

ECG

EVALUATION COOPERATION GROUP



Gender Equality and Development Evaluation Units: Lessons from Evaluations of Development Support of Selected Multilateral and Bilateral Agencies

ECG Paper #5



ABOUT the Evaluation Cooperation Group

The Evaluation Cooperation Group is dedicated to harmonizing evaluation work among multilateral development banks. Its members are the evaluation departments of the African Development Bank, Asian Development Bank, European Bank for Reconstruction and Development, European Investment Bank, Inter-American Development Bank, International Fund for Agricultural Development, International Monetary Fund, the Islamic Development Bank and the World Bank Group. Observers are the Evaluation Office of the Global Environment Facility, the Evaluation Network of the Development Assistance Committee of the Organisation for Economic Co-operation and Development, and the United Nations Development Programme Evaluation Group.

Other evaluations and more information about the ECG can be found at: <http://www.ecgnet.org>.

This publication was led by the Independent Office of Evaluation of the International Fund for Agricultural Development.

Published as an Evaluation Cooperation Group paper, it does not reflect the ECG policy, nor the views of ECG members.

Gender Equality and Development Evaluation Units:
Lessons from Evaluations of Development Support of
Selected Multilateral and Bilateral Agencies

ECG Paper 5

©2012 Independent Office of Evaluation of IFAD (IOE)
International Fund for Agricultural Development
Via Paolo di Dono, 44
Rome, Italy
E-mail: evaluation@ifad.org
Telephone: (39) 06 54591
Fax: (39) 06 5043463
<http://www.ifad.org/evaluation>

This publication was prepared by the Independent Office of Evaluation of IFAD following a request from the Evaluation Cooperation Group. The findings and conclusions expressed herein do not necessarily reflect the ECG Policy, nor the views of ECG members and IFAD's Member States or representatives to its Executive Board. The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of IFAD concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. The designations "developed" and "developing" countries are intended for statistical convenience and do not necessarily express a judgement about the stage reached by a particular country or area in the development process.

All rights reserved. IOE encourages the dissemination of its work and will normally grant permission to reproduce portions of the work promptly. For permission to reprint any part of this work, please send a request with complete information to evaluation@ifad.org.

Cover - Photos of activities supported by IFAD-funded projects

Top: Katojo market in Uganda. Photo by Oanh Nguyen, IOE Evaluation Officer.

Left: Rural beneficiary in Surketh district, Nepal. Photo by Linda Danielsson, Assistant to the IOE Deputy Director.

Centre: Fruit and vegetable canning production in the Republic of Moldova. Photo by Jicheng Zhang, IOE Evaluation Officer.

Far right: Indigenous workers collect stones and other material for irrigation infrastructure in Ecuador. Photo by Giuseppe Bizzarri, courtesy of the IFAD Photo Library.

Bottom left: Production of avocados in Burundi. Photo by Anne-Marie Lambert, IOE Senior Evaluation Officer.

ISBN-: 978-92-9072-367-7

Contents

Acronyms	ii
Preface	iii
Executive summary	iv
INTRODUCTION	1
Purpose and scope of the report	1
LESSONS LEARNED	3
Lessons about policy and policy implementation	3
Policies	3
Policy implementation	6
Accountability	6
Institutional arrangements	7
Performance monitoring	9
Partnerships	9
Lessons on results	10
Strengthened human development	10
Increased access to and ownership of economic results	11
Improved voice and decision-making	11
Synergies between the three dimensions	12
COMPARATIVE REVIEW OF THE CORPORATE GENDER EVALUATIONS	14
Background and context	14
Clarity of evaluation objectives and associated results framework	14
Conceptual clarity and consistency	16
Quality of evaluation frameworks	21
MAINSTREAMING GENDER ISSUES IN EVALUATIONS OF EVALUATION UNITS	26
CONCLUSION	31
Annexes	
I. Methodology for assessment of gender integration in evaluations of evaluation units	32
II. References	35
III. United Nations definitions related to gender equality	36

Acronyms

ADB	Asian Development Bank
AfDB	African Development Bank
CIDA	Canadian International Development Agency
CLE	Corporate-level Evaluation
DFID	Department for International Development (United Kingdom)
EBRD	European Bank for Reconstruction and Development
ECG	Evaluation Cooperation Group (of the development banks)
ECOSOC	Economic and Social Council (of the United Nations)
FAO	Food and Agriculture Organization of the United Nations
IDB	Inter-American Development Bank
IED	Independent Evaluation Unit (of the ADB)
IEG	Independent Evaluation Group (of the World Bank)
IFAD	International Fund for Agricultural Development
IOE	Independent Office of Evaluation of IFAD
MDB	Multilateral development bank
NGO	Non-governmental organization
NORAD	Norwegian Agency for Development Cooperation
OED	Office of Evaluation (of FAO)
SDC	Swiss Agency for Development and Cooperation
SOFA	State of Food and Agriculture (report of the FAO)
UN-HABITAT	United Nations Human Settlements Programme
UNDP	United Nations Development Fund
UNICEF	United Nations Children's Fund
WB	The World Bank
WDR	The World Development Report (of the World Bank)

Preface

This synthesis was undertaken at the request of the Evaluation Cooperation Group (ECG) members, following a panel discussion of gender evaluations at the 2011 ECG meeting in Washington D.C. The findings and conclusions expressed herein do not necessarily reflect the ECG policy nor the views of ECG member institutions.

