more precise assessment can be made.

E. Case studies. Case study designs are based on the selection of cases (households, communities, schools and so forth) that are broadly representative of the total project population. Cases can be *descriptive* (illustrating the different types), *illustrative* (describing the different typologies found in surveys or other parts of the evaluation) or *analytic* (seeking to identify or explain different behavior, processes of change or outcomes). Case studies are a very powerful form of evaluation as they can dig deeper and help explain the lived experiences of different sectors of the target population. For gender evaluation, cases can often uncover and help understand subtle processes that are difficult to capture in interviews. The later section on special, in-depth evaluations will discuss the growing use of Qualitative Comparative Analysis (QCA) case studies. Robert Yin (2003, 2004, 2012) provides a useful introduction to case study research, while Byrne and Ragin (2009) provide a comprehensive review of case study methods. However, none of these texts focus directly on gender.

F. Qualitative methods. Almost all GRE incorporate qualitative (QUAL) methods into the evaluation design. These methods include unstructured and semi-structured individual interviews, group interviews and discussions (including focus groups), and observation (participant and non-participant). There are also a wide range of participatory group discussion techniques (PRA) that include many methods designed for groups with low literacy. One set of very useful techniques use social mapping to help understand the social, economic and power structure of a community, and another set of techniques traces historical timelines and the major events in the history of the community.

2. More in-depth methods for special GRE evaluations

Particular projects or themes are sometimes selected for more in-depth evaluation. The following are some of the wider range of evaluation tools that it may be possible to use for these studies when more time and resources are available and where the mandate may be broader. As indicated earlier, there is no clear line between standard and in-depth evaluations.

Theory based evaluation. In-depth evaluations are often able to use some of the more advanced approaches to theory-based evaluation, including:

a. Contribution analysis

Experimental designs are rarely possible, particularly for retrospective evaluations. Contribution analysis (Mayne, 2011) recognizes that it is rarely possible to assess the direct effect of a particular intervention as most projects are implemented in contexts where there are other agencies, other projects, new government policies and external factors, all of which can contribute to observed changes in the area where a project is operating. Consequently, the purpose is to define and assess the most plausible "contribution story". The analysis usually involves six steps:

- Define how the project is intended to contribute to a set of outcomes (the cause and effect relationship).
- Lay out the program theory (develop the "program story").
- Gather all of the supporting evidence.
- Assess and challenge the contribution claim (identify and test rival hypotheses).

- Find additional evidence.
- Strengthen the contribution claim.

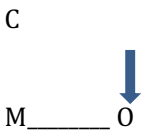
While many TOCs only describe and assess the project's theory of change, contribution analysis argues that it is essential to identify and test plausible rival hypotheses. It also stresses the importance of continually seeking new evidence to support or challenge the evolving theory.

b. Outcome harvesting

The approach is similar to contribution analysis, but while the former is often implemented at the start of a project and continues throughout its implementation, outcome harvesting is used at the end of a project and is consequently very relevant for IEO evaluations (Wilson Grau, 2012). Outcome harvesting collects stories from project beneficiaries and other stakeholders on changes that have occurred over the life of the project. It is quite common for hundreds of stories to be collected. They are organized into groups and their credibility, and their links to the project, are then assessed by the research team. Many evaluators find this approach useful because the stories (potential outcomes) are generated by the stakeholders themselves and not by the research team. Similar approaches to contribution analysis are used to test the credibility of the program story (theory).

c. Realist evaluation

Realist evaluation (Pawson, 2013) believes that the experimental design question "Did the intervention work in this context, with this population and at this point in time?" is too narrow to be of much practical value. Instead realist evaluation asks: "What works, for whom, in what respects, to what extent, in what contexts, and how?" These questions are addressed by defining "generative mechanisms" that explain "how" the outcomes were caused and the influence of context. The basic model is defined as: C (context), M (mechanisms) O (outcomes).



The unique feature of the approach is the belief that the "reasoning" of the actors in response to the resources or opportunities provided by the project is what causes the outcomes. Consequently, the process of understanding this reasoning and the factors that affect it implies a very different research approach from conventional evaluation designs that assume the outcomes are caused by the project interventions. The approach is potentially very valuable for GRE as women's reasoning in response to project opportunities is influenced by a wide range of social, economic, cultural, political, legal and historical factors – which is why project outcomes can be very different in different communities and contexts. The approach is also consistent with the GRE approach of considering women's agency.

d. Broader applications of qualitative methods.

The nature of most qualitative methods is that they rely on building trust with the communities being studied, and consequently their application ideally requires a considerable investment of time. So, although qualitative methods are widely used in standard evaluations, time constraints limit their full application. This is detailed in Appendix 4 below.

Appendix 4: Tools for collecting data for gender evaluations

This Appendix describes the most widely used data collection methods that can be used for the GRE evaluation designs described in Chapter 3 and Appendix 3. Many of the data collection methods are used in several of the evaluations, and Table 4.1 of the Gender Note lists the methods most commonly used for each design. Many of the data collection methods can either be used in a relatively simple way for standard GRE or in more depth for special evaluations. The following discussion does not try to distinguish between standard and more in-depth data collection applications as there is no clear-cut line between the two.

1. Mixed and multi-method approaches

Almost all IEO evaluations require an assessment of both *quantitative* (how much? how many? who is included and excluded?) and *qualitative* dimensions (understanding the implementation processes, the lived experiences of different groups, and mechanisms of social control). Consequently, all GRE should incorporate a mixed/multi-method approach[iii]. QUANT and QUAL methods are often combined in a somewhat *ad hoc* way (e.g. conducting not-very-well selected focus groups at the end of a survey, or commissioning a few case studies that are conducted with very little coordination with a QUANT survey). However, mixed method evaluation should be considered as an integrated evaluation strategy that can combine QUANT and QUAL approaches at all stages of the evaluation (Bamberger, Rugh and Mabry, 2012; Bamberger, 2016).

There are several advantages of a mixed method approach for GRE evaluations. First, *triangulation* (comparing independent estimates of a key indicator) can increase the reliability and validity of evaluation findings. Second, combining different methods may increase the validity of findings when working under budget constraints[iv]. Third, mixed methods help describe and understand interactions among different organizations and actors and to observe processes of behavioral change to complement and help interpret the QUANT findings.

There are several guidelines to keep in mind when using mixed methods:

- Mixed methods often involve professionals from different social science disciplines who have different ways of working. Consequently, it is important to allow more time for team-building and planning in order to integrate the different approaches and to fully benefit from the wider range of findings and analytical methods. Team building is particularly important for GRE as gender and feminist researchers often bring approaches that are unfamiliar to many QUANT researchers, which may initially create some resistance or questioning as to whether some of the QUAL methods really meet "professional" research standards.
- There should be close coordination with respect to design and application of all data collection methods. This is essential to permit triangulation so that one method can be used to validate data collected using a different method. This is only possible if all instruments are collecting comparable data. The situation often arises where estimates of household income obtained from a survey are inconsistent with estimates obtained from in-depth QUAL interviews. Often the questions are asked in different ways so that it is not possible to determine whether the differences are due to the way the question was asked or whether in fact one method is collecting more reliable information[v].
- The timing of the collection of different kinds of data must be coordinated so that the initial findings from one method can be used to correct any issues with another method. Often discrepancies are only found late in the research when it is too late to make any corrections.

2. Secondary data sources

Almost all evaluations begin with a review of available secondary sources. These can include:

a. Project documents. While many of these are included in project files and easily accessible, often project staff have their own files with additional useful reports. For older projects, some of these documents may not have been digitalized and harder to locate.
b. Reports from other agencies. Again, it may be useful to contact agency staff to request documents from their personal files.
c. Systematic reviews provide a synthesis of research and evaluation findings, that meet certain methodological standards, from all studies conducted on a particular type of intervention (e.g. off-grid rural electrification, village water supply). These provide a useful reference point to determine the maximum impact similar interventions have achieved. It is important to be aware that systematic reviews often have quite selective screening criteria, such as only including randomized control trials, so in many areas of gender research it may be the case that the vast majority of studies were excluded because they used qualitative methods.

3. Theory based evaluation (theory of change)

Theory based evaluations used a mixed methods approach for data collection as they draw on all available sources of data. For example, contribution analysis seeks at various points in the analysis to identify any new sources of data that may be relevant to making the "project story" more credible, or alternatively, seeking evidence that could challenge the credibility of the project theory. This will often involve seeking out sources of data that conventional evaluations would not normally use.

There are two kinds of theory-based evaluation that have their own sources of data and data collection methods:

a. *Outcome harvesting* (Wilson-Grau and Britt, 2012)*:* Outcome harvesting collects stories from project beneficiaries and other stakeholders on changes that have occurred over the life of the project. It is quite common for hundreds of stories to be collected. They are then organized into groups and their credibility; their links to the project are then assessed by the research team. Many evaluators find this approach useful because the stories (potential outcomes) are generated by the stakeholders themselves and not by the research team. Similar approaches to contribution analysis are used to test the credibility of the program story (theory).

b. *Realist evaluation* (Pawson, 2013) seeks information to answer the questions: "What works, for whom, in what respects, to what extent, in what contexts, and how?" This requires a creative use of mixed methods often with a strong reliance on QUAL methods to understand difficult-to-measure concepts, such as behavioral change and how mechanisms of social control influence decisions and actions of individuals and groups. Realist evaluation also examines how the context affects program outcomes and consequently draws on some of the kinds of data used in complexity evaluation and systems analysis. Process analysis is also used to understand how the *generative mechanisms* operate[vi]. These are the contextual factors that cause people to reason in a certain way about a project and which results in their acting in a specific way in response to the project.

4. Reconstructing baseline data

As almost all IEO evaluations are conducted retrospectively after the project has been completed, it is not possible to conduct a baseline study to collect data for a pre-test: post-test comparison. However, there are a number of ways that baseline data can be "reconstructed":

a. Using data collected by the project for project selection and design. For some projects, such as low-cost housing or micro-credit, applicants have to complete questionnaires which may include quite detailed socio-economic data (Bamberger, Rugh and Mabry, 2012, Chapter 5).
b. Use secondary survey data.
c. Conducting retrospective surveys where respondents are asked to recall their situation at the time the project began.
d. Key informant interviews.
e. PRA and other qualitative methods.
f. GIS data and satellite images.

5. Surveys.

For some larger evaluations, it may be possible to conduct surveys. These include a range of approaches: short, rapid surveys, unstructured or semi-structured interviews or large-scale structured sample surveys. Experience shows that if the evaluators have local counterparts it is often possible to conduct a short survey of several hundred households in one to two weeks and at a modest cost. Sometimes interviews can be conducted by local teachers or student nurses, or in some cases even high-school students.

Integrating gender into standard QUANT surveys will often require careful coordination. In many cases collecting the required sex-disaggregated data may be more difficult than it might seem. When household interviews are only conducted with the "head of household", in many cultures the majority will be men. Often the man may not know about the activities of female members (for example, how many hours they spend each day on household chores and collecting water and fuel?) Studies have shown that men often under estimate the time burden of such activities on their wives and daughters. In many cultures men also have little information on their children's education but are unwilling to acknowledge this fact. A number of studies have also found that men do not mention the several hours their wives spend each day collecting water or fuel as a problem[vii]. For all of these reasons it is important to plan how reliable sex-disaggregated data can be collected. In some cases, it might require a male and female interviewer to work together as a team, or it may mean arranging a follow-up interview with one or more female members of the household.

Observation is an important tool to complement surveys. For example, when asked who makes major decisions on household purchases and children's education, the wife will often say it is her husband. However, once the evaluator has gained the confidence of the wife and is invited into the house, she will often observe that in fact the wife is actively involved in these decisions.

6. Qualitative methods.

There are a wide range of qualitative techniques that evaluators can draw on. Many of these have the advantage of being relatively economical and have the flexibility to adapt to local conditions. They are also useful for the analysis of the process of project implementation, relationships among different agencies, and for understanding processes of behavioral change. Howard White used the term "ethnographic economics" to refer to the collection of insights on household or community dynamics from conversations with staff, community members and casual observation that can provide background for understanding factors that influence behavior and project outcomes that tend to get missed by conventional surveys and project visits[viii]. Some of the most widely used techniques include:

a. Key informant interviews.
b. Focus groups.
c. Observation checklists (for example, for studying how women participate in meetings and other project activities).

d. In-depth interviews.
e. PRA and other participatory group consultation methods.
f. Diaries: participants or project staff are asked to keep detailed records of their activities , such as time use during a typical day, use of different sources of fuel and power, expenditures.
g. Photographs and artefacts (such as ornaments and religious relics, furniture and household possessions, photographs and graffiti.
h. Audio and video recordings.

7. Broader applications of qualitative methods used in special, in-depth GREs

The following illustrate some of the ways that application of qualitative tools can be improved when time and resources permit.

Initial diagnostic studies

For community based projects, it is very useful to assign a researcher to spend several weeks in the community in order to help understand the daily lives, problems and concerns and attitudes to the proposed interventions (Pillow and Mayo, 2012; Clarke, 2012; Salmen, 1987). The goal is to be able to observe and to experience community life rather than to conduct interviews. It can also be helpful to understand the language and concepts used to think about and describe key concepts such as wellness and sickness, poverty and vulnerability, public agencies and the services they provide, and people's aspirations and fears. The researcher will also explore local history and how this affects attitudes to the proposed project.

It is important to note that in many cases the study can be conducted by a local researcher (with appropriate guidance) so the study does not necessarily have to be very expensive.

Key informants and informal panel studies

The luxury of more time makes it possible to better select key informants and to build confidence with them. Many of the most valuable informants prove not to be the people the evaluator met during the typical one to two day visit to a community.

If the evaluator is able to make several visits to the community over a period of time, it is often useful to develop a network of informants who can be visited periodically to report on the actual progress of the project on the ground, and what people are saying about it. For example, people often have different perceptions and expectations that what the project believes they are expecting.

Participatory planning and group consultation techniques

Participatory methods can be used both to involve communities in the planning, design, analysis and dissemination of surveys or as a participatory evaluation tool to obtain community perspectives on priority issues, the social structure of the community and the changing contribution made by the project over time.

The World Bank Social Observatory project in India has made extensive use of participatory techniques in their research projects. For example, village women designed a survey instrument to identify family needs in poor communities in India. The survey was then administered to almost one million people and the results were analyzed in cooperation between the women and the Social Observatory team. The women then designed and implemented participatory dissemination strategies, using pictures, marches and talks (World Bank, Social Observatory blog).

With respect to participatory group consultation methods for evaluation, there is a wide arsenal of tools and techniques that base planning and evaluation studies on feedback obtained through group

consultations. Participatory rural appraisal (PRA) is one widely used set of approaches (Lykes and Herschberg, 2012; Kumar, 2002). These techniques elicit perceptions of the community and its surroundings, power structures, constraints on women and other vulnerable groups, and a historical perspective on the evolution of the community. While there are risks of bias, of cooption of the process by a small group of more powerful people, or interpretations imposed (intentionally or inadvertently) by the researchers, these approaches are a very valuable tool. However, they require considerable time for their proper use in order to prepare the group, gain an understanding of the context, conduct the meetings (which can often take three to fives) and work with the community to interpret the findings. These techniques have been used extensively in gender research (Cornwall, 2003, 2008; Kumar, 2002).

Story-telling and sense-making

In recent years there has been an increased interest in story-telling. Often beneficiaries are asked to tell a story about significant recent events in the community or what they have experienced or heard about the project. Often the stories are analyzed using "Sense-making software". This approach can provide a different perspective by capturing and analyzing the perspective of the community rather than by asking them to respond to questions developed by the evaluator. Devault and Gross (2012) provide an overview of feminist approaches to qualitative interviewing, listening and story-telling.

Another approach to story-telling is to make audio or video recordings which can then be analyzed using some of the new data analytics software.

Case-based methods.

Case studies have always played an important role in program evaluation. They provide valuable ways to explain quantitative findings (particularly unexpected findings), and to provide in-depth insights into lived experiences, project implementation processes and behavioral change. A number of different case study approaches, all of which are useful for GRE can be identified:

a. Exploratory and descriptive case studies.
b. Illustrative case studies: These are often used as a follow-up to a quantitative survey. For example, survey analysis may create a typology of outcomes and cases can be prepared to illustrate or explain each type.
c. Analytical: the past decade has seen an increasing use of Qualitative Comparative Analysis (QCA). The unit of analysis is a case which can be as large as a country or as small as an individual. Cases can also be households or organizations. This method identifies the configuration of factors (for example, household attributes or community characteristics, such as access to infrastructure) for each case, that are present when the outcome is present. QCA is considered a useful way to provide an approximate estimate of causality when experimental designs are not possible. QCA also has the advantage that it can be used with small samples (50 or less cases), and it is also useful for the analysis of complex programs as it identifies configurations (combinations) of factors that are associated with an outcome, rather than analyzing a single factor (as is the case with experimental designs). For an overview of QCA methodologies see Byrne and Ragin (eds) 2009. UN Women (2014) illustrates how QCA was used to evaluate the impacts of UN Women's country strategies for promoting women' empowerment.

Experimental and Quasi-experimental designs. It is normally not possible to conduct an experimental design for IEO evaluations as they are conducted retrospectively. Yet it may occasionally be possible to use the findings of experimental designs conducted by other parts of the organization. For example, the World Bank's Africa Gender Innovation Lab conducts RCTs and quasi-experimental designs to test innovative approaches for the design and implementation of gender projects (Africa Gender Innovation Lab, 2016).

However, it may be possible to use a quasi-experimental design (QED). While these do not have the statistical rigor of RCTs with respect to internal design validity, QEDs often provide useful

approximations for causal analysis, and some would argue that a well-designed QED may be able to address external validity issues better than an RCT. The following are examples of QEDs that could be considered for GRE when resources and time permit:

a. *Post-test comparison design*. For example, many evaluations of the gender impact of microcredit programs use cross-sectional studies where women who used a village bank are matched with those who did not, where possible using propensity score matching. The challenge with these designs is how to address initial differences between the two groups (before the village bank began to operate) that might explain some of the differences in outcome indicators (expenditure on food, education and other household essentials}; investment in housing; profits generated by small businesses; women's role in household decision-making. For example, the women who took out loans might have had more small business experience, or they might come from families that were more supportive of women running a small business. Post-test evaluations find it difficult to control for these initial differences. However, if a mixed methods design is used there are many qualitative techniques that can explore possible initial differences and how they could affect outcomes.

b. *Natural experiments and pipeline designs*. Sometimes a project or policy is intended to reach all of the population of a district, province or the whole country. However, due to administrative problems, budget cuts or problems such as flooding, some sectors of the target population may not be reached or may be subject to long delays. In these cases, the outcomes for the groups that received the intended services can be compared with those that did not, thus providing an approximate estimate of project outcomes. While this approach is useful, and quite widely used, it must be interpreted with care as there may be systematic differences between beneficiaries and non-beneficiaries. For example, administrative problems may be more likely to affect poorer or more remote areas.

   A similar logic can be used in cases where projects are implemented in phases over a period of time. An example would be the installation of water and sanitation in a large urban slum. The project will usually be implemented in phases (starting at one end of the community and moving to the other) over a period of years. Another example is the construction of a road which again can take several years. The sectors of the community that will not receive the water and sanitation until Year 2 or 3 can be used as a comparison group for the Phase 1 areas that will receive the services in Year 1. Assuming that the characteristics of households in each phase of the project are similar (which is not always the case), then a baseline survey can be conducted in both areas at the start of Year 1 and then repeated at the end of Year 1. Comparing changes in the two groups can provide an approximate estimate of project impacts. The analysis can be refined to compare the changes for women and men.

c. *Reconstructing baseline data*. As discussed earlier, it may be possible to find baseline data so that a pre-test: post-test comparison design can be used.

Bamberger, Rugh and Mabry (2012) Chapter 11 and Appendix F review, with examples, all of the most common experimental and quasi-experimental designs.