Since 2005 there have been a number of corporate-level evaluations documenting the first decade of the implementation of the 1995 Beijing resolutions to mainstream gender in multilateral and bilateral banks and development organizations. This synthesis is the first effort to draw out lessons for evaluators on designing and implementing these types of evaluations. There has been a significant diversity in how evaluators have approached assessing progress in gender equality: diversity in understanding the goals of gender equality, in how to bring about progress on gender equality, and in how to measure that progress. The synthesis examines these differences in evaluation approaches and methodologies. The diversity reflects both agency understanding of gender equality as well as evaluator understanding. For this reason it will be of interest to evaluators, as well as to those responsible for managing the integration of gender equality in the work of bilateral and multilateral agencies. In addition, the report provides a summary of the major findings of the evaluations. A final contribution of this synthesis is to examine a sample of evaluations not specific to gender equality to see how well gender equality was integrated in those evaluations. This will also be of interest to evaluation units.

The report was prepared under the leadership of Ashwani Muthoo, Acting Director of the Independent Office of Evaluation of the International Fund for Agricultural Development (IFAD), who was ably supported by Rieky Stuart and Gita Gopal, consultants. The work benefited from feedback and advice provided by ECG members and others. Special thanks are due to respondents from the African Development Bank, the Asian Development Bank, the European Bank for Reconstruction and Development, IFAD, the World Bank, and the March 2012 discussion by ECG group members in Luxembourg.

Other non-ECG members also provided useful comments during the preparation of the synthesis including the Food and Agriculture Organization of the United Nations, the UN Office of Internal Oversight Services, United Nations Educational, Scientific and Cultural Organization, UN Women, and the United Nations Development Programme.

The synthesis team greatly appreciates the efforts of all of these people in helping complete this report. The successful coordination among the members of the ECG in producing it is an important step toward achieving the ECG's mission of fostering collaboration and harmonization of evaluation work among the evaluation units of its members and disseminating evaluation findings to help maximize the relevance and effectiveness of development assistance.

Jan Willem van der Kaaij

Chair, Evaluation Cooperation Group
and Inspector General, European Investment Bank

Executive summary

This report presents the findings of an examination of recent evaluative work on gender equality undertaken by selected multilateral and bilateral organizations commissioned by the Evaluation Cooperation Group of the multilateral development banks. The purpose of the assessment is two-fold: (i) to present evaluation findings on what works on the ground in terms of strengthening gender equality in-country and what organizational systems and investments support that success; and (ii) to better understand to what extent evaluation units of international development agencies can improve their gender-related methodologies and better integrate gender considerations into their evaluations to support agencies in fulfilling their gender equality mandates. The main added value of this report is to examine these evaluations to better understand the strengths and weaknesses of their evaluation approaches and methodologies.

There is an important link between the gender equality work that is being evaluated and the evaluation methodology. This link is the ‘theory of change’ that is implicit or explicit in agency policies or directives on gender equality – and therefore affects the outcomes and mainstreaming strategy - and the theory of change expressed in the evaluation approach and methodology. To explore this link, the paper examines 11 thematic gender evaluations to briefly synthesize findings about what works well in supporting gender equality and what does not, at the level of outcomes and also at the level of organizational systems and infrastructure. The paper then explores the theory of change about gender equality underlying the evaluations and their methodologies. While the African Development Bank review of gender equality focused mainly on improvements for gender mainstreaming, and the World Development Report focused on impact, this report links those two areas to evaluation and how this can be strengthened.

In addition, the report presents the findings of a review of 59 evaluations undertaken by evaluation units of 6 multilateral agencies to better understand how evaluation units are integrating gender into their work.

Examining reported outcomes, the assessment finds that while the evaluations indicate that the agencies have made some important contributions to gender equality, particularly in education, in the integration of women in economic activities, and in addressing inequalities in access to economic opportunities and resources, there remains much room for improvement. Particularly weak are supporting advances in women’s reproductive health, replicating successful community-level efforts at economically empowering women, increasing voice of women in decision-making, and redistributing household work more equitably.

The evaluations found that performance was not as strong as it could be for a number of reasons. In particular, there was weakness in the policy and associated results frameworks or action plans that set objectives, targets, or measures that could be used to assess progress. This weakness was exacerbated by weak institutional systems (planning, implementation, monitoring and evaluation), lack of leadership and accountability, and inadequate investment in capacity and training, which were found to hamper translating institutional gender policies or action plans into support for gender equality.