Systems and complexity science-based approaches (Williams and Hummelbrunner, 2011, review the main systems analysis approaches)**.** There are a number of new evaluation approaches developed by complexity science that could be applied in GRE. All of these identify the main stakeholder or actors, the linkages between them, how information flows, leadership patterns and how decisions are made. Systems maps and models identify the linkages between organizations that facilitate or constrain desired processes of change. While some of the approaches, such as sociometric analysis require the collection of considerable amounts of survey data, other techniques are visualization tools to help conceptualize the nature of the system within which a particular program operates. Many of the approaches can be linked to a theory of change. Systems approaches are particularly valuable for gender

analysis as they help understand the complexity of the social system within which a project intervention is seeking to promote social change. These are usually tools to make clients (as well as evaluators themselves) aware that outcomes are influenced by many more factors than are usually taken into consideration in program design and evaluation. The following are some of the most widely used approaches, all of which are described with examples in Williams and Hummelbrunner (2011).

a. *Systems mapping:* This covers a number of approaches that provide a visual representation of a system. Systems mapping helps identify the different parts of a system and the linkages between the different parts that are likely to change. Systems maps can also identify positive and negative linkages between different parts of the system and the strength of the linkages. The approach is closely linked to a theory of change and provides a tool for examining in more detail the different parts of the system that affected the intended processes of change. A systems map could also be used to model all of the institutions and processes that comprise the system of social control that constrains processes of women's empowerment.

b. *Social network analysis.* The approach is useful for modelling stakeholder relationships and describing how information flows through the system and decisions are made. The approach is usually based on surveys or observations that calculate interactions between different individuals or groups and the structure of power and decision-making within a group or among different groups.

c. *System dynamics*: This deal with interconnectedness and dynamic relationships among different parts of a program system. The approach draws on engineering and management and involves developing a figure to represent stocks and flow variables. The approach can be used in evaluation to capture how complex systems are affected by, and respond to, development interventions. For example, Williams and Hummelbrunner (2011) used Systems Dynamics to evaluate the effects of microloan programs targeting sex workers in West Africa as part of a larger program to combat HIV/AIDS. The analysis was able to identify and help explain violent swings in the popularity of the microloan program with sex workers.

d. *Critical systems heuristics*: This approach focuses on understanding the factors that determine what is considered to lie within the system being studied that is what gets evaluated? This is principally an ethical decision based on a value judgment. This is important for GRE because in many cases gender issues are not considered to lie within the (project) system so gender is not included in the evaluation.

Concept mapping (Kane and Trochim, 2007). Concept mapping is a technique that uses interviews with stakeholders or experts to obtain an approximate estimate of program effectiveness, outcomes or impacts. A group of experts or stakeholders are asked to list the characteristics of a successful program (e.g. to promote women's empowerment). The listed items are sorted into groups (manually or by computer) and these are organized into a set of dimensions. The same or different group is then asked to rate (usually on a 1-5 scale) actual programs on these dimensions. The ratings can either be produced at one point in time (as would usually be the case for an IEO evaluation, or the ratings can be produced at the start and end of a project to measure change. Concept mapping has several advantages. First, the dimensions on which a program is evaluated are defined by stakeholders (or experts) and not by the evaluation team. This is particularly important for evaluating multi-dimensional, and difficult to define, concepts such as empowerment. Second, software is available to conduct the concept development and evaluation online so it can be much more economical and can involve a wider range of stakeholders or raters. The process can also be carried out relatively quickly.

Using new information technology for GRE (Bamberger, 2017). The past few years have seen a rapid development of exciting new sources for data collection and analysis that are opening up new approaches to program evaluation that it would have been difficult to have imagined even ten years ago. These are based on ICTs (smart phones, internet and other portable devices that can be used in the field and by even the poorest households living in remote regions), and on big data (generated from satellite

images, from twitter and other social media, electronic transfers, remote sensors and the internet of things). These are complemented by powerful new data analytics tools that can analyze vast quantities of data far beyond the capacity of office computers. Most of these involve new kinds of data, much of it collected for a completely different purpose (such as an ATM transaction or a social media discussion), are distinct from the kinds of information currently use by evaluators. In fact, evaluators have been much slower to adopt new information technologies than market researchers, medical researchers, program planners, social marketers, and agencies working in emergency relief.

Despite evaluators having been much slower on the uptake, there are now examples from all sectors and regions on the immense potential of new information technology for development evaluation. Many of the new technologies are of particular interest to IEOs, including for GREs, as they can overcome many of the constraints discussed earlier. Some of the potential applications and advantages include:

a. Much of the information is very fast and cheap to collect and analyze as it is derived from existing data sources and the evaluation does not have to pay for collection.
b. Economical access means that evaluations no longer have to rely on relatively small samples, but data can often cover the total population. This makes it possible to conduct kinds of data disaggregation that were not previously possible, as the small sample size did not provide sufficient numbers for disaggregated data analysis.
c. Ease of data access also makes it possible to incorporate many contextual factors which were previously inaccessible or to extensive to use. This permits placing programs in their broader contexts, which is very important for understanding the multiple factors that affect or constrain social change.
d. Longitudinal data sets are starting to become available that offer the potential to overcome the major constraint on retrospective evaluations, namely the lack of baseline data. Examples of longitudinal data sets include: an analysis of changing attitudes and information on major social issues as reflected on twitter (which now cover a number of years); data streams from satellite images that capture population movements, indicators of poverty and economic growth (such as vehicular traffic, types of house construction, nocturnal light emissions in poor communities, areas under cultivation).
e. Data analytics makes it possible to construct an integrated data platform that brings together many different sources of data using a common metric. This makes it possible to identify patterns of relationships between different data sets that were previously difficult to detect.
f. Real-time data feedback makes it possible to detect changes and trends even in post-project data

Portfolio analysis. Portfolio analysis was described in Section 2.4B. Data is collected from the following sources:

- Policy, planning and project documents from country programs. Documents are reviewed to determine whether and how gender issues are addressed. Sometimes the assessment simply indicates whether there is a reference to gender, but in other cases there is a rating of whether gender was a central priority or how thoroughly it was addressed.
- Key informant interviews may be conducted by phone, e-mail or in person to obtain opinions on how gender was approached in different parts of the program.
- Focus group interviews may be used in a small number of countries to dig deeper.
- A few countries may be selected to conduct field studies that may include project surveys, site visits and interviews with different stakeholder groups. A small number of representative projects may also be selected for more in-depth analysis. For example, the World Bank assessment of the implementation of their global gender policy (World Bank, 2009) began with the assessment of how well gender was incorporated into policies and projects in 93 countries where the Bank had active programs. Rating scales were used and 1,153 projects were analyzed. Based on these ratings, 12 countries were selected for more in-depth analysis (e.g. stakeholder

interviews), of which three were selected for focus group interviews and two for intensive country studies.

Gender flags. Many agencies have developed checklists for assessing whether and how well gender issues have been addressed in country program frameworks, sector programs or projects[ix]. The indicators either use a "Yes/No" format or rate how well the issue was addressed. Box 2.3 in Chapter 2 illustrates the questions included in the World Bank IEG *Gender Flag* country evaluation template. Checklists can either be used for self-assessment by operations staff or for external assessment by IEOs or consultants.

Data collection for case studies. *Descriptive case studies* mainly use QUAL methods to describe the context in which the project operates. *Illustrative case studies*, on the other hand normally used mixed methods data collection as surveys are analyzed to identify a typology of subgroups for example, farmers, village bank members, mothers using a childcare clinic. Cases are then selected within each typology and are studied using QUAL methods.

Qualitative Comparative Analysis (QCA) case studies use a specialized data collection method (Byrne, 2009; Byrne and Ragin, 2009). A matrix is constructed where each case is a row and each column represents either an attribute of the case, or of the context in which it is located, or the outcome variable. In the simplest form of QCA, all attributes and outcomes are dichotomized. For example, QCA might be used to assess how girls enrolment in secondary school (the outcome) is affected by: (i) whether the mother has completed secondary school; (ii) whether the girl lives less than 1,000 metres from the school; and (iii) whether the school has satisfactory toilets (as defined by the Ministry of Education) for girls. The columns would be defined as follows:

- Column 1: Case number.
- Column 2: Mother has completed secondary school [Yes = 1, No = 0].
- Column 3: Girl lives less than 1,000 meters from the school [Yes = 1, No = 0].
- Column 4: The school has satisfactory (as defined by the Ministry of Education) toilet for girls [Yes = 1, No = 0].
- Column 5 [outcome variable]: The girl is enrolled in secondary school [Yes = 1, No = 0].

The data may be collected in a special survey or it may be available from previous surveys or records. Often the variables are selected to reduce the time and cost of data collection.

Appendix 5: A data collection planning matrix for a GRE: Application to a hypothetical Village Development Project in Central Asia

Note: This is a hypothetical example to illustrate the use of the data collection planning matrix. This example only illustrates a few of the possible indicators and data collection methods that could be considered. This matrix could be used in the design of all three case studies discussed in the Reference Document.

	Evaluation design	Indicators	Data sources
Gender question 1: Did the project address women's needs?			
1. Did the project staff believe that women's needs were addressed?	a. Desk review combined with quantitative or qualitative surveys. b. If possible baseline data will be reconstructed to permit a pre-test: post-test design.	a. Did the project staff believe women's needs were addressed? b. Which needs were addressed?	**Addressing both questions** i. Interviews with project staff. ii. Monitoring and other project reports.
2. Did women believe that their needs were addressed?		a. The proportion of women who say projects: • Responded directly to their needs. • Responded somewhat to their needs. • Did not respond to their needs.	i. Sample survey with women. ii. Focus groups.
Notes on the feasibility of the proposed data collection methods: 1. Review monitoring and other project reports to check the kinds of information included on projects perceived to satisfy women's needs. 2. Are the original project staff still available to be interviewed? 3. Will time and resource permit the application of a survey? 4. What has been the experience interviewing women? Is this feasible? Do surveys or focus groups tend to work better with women?			
Gender question 2: How did the project affect women's empowerment?			
1. Participation in project-related community organizations.	a. Desk reviews combined with quantitative or qualitative surveys. b. Reconstructing baseline data if possible.	a. Number of women and men participating in different project organizations for each year of the project. b. Regularity of attendance.	a-i. Project monitoring reports. a-ii. Interviews with project staff. a-iii. Project profiles. b-i. Monitoring reports.

			b-ii. Attendance records b-iii. Visits to meetings to check records
2. Participation in decision-making in community organizations.		a. Number of women and men on leadership committees in project organizations. b. Level of participation of women compared to men measured by: i. Frequency of speaking. ii. Number of women's ideas approved.	a-i. Project monitoring reports. b-i. Observation checklists used in committee meetings. b-ii. Audio and video recordings (if permitted).
3. Impacts on women's status in the household			
a. Mobility.		Places visited: (adapt list to context): [1 point for each] a. Market. b. Clinic. c. Movies. d. Outside the village.	Covering all items: i Rapid sample survey with women. ii. Key informants. iii. Focus groups.
b. Ability to make small purchases.		Things purchased (adapt list) [1 point for each] a. Items for daily use. b. Items for self. c. Treats for children.	Covering all items. i. Observation during visits to families. ii. Rapid sample survey with women. iii. Requesting a group of women to complete an expenditure diary.
c. Involvement in major household decisions.		(Adapt list) [points to be decided]. May give different scores for decisions made on own or made with husband. a. Purchase of goat or small animal b. Household repair. c. Leasing land. d. Purchase of land or equipment for farm or business.	
Notes on the feasibility of data collection methods: 1-4 As for question 1. 5. Is there any experience with the application of multiple choice questions? How well do they work?			

Appendix 6: Example of a GRE design matrix: Evaluating a hypothetical village development project with defined gender objectives in Central Asia.

1. The project's gender objectives

It will be necessary to confirm what the stated gender objectives of the project are, and whether there are additional gender objectives that are implicit but not specifically stated which could be included in the assessment. Box A 6-1 identifies the gender objectives that might be included.

In the box these are divided into *outputs* and *outcomes* (following the categories used in most TOCs) but the distinction between the two categories is not very clear. For example, participation in a group is considered an output, whereas assuming a position of leadership is considered an outcome.

The primary outcomes are considered as the project's defined gender objectives. While some of these are clearly stated in the project document, others are inferred as possible gender outcomes that might be produced and which go beyond the defined project objectives. These could be classified as primary (defined) gender objectives and secondary outcomes (other potential effects on women in the project) and tertiary gender effects (on women in other communities). So, in the design of the project evaluation it will be important to make this distinction and also to agree with management which secondary outcomes should be assessed. It will be important for political reasons to clarify that managers are not being assessed on whether secondary and tertiary gender outcomes are achieved, but that these are being reviewed to better understand the broader potential outcomes that future Community Driven Development projects can (and cannot) be expected to achieve.

Box A 6-1: Possible gender objectives to be included in the assessment

Outputs

a. Women's active participation in the different project committees.
b. Women's active participation in the private group enterprises.
c. Women's equal access to project services and benefits.

Outcomes

Primary gender outcomes (defined in the project design)

a. Organization and leadership
   - Women assume leadership positions in different committees and groups.
   - Involvement in decisions on the selection of projects.
   - Infrastructure projects are selected that directly benefit women.

b. Access to, and control of economic and productive resources
   - Involvement in decisions on control of project resources.

c. Equal access to, and control of, resources for private group enterprises.

Secondary outcomes (effects on women in the project that are not identified in the project design) and tertiary outcomes (effects on women in other communities)

a. Women's position strengthened within the household
   - Women's role strengthened in control of household resources.
   - Women enjoy greater independence (for example geographical mobility).

b. Economic
   - Increased employment opportunities.
   - Increased income and earnings opportunities.

c. Social
   - Increased access to education for women and girls.
   - Greater personal security (for example, reduced domestic violence and sexual harassment outside the home).

3. The project's gender responsive interventions

The project has a number of components/interventions that are specifically intended to promote the equal participation of women and to promote their economic or social well-being, and a number of others that have the potential to advance the status of women (see Box A 6-2).

Box A 6-2: Project components and interventions designed to promote gender equality, and the economic and social well-being of women

1. Social inclusion strategy
    a. *Social mobilization*:
        i. Women are defined as one of the priority groups to receive training and other kinds of support to ensure their participation in project activities.
        ii. Annual meetings will be held with all partner agencies to review progress in achieving the participation of women and other target groups.
        iii. Efforts will be made to recruit women as social mobilizers.
    b. *Village profiles*:
        i. Will focus on issues of exclusion, including exclusion of women and will examine social, political, legal, economic and other factors that cause or sustain exclusion.
        ii. A social capital analysis will examine differences in social capital for women and men.
    c. *Village meetings*:
        i. It is required that at least two of the six committee members should be women.
        ii. A series of structured meetings are planned and the participation of women and men can be monitored.
    d. *Preparation of groups and statements of needs:* efforts will be made to ensure that women members will have the opportunity to state their needs.
2. Identification of priority social infrastructure
    a. It will be important to monitor/assess whether women's priorities are reflected in the selection of infrastructure [note: the PAD does not state whether this is a priority and how it will be achieved].

4. Definition of the gender questions to be addressed

Box A 6-3 identifies an initial list of gender-related questions relating to five dimensions:

a. History and context and how these affect the implementation and likely gender outcomes of the project.
b. Relevance of the project design to the achievement of important gender objectives.
c. Efficiency of project implementation with respect to the achievement of gender objectives.
d. Achievement of gender objectives (efficacy). Objectives are defined in terms of outputs and outcomes.
e. Sustainability and resilience. While the Appraisal Report discusses sustainability, it does not discuss resilience – which is a concept that has been introduced since this project was launched, but which is now frequently discussed together with sustainability.

Box A 6-3: Gender-related questions to be addressed

1. History and context

a. How was women's participation and access to benefits affected by the transition from the Soviet era?
b. How did high unemployment affect women's participation and access to program benefits?

2. **Relevance**

a. How does the program expect to strengthen women's empowerment and ensure women have access to program benefits?
b. Are the interventions relevant to the achievement of these objectives?
c. How relevant were inclusion, social mobilization and provision of infrastructure to achieving objectives?
d. How relevant was small scale private group enterprises?
e. How relevant are the gender objectives and their implementations strategies to the achievement of overall project objectives?

3. **Efficiency [not all agencies include efficiency in the evaluation]**

a. Were women consulted and involved in project design?
b. Did the selected projects reflect the priorities of women?
c. Were project implementation strategies conducive to the participation of women?
d. How did women's participation in design and implementation affect overall project outcomes?
e. The effectiveness of the social mobilization strategy.

4. **Achievement of gender objectives (efficacy)**

a. Outputs
   i. Women's participation in project leadership and management
   ii. Women's role in decision-making
b. Primary Outcomes
   i. Project effects on women's empowerment.
   ii. Assessing social and economic outcomes for women.
   iii. Did projects respond to women's needs and priorities?

c. Secondary and tertiary outcomes

   i. Were there positive and/or negative unintended project outcomes for women?
   ii. What effects did the project have on women' empowerment?

5. **Sustainability and resilience**

a. What evidence is there that the different gender outcomes will be sustainable?
b. Do individual women, and women's groups develop strategies to adapt to stress and shocks and do they learn from experience and improve how they adapt?

5. Initial proposal for the evaluation methodology

5.1. Table A 6-1 presents initial ideas for the evaluation design. This only provides a framework that will need to be elaborated once agreement has been reach on the general approach. The table includes three columns:

- The questions covering each of the five dimensions of the evaluation that are listed in Box 4.
- The indicators used to address each question.
- The data collection methods.

The framework identifies the following possible data collection methods. The application of each of these data collection tools will be explained once there is an initial agreement on the range of feasible and appropriate methods for this evaluation. It is recognized that not all of the listed methods will be used in this evaluation. The initial list of data collection and analytical methods includes:

1. Constructing a gender TOC that can help identify key evaluation questions, indicators and the pathways through which it is expected that gender outcomes will be achieved.
2. Constructing a data collection planning matrix. Appendix 5 presents an example of a data collection planning matrix that could be used in the design of the present evaluation. The matrix lists all of the key gender (and other) questions, the proposed evaluation designs, the indicators and the data collection methods. Importantly, it also assesses the feasibility of collecting the proposed information within the budget, time, methodological and organizational constraints within which the evaluation will be conducted.
3. Secondary sources:
    a. The Appraisal Report.
    b. The Completion Report.
    c. Project monitoring reports.
    d. Project profiles.
    e. Government reports.
    f. Partner reports.
    g. Civil society reports.
4. Consultations:
    a. Expert and key informants.
    b. Focus groups.
5. Surveys and questionnaires:
    a. Rapid, short sample survey.
    b. Rating scales to be completed by beneficiaries, project staff or partners.
6. Case studies
    a. Descriptive case studies.
    b. Analytical case studies (QCA).
7. Qualitative field work:
    a. Project visits
    b. Informal interviews.
    c. In-depth interviews.
    d. Observation.
8. Social media analysis (if this is feasible):
    a. Social media analytics (for example, analysis of twitter and other social media).
9. Pipeline designs to construct a counterfactual.

5.2. Evaluating project outcomes promoting women's empowerment. The project refers to social mobilization as a mechanism for strengthening the participation of women in community organizations promoted by the project and in the management of these organizations. It also promotes the participation of women in the private group enterprises. Although the project documents do not refer directly to gender empowerment, these objectives could be considered important elements in what many agencies would call the economic and social empowerment of women. Chapter 8 of the Reference Document discusses how these objectives could be incorporated into a gender empowerment framework and illustrates how an empowerment framework could be development to assess the effects of the project on promoting the different dimensions of women's empowerment.