These weaknesses then affect the evaluations themselves. Methodologically, the evaluations assess achievements against the stated objectives of the agencies, which for most agencies is “gender equality” and/or “women’s empowerment”. Without a clear results framework, it is highly challenging to design and implement a strong and useful evaluation because it is unclear what these terms mean. Lack of understanding of the desired outcomes results in all actions

related to women, and subsequent results considered as sufficient to support gender equality or women's empowerment. There is some evidence that agencies with more robust policies and results frameworks are able to capture results better and potentially use this information for learning and improvement. While most evaluations examined 'gender mainstreaming' as the strategy of choice used by the agency, the evaluations had widely differing interpretations of what gender mainstreaming is, (in some cases reflecting agency interpretations). This further diminishes comparability and learning across evaluations. Furthermore, the multiple interpretations of "gender mainstreaming" create a potential source of confusion for their common client.

In addition, many in the sample of 59 non-gender evaluations reviewed are not strong in incorporating support for gender equality in the social and economic sectors. Although the sample is small, evaluation units that give clear directives for integrating gender considerations in all evaluations perform better.

In conclusion, what stands out sharply when these evaluations are reviewed for results and through the lens of good evaluation methodology are two interlinking aspects. The first is the weakness in results frameworks (namely, lack of clarity in gender equality objectives) – this affects both results on the ground as well as evaluation methodologies. The second is the weakness in evaluation frameworks caused by confusion about the theory of change at the level of outcomes and the understanding of gender mainstreaming, as well as lack of targets and clear indicators with transparent rating scales. This also diminishes the ability of evaluative work to generate strong and credible lessons that can ultimately enhance the quality of support for gender equality and the results thereof.

There is currently considerable work being done to sharpen the policy and results frameworks for gender equality and women's empowerment (International Fund for Agricultural Development, United States Agency for International

Development with the International Food Policy Research Institute and the Oxford Policy Group, Cooperative for Assistance and Relief Everywhere [CARE], UN Women, Food and Agriculture Organization of the United Nations, Asian Development Bank, World Bank) and Evaluation Cooperation Group members can build on this work to harmonize their own evaluative practice and to encourage agency management to adopt more robust results frameworks and indicators in their ongoing portfolio management.

Introduction

“Assessing progress in the mainstreaming of a gender equality perspective is a bit like picking up mercury. It all too quickly slips through your fingers. There is often no agreement on what to look for, how to measure progress, how ‘high the bar’ should be. Until organizations have clear objectives and targets of what they hope to achieve and how they will monitor and measure those achievements, it will be up to evaluators to sort out what they are looking for.”
(United Kingdom Department for International Development Synthesis Report, 2006, page 5)

Purpose and scope of the report

This report presents the findings of an examination of recent evaluative work undertaken by selected multilateral and bilateral organizations. It examines 11 thematic gender evaluations to share findings about what works well in supporting gender equality and what does not, at the level of outputs and outcomes and also at the level of organizational systems and infrastructure. The report was commissioned to share knowledge and learning from a series of recent corporate-level gender evaluations. Most of the organizations evaluated now have 15 or more years of experience in using a gender mainstreaming approach to contribute to gender equality. Boards of Directors have asked for these evaluations to assess the success of the approach, as well as to provide evidence of agency contributions to gender equality, for purposes of accountability and for improving guidance and practice. This report also examines these evaluations to better understand the strengths and weaknesses of their evaluation methodologies. In addition, the report presents the findings of a review of 59 evaluations undertaken by evaluation units of 6 multilateral agencies to better understand how evaluation units are integrating gender into their work.

The purpose of the assessment is two-fold: (i) to present evaluation findings on what works on the ground in terms of strengthening gender equality in-country and what organizational systems and investments support that success; and (ii) to

better understand to what extent evaluation units of international development agencies can improve their gender-related methodologies and better integrate gender considerations into their evaluations to support agencies in better fulfilling their mandates.¹ The report will therefore be of interest to evaluators, Boards of Directors, and those responsible for achieving gender equality results in a range of development agencies. Recent reports like the World Bank’s World Development Report (WDR) 2012 present greater depth on overall outcomes of gender equality work, and the review of gender evaluations of the African Development Bank (AfDB) is more thorough on the mechanisms and processes of gender mainstreaming. This report’s emphasis is on evaluation department approaches and methodologies, and how evaluation practice can strengthen agencies’ efforts to contribute to gender equality.

All 11 agencies have undertaken recent corporate-level evaluations of their performance related to gender equality (table 1). The World Bank (WB) has undertaken two such evaluations over the span of the last decade, giving an overview of almost 20 years of efforts to address gender equality in the Bank’s portfolio. Evaluations undertaken by the three bilateral agencies (Canadian International Development Agency [CIDA], Department for International Development [DFID], Swiss Agency for Development and Cooperation [SDC]) were included in this review to provide a broader base for comparison in gleaned lessons. Table 1 lists the 11 evaluations covered by this assessment.² In addition, to assess how evaluation

1. This paper uses the standard UN definitions, which can be found in annex III, and at <http://www.un.org/womenwatch/osagi/gendermainstreaming.htm>.