Table A 6-1: Key gender-responsive questions, indicators and data collection methods:		
PART I: SOME GENERAL APPROACHES THAT CAN BE USED THRUGHOUT THE EVALUATION		
1. Gender theory of change: Developing a gender theory of change that is used to identify the key evaluation questions and to define the processes through which outputs and outcomes are to be achieved. This also defines key assumptions to be tested.		
2. *Results framework* (if it has been used in the project): This defines the intended gender objectives to be assessed.		
3. *Descriptive case studies*: Usually a relatively small number of case studies that are broadly representative of the main project scenarios and which are used to illustrate how the projects evolved and the lived experience of the project populations. Case studies can be longitudinal (conducted over a relatively long period of time), or (as is usually the case with ex-post evaluations) conducted at one point in time, and relying extensively on recall.		
4**.** *Analytical case studies*: These use techniques such as QCA (qualitative case analysis) to identify the necessary and sufficient conditions for project outcomes to occur, and also the necessary and sufficient conditions where project outcomes do NOT occur. Normally at least 30 cases are required. While QCA are rarely (if ever) used by IEG, they could be a potentially powerful analytical tool that permits attribution analysis to be used.		
5**.** *Counterfactual analysis*: This compares provinces or regions where the project has been implemented with regions where the project has yet been implemented. Many projects are rolled-out in phases so that a *pipeline design* can be used where regions where the project has not yet been implemented are compared with areas where it has been implemented. Sometimes the design is based on planned phasing-in of different regions while in many cases it is based on natural experiments where delays are caused by unplanned circumstances, such as delays in funding or administrative problems. While there are methodological limitations in the use of pipeline designs, they offer a useful tool for assessing attribution where experimental and quasi-experimental designs are not feasible.		
PART II: DESIGNS FOR ASSESSING EACH DIMENSION OF THE EVALUATION		
Dimension/ Question	**Indicators**	**Data collection methods**
1. HISTORY AND CONTEXT		
1. How was women's participation and access to benefits affected by the transition from the Soviet era?	Opinions from experts and key informants on the effects of high unemployment.	a. Expert and key informant interviews.
2. How did high unemployment affect women's effective participation?	Opinions from experts and key informants on the effects of high unemployment.	a. Expert and Key informant interviews. b. Appraisal Report and ICR.
2. RELEVANCE: The relevance of the project concept and design for promoting women's empowerment and social and economic benefits		
1. How does the program expect to	a. Develop a Theory of Change (TOC) that identifies the	a. The draft TOC would be developed by

strengthen women's empowerment and ensure women have equal access to services?	intended gender outputs and outcomes and the processes/steps through which these are to be achieved. Some of the outputs and outcomes may include: **Outputs.** a. Women participating in the different committees. b. Women participating in the private group enterprises. c. Women's access to benefits and services provided through the project. **Outcomes.** a. Women in leadership positions in the different groups. b. Women owning or gaining management positions in the private group enterprises. c. Women's access to, and control of economic and productive resources at the level of the household, community and enterprise. d. Women's and girl's increased access to educational opportunities. e. Women's increased geographical mobility.	consultants on the basis of project documents and interviews. Feedback would be obtained from project staff, beneficiaries and civil society. b. ICR. c. Interviews with project staff and partner implementing agencies. d. ICR Annexes 11 and 12.
2. How relevant are these intended outputs and outcomes for women's empowerment?	a. Women's opinions on the relevance of each output and outcome (see point 1 above) to their lives and to their feeling of empowerment [*** Note: an appropriate and understandable term for "empowerment" must be identified]. b. The opinion of key informants and women's organizations on the relevance of the outputs and outcomes for different groups of women.	a. Review theory of change. b. Focus groups with beneficiaries, project staff and civil society. c. An appropriate locally understood translation of "empowerment" would be developed through in-depth interviews and possibly discussion groups with beneficiaries. d. One approach used for the evaluation of the impacts of microcredit on women's empowerment in Bangladesh was to meet with local women in groups and to work with them to identify what for them would be key dimensions of empowerment (for example, being able to travel outside the family compound, to be involved in decisions on purchase of school uniforms, being able to walk through rice paddies without being

		required to step into the water to allow men to pass on the narrow pathway). Each item was then put into a scale and women were asked to locate their present position on each dimension[1]. The advantage of this approach is that the women themselves defined the dimensions that were important to them, and the advances (usually very modest) that they thought were possible.
3. How relevant are the inclusion strategy, social mobilization and providing essential infrastructure, to women's empowerment?	a. Women's opinions on the relevance of inclusion, social mobilization and provision of infrastructure for their lives and their feeling of empowerment. b. The opinion of key informants and women's organizations on the relevance of these project components for the different groups of women.	a. Expert and key informant interviews. b. Interviews with project staff. c. Information from the PAD and ICR. d. Focus groups and in-depth interviews. e. Observation of project activities (for example, do social mobilization training activities help women to participate more actively? f. Audio and video recordings of group activities.
4. How relevant are small scale private group enterprises for women's empowerment?	a. Women's opinions on the relevance of private group enterprises for their lives and their feeling of empowerment. b. The opinions of key informants and women's organizations on the relevance of these enterprises for different groups of women.	a. As for point 3. b. Review of group records (for example, loan sizes for women and men).
5. How relevant were the gender objectives and their implementation for the achievement of overall project objectives?	a. Opinions of project staff on the relevance of the gender objectives for the overall achievement of project objectives. b. Opinions of key informant and women's organizations.	a. As for point 3. b. Review of theory of change to check on critical assumptions.
3. EFFICIENCY: Gender-responsiveness of project implementation		
1. Were women consulted and involved in project design?	a. Proportion of women on different planning groups. b. Proportion of women who were aware of the proposed projects. c. Proportion of women who say they were consulted. d. Proportion of women who were involved in project	

[1] Hashemi, Schuler and Riley (1996) "Rural credit programs and women's empowerment in Bangladesh," World Development 24(4) : 635-53.

	implementation.	*An appropriate combination of the following tools and techniques will be used to address each of these questions.* Project monitoring reports. a. Project profiles. b. Focus groups. c. Interviews with project staff. d. Social media analysis (if appropriate). e. Rapid surveys (if feasible). f. ICR.
2. Did the projects selected reflect the priorities of women?	i. Proportion of women who say projects: ○ Responded directly to their needs ○ Responded somewhat to their needs ○ Did not respond to their needs	
3. Was the way the project was implemented conducive to the participation of women?	a. Were there guidelines on how to involve women in projects? b. How actively were women involved?	
4. How did the participation of women in project design and implementation affect the achievement of the overall project objectives?	a. How actively were women involved in project design and implementation? b. Were there any differences in how well projects achieved their objective depending on how actively women were involved?	
5. The effectiveness of the mobilization strategy.	a. Were there differences in how effectively mobilization strategies involved women and men? b. What were women' views on the effectiveness of the mobilization strategies?	
4. OUTPUTS AND OUTCOMES: Achievement of gender responsive project objectives {efficacy]		
A. OUTPUTS:		
1. Women's participation in leadership and management of projects.	a. The proportion of project organization leaders and managers who are women.	a. Project monitoring reports. b. Interviews with project staff. c. Project profiles. d. Audio and video recordings. e. Observation. f. Expert and key informant interviews.
2. Women's role in project decision-making.	g. What is the level of contribution of women to project decision making: ○ Women play a major role. ○ Women play some role. ○ Women play very little role.	
B. OUTCOMES: Assessing gender responsive outcomes.		
1. Project effects on women's empowerment.	a. Participation in community organizations. b. Participation in decision-making in community organizations. c. Did the selection of village projects reflect women's needs and priorities? d. Impacts on women's status in the household: • Access to and control over resources. • Participation in decision-making.	a. Theory of change. b. Monitoring reports. c. ICR. d. Experts and key informants. e. Case studies. f. Analysis of social media. g. Audio and video recordings. h. Focus groups.

	e. Geographical mobility. f. f. Strengthening social capital.	i. Observation. j. Rapid survey. k. Project profiles. l. Self-reported rating scales. m. Harvard-type time use and access to and control of resources checklists.
2. Assessing social and economic outcomes for women.	a. Girl's and women's education. b. Personal security. c. Geographical mobility. d. Strengthening social networks and social capital. e. Access to information about the outside world and about the community. f. Reducing time burdens.	
3. Do projects address women's needs and priorities [*** Note: this may be included under efficiency].	a. Proportion of women who say projects: • Responded directly to their needs. • Responded somewhat to their needs. • Did not respond to their needs	
4. Identifying unintended outcomes.	a. Did the projects have any unintended (unanticipated) positive outcomes for women? b. Did the projects have any unintended (unanticipated) negative outcomes for women?	
SUSTAINABIITY AND RESILIENCE:		
1. What evidence is there that different gender outcomes will be sustainable?	a. Have projects continued to deliver the same level of services over the life of the project (up-to to time of the evaluation?)	a. Sustainability checklists. b. Case studies. c. Observation.
2. Do individual women, and women's groups develop strategies to adapt to stress and shocks and do they learn from experience and improve how they adapt?	a. What lessons have been learned with respect to ways to anticipate and adapt to stresses and shocks?	

Appendix 7: Example of a GRE design matrix: Evaluating a hypothetical public transport project in Eurasia with defined gender objectives.

1. Framework of the evaluation report

It is proposed that the framework of the evaluation should follow the structure of a typical Independent Evaluation Group project assessment report. The gender assessment will be built into the respective sections of the overall evaluation. The proposed outline presented in Box A 7-1 also suggests some additional sections that might be included in a public transport project.

Box A 7-1: Proposed outline for the gender assessment component of the Public Transport Project in Eurasia.

[*Possible additional sections not normally included in this type of evaluation are indicated by* ***].

1. **History and context:**
   - **a.** *** The project's gender objectives and the project design through which the objectives will be achieved. This might include both the *explicit gender objectives* and some *secondary gender outcomes [objectives*] that might be achieved.
   - **b.** The Gender Theory of Change. *The IEO will decide if this is required.*
2. **Relevance**: The relevance of the project concept and design for promoting women's empowerment and access to the project social and economic benefits.
3. *** **Efficiency**: The gender-responsiveness of project design and implementation.
4. **Achievement of the project gender objectives** [efficacy].
   - a. Outputs.
   - b. Outcomes: primary (project gender objectives), secondary (other potential benefits for women participants not included in project design), and tertiary (effects on other women not involved in the project).
5. *** **Sustainability and resilience:** The likelihood that gender outputs and outcomes will be sustained and the bus company and other stakeholders will have the resilience to learn from experience and to adapt to the evolving context within which the project operates.
6. **Lessons learned:** how to design future projects to strengthen women's empowerment and gender equality.

2. The project's gender objectives.

2.1. Primary objectives

The Project Appraisal Document identifies several objectives relating to promoting equality of opportunity in the processes of recruitment, training and career advancement within the public transport company. These will be defined as the *direct gender objectives*. Following the practice of most theories of change, these objectives are divided into *outputs* and *outcomes* (see Box A 7-2).

However, there is extensive evidence that women and men have different travel and transport needs. Furthermore, many projects tend to focus more on men's transport needs (getting quickly from home to

work) and to overlook women's more complex needs to combine travel to work with taking children to school and the doctor and shopping (World Bank, 2010)[2]. These *multi-chaining* needs are ignored by many transport projects, resulting in the loss of many potential women passengers who have to rely on informal transport services. There are also issues of security as many studies have found that high proportions of women have experienced sexual harassment on public transport.

The Asian Development Bank (AsDB) Gender and Transport Toolkit reports that there are gender differences in travel patterns, use of transport modes, time use and time poverty, access to resources for travel, mobility and safety and that there are a number of gender barriers and differences in benefits[3].

As an objective of the project is to provide high quality service to passengers, Box A 7-2 identifies a number of *secondary and tertiary outcomes* that could be considered for inclusion. If any of these outcomes are included in the evaluation, it will be important to clarify that project managers and staff are not being assessed on whether secondary outcomes are achieved as these were not included in the project objectives. The reason for their inclusion is to help the European Bank for Reconstruction and Development (EBRD) learn lessons to improve the design and gender-responsiveness of future projects.

2.2. **Defining secondary and tertiary outcomes**

Chapter 9 discusses in more detail how secondary and tertiary outcomes could be defined and measured for this project. The proposed approach is strategically helpful because it shows that many of these outcomes are potentially positive and could significantly increase the estimated project benefits (rates of return). This is important because secondary and tertiary outcomes are often discussed in terms of "unintended outcomes" where the focus has usually been on the negative outcomes, such as increases in domestic violence or increased demands on women's time.

3. The project's gender responsive interventions

The project has several components/interventions that are specifically intended to promote the equal participation of women in recruitment, training and job advancement within the public transport company (see Box A 7-3). There is also a reference to ensuring the proposed specifications for the buses satisfy EU requirements for environmental and social impact. This could be interpreted to include strategies to prevent sexual harassment (a common problem on public transport) and possibly also to consider safety and convenience considerations with respect to the location of bus stops.

One important set of issues that do not seem to be addressed concerns the different transport and travel needs of women and men (discussed in the previous section). However, as there is no reference to these issues in the project document, these questions can presumably not be addressed when assessing project interventions.

[2] World Bank (2010). *Mainstreaming gender in road transport*. Chapter 2 Gender trip patterns and mobility constraints.

[3] ADB (2013) *Gender toolkit: Transport. Maximizing the benefits of improved mobility for all.*

Box A 7-2: Identifying gender objectives: Direct project gender outputs and outcomes, and secondary and tertiary gender outcomes.

Gender Outputs

Direct Gender Outputs

a. Equality of:
    - Recruitment opportunities.
    - Career progression.
    - Access to training.
b. Increase the number of women hired.
c. Workshop with stakeholders to disseminate lessons learned on how to promote gender equality.
d. Ensure buses meet European Union standards, including on gender considerations (such as security and comfort).

Secondary gender outputs [not defined in the project design]

a. Route planning and service frequency takes into consideration women's specific transport needs.
b. Service planning takes into consideration gender-related security and safety. For example:
    - Drivers and conductors are trained how to address sexual harassment on the buses.
    - Bus stops are well lit are not located close to bars or other areas with greater risk for women.
    - Additional security is provided for women travelling to and from work at night.
    - Drivers and staff are trained to respect women pedestrians [studies in a number of countries have found that drivers do not slow down to allow women pedestrians to safely cross the road, and sometimes drivers will not completely stop so that women have to get off when the bus is still moving].
c. Number of women passengers steadily increases.

Primary Gender Outcomes (program objectives)

Direct gender outcomes

a. More women are hired.
b. More women receive training.
c. More women are promoted.
d. Women's job stability increases.
e. More women become drivers and mechanics.
f. Gender good practice lessons are disseminated to, and adopted by other transport companies.

Secondary gender outcomes

a. Women's safety improves compared to other bus companies and forms of transport.
b. Gender responsive route planning contributes to reducing women's time burden.
c. Women' income increases (due to greater access to job opportunities through more convenient transport].
d. Children's health improves as gender-responsive transport planning makes it easier for working mothers to take children to the doctor.
e. Children's school attendance improves as it is easier for working mothers to take children to school.

Box A 7-3 Project components and interventions designed to promote women's access to employment opportunities and attention to women's concerns in the design of buses (and possibly bus stops).

1. Interventions to promote women's access to employment opportunities in the bus company.

a. Promoting gender-responsive recruitment policies.
b. Putting in place career advancement policies for women.
c. Training for women candidates to promote career advancement and increasing the number of women employed at the company.

2. Gender-sensitive service design.

a. Ensuring bus design responds to EU requirements on social and environmental impacts.
b. This requirement might also include locating bus stops and their design (for example, adequate lighting) to take into consideration women's safety and convenience.
c. Note: there does seem to be any reference to planning transport routes to respond to women's *multi-chaining* needs to combine travel to work with the need to take children to school and to the doctor and to go shopping.

3. Disseminating lessons on how to make transport programs more gender-responsive.
Workshops promoting knowledge sharing across the country and the wider region.

4. Definition of the gender questions to be addressed in the evaluation

Box A 7-4 identifies an initial list of possible gender-related questions relating to dimensions listed in Table A 7-1. These cover five dimensions:

a. History and context and how these affect the implementation and likely gender outcomes of the project. This includes information on earlier programs to incorporate women into the transport sector and broader efforts to broaden women's access to labor markets. Relevant legislation and government policies will also be described.
b. Relevance of the project design to the achievement of important gender objectives.
c. Efficiency of project implementation with respect to the achievement of gender objectives.
d. Achievement of gender objectives (outcomes/efficacy). Objectives are defined in terms of outputs and outcomes.
e. Sustainability and resilience. There is no direct reference to sustainability or resilience in the project document. However, these are important considerations as many initiatives to promote women's economic equality begin well but often encounter organizational, cultural, political and economic challenges that reduce their longer-term impact. So, a decision must be made as to whether it is appropriate to include this dimension in the evaluation.

Box A 7-4: Gender-related questions to be addressed in the evaluation

1. History and context

a. Have there been earlier initiatives to promote women's employment in the transport sector? What were the results?
b. Have there been initiatives in other sectors and what were the results?
c. Is there any legislation or government policies affecting women's labor market access? What have been the effects of these measures?
d. How did these earlier experiences affect the design of the present project?

2. **Relevance**

a. How does the program expect to strengthen women's access to employment and job advancement in the bus company?
b. Are the interventions relevant to the achievement of gender equality in the bus company and perhaps more widely?
c. How relevant are the gender objectives and their implementations strategies to the achievement of overall project objectives?

3. **Efficiency**

a. Were women consulted and involved in project design?
b. Do the selected intervention reflect the priorities of women?
c. Are there any additional design or implementation elements that should have been included?
d. Were project implementation strategies conducive to the participation of women?

4. **Achievement of gender objectives (efficacy)**

a. **Outputs:**
   i. Increased number of women recruited.
   ii. Increased number of women trained.
   iii. Rate of women's advancement increases.
   iv. Consultation mechanisms with women are put in place.
   v. Bus and service design takes gender issues into consideration.
   vi. Gender issues are included in workshops to disseminate lessons from the project.

b. **Primary gender outcomes.**
   i. Increase in proportion of women employed by the bus company. Proportion increase over the life of the project.
   ii. Increase in proportion of women in non-administrative positions (drivers, mechanics).
   iii. Increase in women's earnings.
   iv. Increase in proportion of women in managerial positions.

c. Secondary gender outcomes.

   i. Improved work-life balance.
   ii. Enhanced women's empowerment (see Chapter 9).

d. Tertiary gender outcomes

   i. Improved comfort and safety for women passengers.
   ii. Time saving.
   iii. Increased access to urban services and entertainment.

5. **Sustainability and resilience**

a. What evidence is there that women's advances in the company are likely to be sustained?
b. Have women in the company learned coping mechanisms for working and advancing in a male-dominated work environment?

5. Initial proposal for the gender evaluation methodology.

5.1. Table A 7-1 presents initial ideas for the evaluation design. This provides an initial framework that will need to be refined once agreement has been reached on the evaluation approach. The table includes three columns:

- The questions covering each of the five dimensions of the evaluation listed in Box A 7-4.
- The indicators used to address each question.
- The data collection methods.

The framework identifies a range of possible data collection methods from which an appropriate set will be selected for each stage of the evaluation. All of these methods are described in the ECG Reference Document. It is likely that not all of the listed methods will be used in the present evaluation. The initial list of data collection and analytical methods includes:

1. *Constructing a gender theory of change* [TOC] that can help identify key evaluation questions, indicators and the pathways through which it is expected that gender outcomes will be achieved. Not all gender evaluations use a TOC so a decision will be needed as to whether it is appropriate for the present evaluation.
2. *Constructing a data collection planning matrix*. Appendix 5 presents an example of a data collection planning matrix that could be used in the design of the present evaluation. The matrix lists all of the key gender (and other) questions, the proposed evaluation designs, the indicators and the data collection methods. Importantly, it also assesses the feasibility of collecting the proposed information within the budget, time, methodological and organizational constraints within which the evaluation will be conducted.
3. *Secondary sources.*
    a. The project document.
    b. Project monitoring and progress reports.
    c. Government reports.
    d. Partner reports.
    e. Civil society reports.
    f. Academic research and publications.
    g. Systematic reviews.
4. *Consultations.*
    a. Expert and key informants.
    b. Civil society organizations.
    c. Other funding agencies.
    d. Focus groups.
5. *Surveys, rating scales and checklists.*
    a. Rapid, short sample survey.
    b. Traffic and passenger surveys.
    c. Travel safety audits.
    d. Rating scales to be completed by beneficiaries, project staff or partners.
    e. Checklists.
6. Case studies.
    a. Descriptive case studies.
    b. Analytical case studies (QCA).
7. Qualitative field work.
    a. Project visits.
    b. Informal interviews.
    c. In-depth interviews.
    d. Observation.
    e. Participant observation.

    f. Travel diaries.
8. Social media analysis (if this is feasible).
    a. Social media analytics (for example, analysis of twitter and other social media).
    b. Internet surveys.
9. Pipeline designs to construct a counterfactual.