2. The evaluations indicated that the 13 agencies that were considered for this report have gender policies or gender action plans. However two of the 13, European Bank for Reconstruction and Development and Inter-American Development Bank, have not undertaken recent corporate gender evaluations. Thus, the gender evaluations of only 11 agencies have been covered for this section.

departments dealt with evaluating gender equality in their evaluative work, 59 recent evaluations of multilateral agencies whose main purpose was not to assess gender-equality-related performance were selected and reviewed for their attention to gender equality (The list of these evaluations can be

found in annex I). While not an evaluation, the WB's 2012 WDR³ presents an in-depth analysis of progress and barriers to gender equality worldwide, and was also utilized to provide additional information, as relevant, to deepen this review of evaluations.

Table 1. Evaluations reviewed

#	Agency	Year	Name of evaluation (and evaluation unit)
1	Asian Development Bank	2009 2010	ADB's Support to Gender and Development - IED Evaluations of 2009 and 2010 (Independent Evaluation Department [IED])
2	African Development Bank	2010 2011	Mainstreaming Gender Equality: A Road to Results or a Road to Nowhere? An Evaluation Synthesis (A Meta Evaluation by Operations Evaluation Department (OPEV))
3	World Bank	2001 2009	Evaluation of World Bank Support for Gender and Development (Independent Evaluation Group ([IEG]))
4	IFAD	2010	IFAD's Performance with Regard to Gender Equality and Women's Empowerment (Independent Office of Evaluation [IOE])
5	UNICEF	2007	Evaluation of Gender Policy Implementation in UNICEF (UNICEF Evaluation Office [UNICEF EO])
6	UNDP	2006	Evaluation of Gender Mainstreaming in UNDP (Evaluation Office [EO])
7	UN-HABITAT	2011	Evaluation of Gender mainstreaming in UN-HABITAT (Monitoring and Evaluation Unit [MEU])
8	FAO	2011	Evaluation of FAO's Role and Work Related to Gender and Development (Office of Evaluation [OED])
9	CIDA	2008	Evaluation of CIDA's Implementation of its Policy on Gender Equality (Evaluation Division of the Performance and Knowledge Management Branch (PKMB))
10	SDC	2009	Evaluation of SDC's Performance Mainstreaming Gender Equality (Evaluation and Controlling Division [E+CD])
11	DFID	2006	Evaluation of DFID's Policy and Practice in Support of Gender Equality and Women's Empowerment
Other current sources of information on gender equality results			
	World Development Report 2012	2011	Not an evaluation, but strong on impact analysis
	State of Food and Agriculture	2011	Not an evaluation, but strong research and analysis on the gender dimensions of agricultural development
	UN Women Action Plan 2011-13	2011	Results framework for the UN for gender equality

This report does not intend to rank the eleven evaluations in terms of quality, which is difficult given the differences in mandates and support provided by each agency. It is also beyond the scope of a desk review. The assessment aims to glean lessons from comparative review of these 11 evaluations

and provide a few recommendations to the ECG for discussion and consideration among its members. Throughout the text, the acronym of the evaluation unit is used when the responsibility lies with the evaluation unit (as listed in table 1). Otherwise, the acronym of the agency is used.

3. <http://go.worldbank.org/CQCTMSFI40>.

Lessons learned

Several lessons that emerge from these evaluations are worth noting. Multilateral and bilateral agencies need to take steps to strengthen the clarity of their gender policies based on the findings of the evaluations. They also need to strengthen the implementation of their respective gender policies through developing clear statements of expected results (including targets and indicators) at output and outcome levels, implementation plans and processes, ensuring availability of adequate resources, and enhancing staff skills. Most importantly, monitoring and evaluation systems need strengthening – and this includes systematic monitoring of inputs, outputs, and at least intermediate outcomes, as well as monitoring that activity completion reports integrate gender considerations. Such reporting will help to increase accountability as well as to facilitate fine-tuning of support to enhance development effectiveness. Dissemination of good practices and recognition of staff performance are also needed to demonstrate management commitment to gender equality. The lessons in this section draw particularly on the findings of the AfDB synthesis and are corroborated by similar findings in the Office of Internal Oversight Services Inspection and Evaluation Division's gender thematic evaluations in the UN Secretariat.⁴

Other lessons obtained from these evaluations are organized around the following dimensions: (i) Gender Policies and Policy implementation, and (ii) Results of Agency Support.

Lessons about policies and policy implementation

POLICIES

The evaluations indicate that only 8 of the agencies have a gender policy.