5.2. Measuring secondary and tertiary outcomes

Chapter 9 discusses in more detail how secondary and tertiary outcomes can be defined and measured.

Table A 7-1: Key gender-responsive questions, indicators and data collection methods		
•		
• **PART I: SOME GENERAL APPROACHES THAT CAN BE USED THRUGHOUT THE EVALUATION**		
1. Gender TOC: Developing a gender TOC that is used to identify the key evaluation questions and to define the processes through which outputs and outcomes are to be achieved. This also defines key assumptions to be tested.		
2. *Results framework* (if it has been used in the project): This defines the intended gender objectives to be assessed.		
3. *Descriptive case studies*: Usually a relatively small number of case studies that are broadly representative of the main project scenarios and which are used to illustrate how the projects evolved and the lived experience of the project populations. Case studies can be longitudinal (conducted over a relatively long period of time), or (as is usually the case with ex-post evaluations) conducted at one point in time, and relying extensively on recall.		
4. *Analytical case studies*: These use techniques such as QCA (qualitative case analysis) to identify the necessary and sufficient conditions for project outcomes to occur, and also the necessary and sufficient conditions where project outcomes do NOT occur. Normally at least 30 cases are required. While QCA are rarely (if ever) used by IEOs, they could be a potentially powerful analytical tool that permits attribution analysis to be used.		
5. *Counterfactual analysis*: This can be used at the national level for programs intended to cover the whole country, or at the project level. For the evaluation of the Transport Project, counterfactual would probably compare the project with other bus companies (assuming there any other companies that are of a similar size with common characteristics). The analysis would require a before and after comparison.		
PART II: DESIGNS FOR ASSESSING EACH DIMENSION OF THE EVALUATION		
• **Dimension/ Question**	• **Indicators**	• **Data collection methods** **[*appropriate tools will be selected for each phase of the evaluation from this list of options*]**
1. HISTORY AND CONTEXT		
a. Have there been earlier initiatives to promote women's employment in the transport sector? What were the results?	a. Transport sector projects with gender components/objectives. b. Gender outcomes. c. Transport projects that had unintended gender outcomes.	a. Expert and key informants interviews. b. Government reports. c. Civil society reports. d. Donor reports. e. Academic research.
b. How did these earlier experiences affect the design of the present project?	a. References in the project documents on previous gender initiatives.	

c. Have there been initiatives in other sectors and what were the results?	a. Other sector projects with gender employment components. b. Project gender outcomes.	
d. Is there any legislation or government policies affecting women's labor market access? What have been the effects of these measures? •	a. Legislation concerning women's labor market access. b. Opinions on policy outcomes.	
e. Are there any important economic, political or cultural contextual factors that have affected how the gender components were designed or implemented or that affected their outcomes?	a. Reports and indicators on contextual factors affecting women's labor market access.	•
2. RELEVANCE: The relevance of the project concept and design for promoting women's empowerment and social and economic benefits		
a. How does the project expect to promote women's access to employment and to improve the quality of transport services for women?	a. Develop a TOC that identifies the intended gender outputs and outcomes and the processes/steps through which these are to be achieved. Some of the outputs and outcomes may include: i. Direct Gender Outputs [see Section 4A]. ii. Secondary gender outputs [see Section 4B]. iii. Direct gender outcomes [see Section 4C]. iv. Secondary gender outputs [see Section 4D.]	a. The draft TOC would be developed by consultants on the basis of project documents and interviews. Feedback would be obtained from project staff, beneficiaries and civil society. b. Project document. c. Interviews with project staff and partner implementing agencies. d. Experts and key informants. e. Household income and expenditure surveys. f. Passenger and transport surveys. g. Social media analysis (twitter, and such like) if feasible. h. Participant observation (observer travelling on buses).

b. How relevant are these intended outputs and outcomes for women workers and passengers?	a. Women's opinions on the relevance of each output and outcome (see point 1 above) to their lives [*** Note: an appropriate and understandable term for "empowerment" must be identified]. b. The opinion of key informants and women's organizations on the relevance of the outputs and outcomes for different groups of women.	a. Review theory of change. b. Focus groups with beneficiaries, project staff and civil society. c. Expert and key informant interviews. d. Interviews with project staff. e. Information from the project document. f. Audio and video recordings of project activities. g. Project monitoring and progress reports. h. Application of OECD-type rating scale.
c. How relevant were the gender objectives and their implementation for the achievement of overall project objectives?	a. Opinions of project staff on the relevance of the gender objectives for the overall achievement of project objectives. b. Opinions of key informant and women's organizations.	

3. EFFICIENCY: Gender-responsiveness of project implementation		
a. Were women consulted and involved in project design?	a. Were stakeholders consulted on project design? b. Were there provisions to involve women in the consultations? c. Which groups of women were consulted? - Actual or potential employees - Civil society f. How many women were actively involved the consultations? •	a. Project monitoring reports b. Project profiles c. Focus groups d. Interviews with project staff e. Social media analysis (if appropriate) f. Rapid surveys (if feasible) g. Observation
b. Did the projects selected reflect the priorities of women	Proportion of women who say projects: o Responded directly to their needs. o Responded somewhat to their needs. o Did not respond to their needs.	
c. Was the way the project was implemented conducive to the participation of women?	a. Were there guidelines on how to involve women in projects? b. How actively were women involved?	
d. How did the participation of women in project design and implementation affect the achievement of the overall project objectives?	a. How actively were women involved in project design and implementation? b. Were there any changes in design in response to feedback from women? c. How did these changes affect the overall efficiency of the different project components?	
4. OUTPUTS AND OUTCOMES (IMPACTS): Achievement of gender responsive project objectives		
A. DIRECT GENDER-RESPONSIVE OUTPUTS:	**a. Direct Gender Outputs** i. Increased number of women recruited. ii. Increased number of women trained. iii. Rate of women's advancement increases. iv. Consultation mechanisms with women are put in place. v. Bus and service design takes gender issues into consideration. vi. Gender issues are included in workshops to disseminate lessons from the project.	a. The project document. b. Interviews with project staff and partner implementing agencies. c. Experts and key informants. d. Household income and expenditure surveys. e. Passenger and transport surveys. f. Social media analysis (twitter and such like) if feasible. g. Participant observation (the researcher spends time travelling on buses to

		observe sexual harassment and other issues affecting women passengers and drivers/conductors. h. Participant observation (observer travelling on buses). i. Focus groups with beneficiaries, project staff and civil society. j. Audio and video recordings of project activities. k. Project monitoring and progress reports. l. Application of OECD-type rating scale.

B. SECONDARY GENDER-RESPONSIVE OUTPUTS	**b. Secondary gender outputs.** i. Route planning and service frequency takes into consideration women's specific transport needs. ii. Service planning takes into consideration gender-related security and safety. For example: o Drivers and conductors are trained how to address sexual harassment on the buses. o Bus stops are well lighted and are not located close to bars or other areas with greater risk for women. o Additional security is provided for women travelling to and from work at night. o Drivers and staff are trained to respect women pedestrians [studies in a number of countries have found that drivers do not slow down to allow women pedestrians to safely cross the road, and sometimes drivers will not completely stop so that women have to get off when the bus is still moving].	

	iii. Number of women passengers steadily increases.	
C. DIRECT GENDER RESPONSIVE OUTCOMES	**c. Direct gender outcomes** i. Increase in proportion of women employed by the bus company. Proportion increase over the life of the project. ii. Increase in proportion of women in non-administrative positions (drivers, mechanics). iii. Increase in women's earnings. iv. Increase in proportion of women in managerial positions. *v.* Has the work/live balance improved for women workers?	
D. SECONDARY GENDER RESPONSIVE OUTCOMES	***d. Secondary gender outcomes*** i. Women's safety improves compared to other bus companies and forms of transport. ii. Gender responsive route planning contributes to reducing women's time burden. iii. Women' income increases (due to greater access to job opportunities through more convenient transport. iv. Children's health improves as gender-responsive transport planning makes it easier for working mothers to take children to the doctor. ***v.*** Children's school attendance improves as it is easier for working mothers to take children to school.	
3. Women's participation in leadership and management of projects	a. The proportion of project organization leaders and managers who are women	a. Project monitoring reports. b. Interviews with project staff. c. Project profiles. d. Audio and video recordings. e. Observation. f. Expert and key informant interviews.
4. Women's role in project decision-making	a. What is the level of contribution of women to project decision making: i. Women play a major role.	

	ii. Women play some role. iii. Women play very little role.	
B. OUTCOMES: Assessing gender responsive outcomes.		
5. Project effects on women's empowerment	a. Participation in community organizations. b. Participation in decision-making in community organizations c. Impacts on women's status in the household. d. Access to and control over resources. e. Participation in decision-making. f. Geographical mobility. g. Strengthening social capital.	a. Theory of change. b. Monitoring reports. c. Project document. d. Experts and key informants. e. Case studies. f. Analysis of social media. g. Audio and video recordings. h. Focus groups. i. Observation. j. Rapid surveys. k. Project profiles. l. Self-reported rating scales. m. Harvard-type time use and access to and control of resources checklists.
6. Assessing social and economic outcomes for women	i. Girl's and women's education. ii. Personal security. iii. Geographical mobility. iv. Strengthening social networks and social capital. v. Access to information about the outside world and about the community. vi. Reducing time burdens.	
7. Do projects address women's needs and priorities [*** Note: this may be included under efficiency].	a. Proportion of women who say projects: i. Responded directly to their needs. ii. Responded somewhat to their needs. iii. Did not respond to their needs.	
8. Identifying unintended outcomes.	a. Did the projects have any unintended (unanticipated) positive outcomes for women? b. Did the projects have any unintended (unanticipated) negative outcomes for women?	