Three of the agencies have a gender action plan. Approaches of all but one of the 11 agencies continue to give priority to gender mainstreaming – an approach enshrined in

the 1995 Beijing Declaration and Platform for Action and subsequent UN policies, although a number of agencies also support women-specific programming.⁵ In 2010, the Inter-American Development Bank (IDB) went a step further and treats gender as a safeguard policy. This means that in addition to “proactive action, which actively promotes gender equality and the empowerment of women through all the Bank’s development interventions”, the policy also requires staff to take “preventive action, which introduces safeguards to prevent or mitigate adverse impacts on women or men due to gender resulting from the Bank’s actions through its financial operations”. On the other hand, WB has narrowed its broader gender mainstreaming policy to a country-level mainstreaming requirement in 2003. This approach requires diagnosis of gender issues at the country level, identification of priority sectors and areas for gender equality in the country, and selective gender integration at the sector or project level only in sectors and areas identified as priorities in the country assistance strategy.⁶ Thus, among these 11 agencies, IDB has the broadest and most rigorous gender approach, while the World Bank has the most selective approach.

Some evaluations examined the relevance of gender equality and women’s empowerment for effective poverty reduction and economic growth. They indicate a need for stronger evidence to explain the observed correlations among gender equality or women’s empowerment, poverty reduction, and economic growth, since the availability and quality of existing data is inadequate to explain the dynamics and the causality of these correlations. A literature review undertaken for the IEG evaluation finds that the evidence on the extent to which gender inequality is linked to growth outcomes is tentative at best. The reliance on cross-country regressions in some of the analyses during the late 1990s is problematic for many reasons. The evaluation notes that recent reviews highlight the conceptual, methodological, and data challenges faced in seeking to establish empirical— let

4. The Office of Internal Oversight Services evaluation can be found at <http://www.un.org/Docs/journal/asp/ws.asp?m=a/65/266>.

5. Women-specific programming is defined differently by different organizations to include (i) support to women organizing for their rights through policy advocacy, research, and mobilization and/or (ii) affirmative action to ‘level the playing field’ when there is evidence of gender bias (e.g. providing literacy training, leadership training, or entrepreneurial training for women, or undertaking special efforts to recruit women). In some organizations, this type of ‘affirmative action’ is understood to be part of gender mainstreaming.

6. A limited number of WB policies such as that on involuntary resettlement, indigenous peoples, community forests, and development cooperation require project level consideration of gender issues.

alone causal—relationships among gender inequality, poverty, and growth. However, the evaluation concludes that despite the methodological challenges, the literature suggests that gender-based differences in health, education, access to economic assets, and voice affect overall economic choices and opportunities for both men and women. They affect the labour productivity of men and women, the performance and potential of their businesses, and the incentives they face as economic agents. These differences also have longer-term, intergenerational effects, influencing the education, welfare, and future economic potential of children. All these factors can affect the nature, pace, sustainability, and impact of economic growth and poverty reduction— and thereby the effectiveness of development interventions in achieving their goals.

Some evaluations also find that demand for gender projects was low in client countries. For example, both IFAD and FAO evaluations found that progress on gender equality was more difficult in the Middle East/North Africa and in countries of the former Soviet Union. The AfDB evaluation finds that “inconsistent ownership of gender within partner governments, which results in it being given a low priority in poverty reduction strategies and country policy dialogue” is a significant barrier. The IEG evaluation finds that despite the formal acceptance of gender equality or women’s empowerment by client countries, 87 percent of 167 Bank staff respondents to an IEG survey identified the lack of demand in client countries as the most important constraint in addressing gender issues in their work. However, it also notes that several client stakeholders pointed out that this demand depends on whom Bank staff consult in countries. They stressed the need to ensure wider participation (outside the Ministry of Finance and other line ministries) in this dialogue (IOE, OED and SDC had similar findings). Additionally in other agencies, lack of client demand did not completely absolve staff from integrating gender into their support, because gender mainstreaming is considered an integral element of development support and gender policies mandated such integration.⁷ More recently, a few countries have been pressing donors to conform to their agenda

for gender equality. Nicaragua, Ethiopia, and Rwanda are most often cited as examples of countries where central planning ministries insist on attention to gender equality in donor and ministry initiatives.

One key lesson observed by several evaluations (ADB, IFAD, FAO, UNDP, WB) is the importance of a clear results framework⁸ if gender policies are to be effectively implemented. Only a few of the gender policies include or are linked to a results framework. As the ADB evaluation notes, because of this gap policy implementation is input-cum-process oriented and lacks defined outputs and outcomes that could be monitored and measured against targets. The UNDP evaluation finds that the lack of a clear results framework (objectives, indicators, and targets) makes it difficult to assess results. The CIDA evaluation finds that while the gender equality policy sets out clear and focused objectives and provides a conceptual blueprint for action, there is not an explicit contemporary Agency-wide gender equality strategy or action plan for implementing the policy. The IEG evaluation finds that the absence of a results framework in the Gender Strategy weakened country-level results or outcomes. Interventions were focused on outputs rather than on outcomes, and the lack of a considered framework resulted in inconsistent attention to gender issues, particularly outside the human development sectors. Therefore, in many cases, any result or work associated with women is considered to have contributed to gender equality, although a few evaluations draw out examples where work associated with women either did not contribute to gender equality or had perverse effects (IFAD, SDC, WB). IOE notes that recent country-level strategies (country strategic opportunities programmes) tend to have stronger gender dimensions in their results frameworks.

Although there has been no inter-agency comparison of relative success in contributing to gender equality, there is some evidence, based on practitioner perceptions of who is ‘leading’ on gender equality among donor agencies, and based on systematic client-disaggregated monitoring of gender in design by FAO’s Investment Centre, that agencies with

7. As noted earlier, the World Bank follows a policy of selective integration and if the country assistance strategy does not integrate gender considerations, then staff need to integrate gender considerations only in a highly limited set of cases.

8. Some agencies reflect confusion between a results framework and a monitoring framework. A results framework states the desired outcomes and outputs, and helps the project designer to design appropriate activities and identify inputs that are likely to contribute towards the desired outcomes. A monitoring framework focuses on results and assesses whether and how the inputs and activities are leading to the desired outputs, intermediate outcomes and impact.

clearer policies and results frameworks are better able to capture results and are therefore perceived to be better able to learn from and refine their programmes. As a consequence of these findings, a number of the organizations evaluated are revising and updating their gender policies and associated results frameworks and implementation plans (WB, FAO, IFAD, ADB). Although beyond the scope of this review, a more detailed comparative analysis of the old and new (post-evaluation) policies and linked results frameworks/action plans would be very useful for an organization aiming to revise and strengthen its gender policy.

The WDR, drawing on a wide range of research, proposes a ‘theory of change’ that describes the dimensions and levels of gender equality, and how these

dimensions and levels interact to create change (see figure 1). The report presents this framework as one example to illustrate that this level of articulation is very useful for explaining and clarifying the assumptions in gender policies and results frameworks, and for testing the adequacy of policies during implementation. Most of the policies and results frameworks, however, fell short of this standard. In analysing results in the three dimensions, the WDR’s analytical framework enhances understanding of where the constraints lie (some of them may lie across dimensions and/or levels) and how policies can influence changes in and across the three dimensions to strengthen progress towards gender equality and women’s economic empowerment. Gender equality cannot be achieved by addressing only one or two of these dimensions. Inequalities need to be addressed in all dimensions.

Figure 1. WDR analytical framework to analyse dimensions of gender equality



*Accumulation of endowments relates to ensuring that women obtain support to exercise fully their human and physical endowments with special reference to education, health, land and financial assets. Access to sexual and reproductive health services is part of health. (Some frameworks describe land and financial assets under ‘access to economic opportunities’ [Harvard, Moser]. Freedom from violence and recognition of human rights is also sometimes described under endowments, sometimes listed separately [e.g. UN rights-based approaches].)

**Access to economic opportunities relates to how men and women are able to utilize the endowments and whether or not there are systematic differences by gender in economic opportunities and returns to these endowments, focusing on how individuals allocate time and resources to reap returns on their endowments, mostly in labour and asset markets. (Some frameworks include constraints on time [work burden] and mobility in this dimension [Harvard, Moser]. Some frameworks explicitly consider the trade-offs between the productive economy and the care economy [Harvard, Moser].)

***Agency (voice and participation) is about the final dimension relating to the differential agency of men and women—their abilities to make choices and take actions, focusing on voice, political participation and mobility. [Molyneux, Moser].

The WDR identifies these dimensions interacting within the household, in the community, and at a systemic/institutional level. [cf. Gender@Work framework (See www.genderatwork.org/gender-work-framework/)]

POLICY IMPLEMENTATION

Several evaluations (ADB, FAO, IFAD and WB) find that the policy commitments of agencies towards gender equality were not reflected in their institutional arrangements, which needed substantial strengthening in a number of areas, if gender equality is to be supported. All but one of the 11 agency policies have adopted gender mainstreaming as the main method to achieve gender equality.⁹ The AfDB review of 26 evaluations finds that all evaluations point to a similar finding: gender mainstreaming is a complex undertaking that has not been widely carried out by the development community. This assessment also finds that gender mainstreaming has in general not been very successful in delivering results, and that there has not been a serious effort at implementing gender mainstreaming appears in the respective agencies. These findings are grouped under the headings of Accountability, Institutional Arrangements, Performance Monitoring and Partnerships.

Most evaluations continue to value gender mainstreaming as a means to contribute to gender equality. The IEG evaluation takes the view that the WB must return to a position of gender mainstreaming at the project level, although given the limited resources, it finds some merit in focusing on countries where gender inequalities are high. Several other evaluations note that gender mainstreaming does not work because of the lack of adequate human and financial resources to make gender mainstreaming effective (UN-HABITAT, IED, SDC, UNICEF). On the other hand, the AfDB evaluation suggests four options for consideration in lieu of gender mainstreaming, given the lack of effective implementation. The four options presented by the evaluation are as follows:

- Gender focusing: This approach would focus on those sectors where gender equality appears to be sufficiently embedded and has made some progress, and attempt to create linkages with related sectors.
- Women in development plus: Many of evaluations have found that, in practice, when gender equality is integrated into country-level interventions, it is boiled down to a women-centred or women's

empowerment approach. This option would make this approach strategically explicit, building on the experiences that did deliver results, but would incorporate more fundamental analysis of gender power structures, and seek to position interventions to empower women economically and politically. This would entail a return to a concentration on interventions that empower women (and men where appropriate) and facilitate incremental social change over time.