SUSTAINABIITY AND RESILIENCE:		
3. What evidence is there that women's advances in the company are likely to be sustained?	a. Did the proportion of women recruited increase, and did the increase continue throughout the life of the project? b. Did the proportion of women who received advancements (promotion) increase and did these increases continue throughout the life of the project? c. Did the number/proportion of women receiving training increase, and did these increases continue throughout the life of the project? d. Opinions of different stakeholder groups on the likelihood that the improvements in points a-c will be sustained.	a. Sustainability checklists. b. Case studies. c. Observation. d. Focus groups. e. Individual interviews.
4. Have women in the company learned coping mechanisms for working and advancing in a male-dominated work environment?	a. What were the challenges facing women workers to sustain their progress? b. What were the opinions of other stakeholder groups? c. Did women report that they had learned any coping mechanisms to deal with problems identified in point b?	

Appendix 8: Links to the gender indexes

1. Africa Gender Equality Index
https://www.afdb.org/en/topics-and-sectors/topics/quality-assurance-results/gender-equality-index/
2. SDG Indicators
https://sustainabledevelopment.un.org/content/documents/21252030percent20Agendapercent20forpercent20Sustainablepercent20Development percent20web.pdf
3. The Gender-Related Development Index (GDI)
http://hdr.undp.org/en/content/gender-development-index-gdi
4. Gender Empowerment Measure (GEM)
https://en.wikipedia.org/wiki/Gender_Empowerment_Measure
5. Social Watch Gender Equity Index (GGI)
http://www.socialwatch.org/taxonomy/term/527
6. World Economic Forum Gender Gap Index (GGI)
https://www.weforum.org/reports/the-global-gender-gap-report-2016
7. Africa Women's Progress Scorecard
http://www.uneca.org/sites/default/files/PublicationFiles/agdi_2011_eng_fin.pdf
8. Thematic Indicators
FAO
http://www.fao.org/fileadmin/templates/ess/ess_test_folder/Workshops_Events/AFCAS_19/AFCAS_05_7_2_b.pdf
UNESCO
http://uis.unesco.org/en/glossary-term/gender-parity-index-gpi

Appendix 9: Strengthening the dissemination and use of gender-evaluation findings.

1. The underutilization of evaluations

There is extensive evidence that even methodologically sound evaluations are frequently under-utilized (Box A 9.1). Most of these findings are equally applicable to GRE but there are additional factors that can also constrain the utilization of gender-responsive findings. First, many projects do not have specific gender objectives, or gender objectives are defined very narrowly. In both cases, findings relating to potentially important gender-related secondary and tertiary outcomes may not be considered relevant if they were not included in the results framework. The case studies presented in Part II provide multiple examples of excluded gender issues. Second, project staff may be resistant to accepting the findings of gender analysis which they feel unfairly criticize them for not having addressed gender issues that were not included in the project design. Third, the way in which GRE findings are presented is often not linked directly to development objectives, but addresses broader concerns such as women's empowerment or human rights. Fourth, GRE often need to make a stronger case for the value added of the time and resources invested in addressing gender issues. Fifth, IEOs may not have a dissemination strategy that ensures the findings are accessible to the wider range of stakeholders (for example, civil society and women's organizations) that are the potential advocates and implementers of the evaluation findings. Finally, stakeholders often argue that the GRE findings are not based on a full understanding of the local context and that recommendations are not realistic within the local political and cultural scenario.

Box A 9.1 Reasons why evaluation findings are often underutilized

- Bad timing: the report is too late to contribute to policy decisions or it comes too early before agencies are focusing on these issues.
- Lack of flexibility to respond to the information needs of stakeholders.
- Wrong questions are asked and findings not considered useful or relevant.
- Evaluations are too expensive and make too many demands on agency staff and resources.
- Simplistic answers are given to complex issues.
- Evaluators do not understand the complexities of the local context.
- Local experts are not consulted/involved in the evaluation.
- Findings are not presented in a way that is easily accessible to different stakeholders.

Source: Bamberger, Segone and Tateossian (2016)

2. Strategies for promoting the utilization of GRE evaluations

Given the nature of the IEO mandate, not all of the utilization strategies discussed in the literature are applicable to IEOs. In particular, the requirement for independence and the fact that evaluations are conducted ex-post means that many of the *utilization-focused* evaluation strategies (Patton, 2008) that

require regular interaction with implementing agencies are not directly applicable. The following are some of the strategies for promoting GRE utilization applicable to IEOs:

a. Develop a dissemination strategy that addresses all of the issues identified in the previous section, including the specific gender-related constraints and opportunities.
b. Ensure findings and recommendations are aligned with project development objectives. Many GRE findings address broader issues some project staff may not find relevant or cannot implement. This presents a strategic challenge as many of these issues relating to empowerment, human rights and inclusion are critical to development. Findings must be integrated with a capacity-building strategy that can help agencies review and broaden their approaches to social as well as economic development.
c. Related to the previous point, utilization of GRE evaluation findings can only be achieved if they are based on a full understanding of the local political, economic and social context. Changes in deeply engrained systems of social control require the strategic identification of potential intervention points that are realistic within each local context.
d. Using "carrots" (incentives), "sticks" (sanctions and punishments) and "sermons" (show of support from respected figures) to encourage utilization (MacKay, 2007). An example of an incentive might be the availability of a gender fund that could be drawn on to support implementation of some of the recommendations.
e. IEOs may need to develop innovative communication mechanisms (such as use of social media, smart phones, collaboration with civil society, shorter publications targeted to particular audiences).
f. Ensuring that agreed actions on gender findings are included in the management action plans that most IEOs use to monitor implementation of the management agreements and commitments on evaluation recommendations.

Box A 9.2 Useful references on utilization of GRE evaluations

UN Women Independent Evaluation Office (2015) *How to Manage Gender-Responsive Evaluation: Evaluation Handbook.* Chapter 7 Use and follow-up.

Bamberger, M, M. Segone & S. Reddy (2014). *National policies for sustainable and equitable development: How to integrate gender equality and social equity in national evaluation policies and systems.* Eval Partners, UN Women and IOCE.

Karkara N (undated). "Advocating for evaluation: A toolkit to develop advocacy strategies to strengthen an enabling environment for evaluation".

3. Developing an organizational learning strategy

IEOs should coordinate with agency capacity development units to strengthen staff understanding of gender issues and outcomes and how they should be addressed in the design of future projects (programs and policies). GRE reports provide valuable teaching material as they illustrate how agency gender policies and strategies actually operate in the field. This can also help build awareness of the importance of building the collection of data on gender indictors into project M&E and management systems.

4. Building gender indicators and findings into key agency reports

IEO should work to ensure that gender indicators are incorporated into key agency reports and policy documents as well as into most IEG products. Chapter 2 described the different kinds of indicators that could be mainstreamed (see Section 2.7). The following are some of the indicators and evidence that could be incorporated into different agency reports.

a. Disaggregating standard socio-economic indicators by sex.
b. Presenting gender checklists.
c. Incorporating gender into the OECD/DAC standard rating scales.
d. Short illustrative case studies.

[i] Many evaluations rely on data collected at the level of the household. GRE evaluators argue that household level data ignores important differences in how food and other resources are distributed among household members. Disaggregated analysis is critical for GRE as women and girls often receive a smaller share of food so that malnutrition rates can be higher.

[ii] For example, Case Study No. 1 (the Village Investment Project) includes output indicators for each of the Project Components (capacity building and empowerment, village investments and management). For the village investments component, ten sets of output indicators are measured for core sectors including: people at the project level with improved water supply, construction of water supply points, direct female and male project beneficiaries.

[iii] Mixed methods are usually understood to combine quantitative and qualitative methods from two different social science disciplines while multi-method approaches involve combining different research and evaluation methods from within the same discipline. However, the distinction is not very clear and some offers prefer to combine the two approaches.

[iv] Often most of the evaluation budget might be invested in a quantitative survey so as to achieve maximum statistical power. Another option might be to reduce the sample size and to invest some of the resources in qualitative methods such as focus groups, key informant interviews and so forth Combining these independent estimates might produce more useful and meaningful results. However, increasing validity requires that QUAL data is selected to ensure it is representative and can be compared with the QUANT findings,

[v] An example of a common difference is when a QUANT survey is asking about changes in income since the start of the project while QUAL interviews are focusing more on feelings of economic security and vulnerability to economic crises. Sometimes surveys find that income has gone up but that many people feel more vulnerable and insecure. It often takes the researchers some time that these questions are exploring different issues.

[vi] Better Evaluation (http://www.betterevaluation.org/en/approach/realist_evaluation) defines a generative mechanisms as follows: Strictly speaking, the term 'generative mechanism' refers to the underlying social or psychological drivers that 'cause' the reasoning of actors. For example, a parenting skills program may have achieved different outcomes for fathers and mothers. The mechanism generating different 'reasoning' by mothers and fathers may relate to dominant social norms about the roles and responsibilities of mothers and fathers. Additional mechanisms may be situated in psychological, social or other spheres. Context matters: first, it influences 'reasoning' and, secondly, generative mechanisms can only work if the circumstances are right. Going back to our example, there may different social beliefs about the roles and responsibilities of mothers and fathers in different cultures, which may affect how parents respond to the parenting program. Whether parents can put their new learning into practice will depend on a range of factors – perhaps the time they have available, their own beliefs about parenting, or their mental health. Finally, the context may provide alternative explanations of the observed outcomes, and these need to be taken into account during the analysis.

[vii] A typical response is that women actually enjoy this travel time (and carrying heavy burdens on their head) time because they sing and chat with their friends as they walk.

[viii] A frequently cited example concerns the role of the mother-in-law on child nutrition. The mother-in-law is frequently not targeted in nutrition education programs, but she can prevent her daughter-in-law from applying the lesson learned in nutrition education programs.

[ix] A number of organizations such as UN Women, USAID, DIFD 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?)".