- Policy dialogue on gender equality in new aid modalities: Gender mainstreaming and women's empowerment have been aimed, at the operational level, at traditional project-based modalities. However, the increasing use of new aid modalities has created an additional set of challenges for integrating gender equality into interventions. Within the context of option 1, donor organizations and partner governments could enhance the consideration of gender equality and women's empowerment by focused policy dialogue and development in key sectors, such as education and health, private sector development, and the investment enabling environment. Policy-based lending could be used to enhance gender dialogue around reforms in specific sectors such as finance and extractive industries. SWAPs could continue to build from a position of strength in the education and health sectors to support longer-term enhancements in human capital and equitable economic development.
- Improving results reporting and learning through more systematic integration of monitoring and evaluation: Gender monitoring would be strengthened in those sectors where gender equality and women's empowerment are the focus. Gender could be integrated more broadly across all evaluative activities to uncover unintended results, enhance cross-sectoral learning, and reduce blindness.

ACCOUNTABILITY

Greater commitment and leadership of senior management to gender equality is seen as necessary. Several evaluations (FAO, IFAD, SDC, WB) as well as the benchmarking exercise in the IOE evaluation and the

⁹ The only exception as discussed earlier is WB. However, the IEG evaluation recommends that WB should reinstate a broader policy in this regard with the project as an additional entry point to the mandated country-level entry point.

AfDB evaluation synthesis¹⁰ found that a common lesson across evaluations is that consistent leadership and follow-up by senior management and executive boards is important in ensuring support for gender equality. Most evaluations suggest that one effective way of ensuring such accountability is to establish a systematic review process to assess progress in policy implementation and achievement of desired outcomes. While most gender policies recommend such oversight, almost all evaluations found that this is lacking. Neither senior management nor governance bodies systematically review efforts or results (CIDA, SDC, WB). Evaluations recommend tightening accountability of senior managers for policy implementation. The IEG evaluation compared the processes required for implementation of safeguard policies and its gender policies demonstrating the weakness in the latter case. Since gender equality, like environmental sustainability or knowledge management and policy dialogue is a cross-cutting issue, and such issues are not consistently the main focus of leadership, periodic review at the most senior level is seen as necessary to ensure adequate attention is regularly devoted to progress on gender equality by mid-level managers and staff. FAO, as a follow up to evaluation recommendations, ensured that action on gender equality was part of each senior manager's performance contract.

In addition to senior leadership commitment, stronger accountability at the managerial and staff levels is necessary to strengthen the implementation of support for gender equality. The need to shift from a 'permissive' environment or a personal commitment to gender equality is stressed by many evaluations. Staff who see the importance of mainstreaming gender equality in their programming are permitted or encouraged to do so, although strong performers report that they are seldom recognized for their efforts. At the same time, those who do not make any efforts to integrate gender dimensions in their work do not face any negative consequences (AfDB, IFAD, SDC, WB). Strengthening accountability by developing gender-related programming frameworks (at the country level or the sector level), and requiring gender-related objectives

and tasks in individual performance agreements, especially for managers, are seen as measures to strengthen performance (FAO, IFAD, SDC, CIDA). Where agencies set targets (ADB, IFAD, CIDA) they are found to be useful and the evaluations encouraged their refinement and expansion. Examples of targets included quality-at-entry for attention to gender equality (percentage of proposals meeting quality targets) as well as at subsequent milestones in the project cycle, and programme expenditure targets. The DFID evaluation notes the importance of internal gender champions, and the SDC evaluation finds that effective implementation of the gender policy depends on leadership of departmental, divisional, and other managers.

INSTITUTIONAL ARRANGEMENTS

Institutional arrangements for implementation are different among agencies, but evaluations do not have adequate data to determine what type of arrangements worked better. Most agencies have gone through different stages and models in institutionalizing gender integration, according to the United Nations Human Settlements Programme (UN-HABITAT) evaluation. Some institutions have a more centralized arrangement, with a central gender unit that classifies projects according to a gender classification and monitors results. Others have a very decentralized arrangement with a central unit more focused on policy issues, gleaning lessons at a corporate level, and collating the results for senior Management. The UN-HABITAT evaluation notes that high-level political commitment appears to be the variable that makes these varied institutional arrangements effective in making gender equality and mainstreaming most visible and critical. The UNICEF evaluation also makes the point that for gender mainstreaming to work effectively, there is a need for significant transformation in the way an institutional works and a major allocation of human and financial resources.

There is no agreement on adequate staffing levels to support effective contributions to gender equality. The evaluations generally recommend increased gender-specialized human resources for planning and design, quality assurance, monitoring

10. Many of the findings reported in this chapter are documented in greater depth in AfDB's evaluation synthesis of 26 gender evaluations, *Mainstreaming Gender Equality: A Road to Results or a Road to Nowhere? An Evaluation Synthesis*, May 2011, OPEV, African Development Bank.

and evaluation, and for support during programme implementation. However, a number of evaluations report that when task managers and programme staff are convinced of the usefulness of integrating gender equality dimensions, allocating human and financial resources is not a constraint (SDC, CIDA, IFAD, FAO), especially if this can be done at country or regional level. There is a risk in over-emphasis on the use of casual or consultant resources, with the consequence that the knowledge and experience gained may not be well integrated into the organization's ongoing practice and systems.

Agencies also varied widely in their investment in specialist gender resources, in terms of quantity, seniority, location, and investment in learning. There is inadequate data to assess what is 'sufficient' to resource good results. Nor is there guidance in the evaluations for the level of programmatic resources required to deliver gender equality outcomes (for example, additional design costs, or implementation and monitoring costs, which are often coded differently by different units). CIDA designated a budgeted amount for 'gender funds', and the IOE evaluation recommended that the Finance Department undertake an analysis to develop some 'rule of thumb' guidance on the additional costs of programme or project/level investments required to achieve credible gender equality results. Better benchmarking in both of these areas (organizational infrastructure and programmatic investment) would be helpful.

Some evaluations point out the need for clear guidance to, and training of staff, on how to mainstream gender into development support. A number of evaluations report confusion around gender-related concepts (people using the same terms with widely different understanding in the same institution) and difficulty in moving from understanding concepts to applying them to programme design and implementation at various stages. A number of evaluations find that managers rarely attend gender training and that gender is not integrated into orientation training of new staff appointed by the agency. The IED evaluation made the important point that as agencies

shift focus in terms of sectors (a shift to large infrastructure and private sector development in the case of ADB), traditional approaches are likely to be insufficient. The evaluation of CIDA support notes different levels or types of staff may need different levels of awareness and expertise. As the AfDB synthesis notes, there is little assessment of what 'training needs' are, and there is a tendency to deliver one-time, one-size-fits-all, one-shot supply-based training. The evaluations do not discuss or assess the comparative cost-benefit of coaching or 'just-in-time' training, or attention to gender in other types of training (for example, training on results). There is little assessment available of what type of training has been most effective for what purpose. Without a stronger evidence base of training needs assessment, or evidence on the effectiveness of different types of training, it is difficult to design or justify a significant investment in training, despite the findings of the confusion in concepts and inability to apply them.

Some evaluations find that a combination of factors, if adhered to by task teams, appear to enhance policy implementation as well as achievement of gender equality objectives. These are: (i) an initial gender analysis, (ii) clear gender equality objectives/outcomes and associated resources; (iii) planning or a road map around outcome-focused activities; and (iv) specific and relevant sex-disaggregated or gender-related indicators (SDC and WB). Quality at entry reviews that include a meaningful review of attention to gender equality, particularly when done early in the design process, are also seen as useful (IFAD).

Although there is no proven relationship between gender balance in staffing and greater contribution to gender equality,¹¹ such a focus within the institution reiterates commitment by the agency. As the CGIAR notes, on its gender and diversity website, "We cannot achieve on the outside what we do not practice on the inside." The evaluations deal with the issue of gender balance in staffing differently. Some evaluations do not discuss this issue since it is related to civil service staffing policies (CIDA) and extrinsic to the evaluation. Other

11. Although in the private sector there is a correlation between better overall performance and greater gender balance at senior levels of an organization. See Catalyst, 2007.

evaluations see gender balance in staffing as a major pillar of the agency's commitment to gender equality (SDC).

Gender balance in staffing is one of the easiest issues to track, and considerable work has been done in the leading organizations on HR data that allows

managers not only to track progress, but also to diagnose where problems exist. This type of tracking can be done by department or unit, as well as by stage in the process. Figure 2 below sets out some of the good practice found in the evaluations. When the ratios are not balanced (i.e. close to 50/50) there may be a need for remedial action.

Figure 2. Measuring gender balance in staffing

1. Gender balance by level in the organization, by location, and by department (including the use of consultants or other non-permanent staffing where relevant)
2. Gender disparities in staff satisfaction by level, location and department
3. Existence of clear policies in favour of gender equality, against harassment, and in support of work-life balance.
4. Gender disparities in the use of anti-harassment and work-life balance policies and procedures.
5. Gender disparities in promotion, professional development opportunities, or salary levels for women and men in similar positions.
6. Systematic tracking and review of gender ratios in the recruitment process (gender balance of applicants, of long lists, short lists, interview panels, recommendations, decisions).
7. Use, timeliness and perceived fairness of harassment resolution processes.
8. Gender balance in resignations or other departures.

PERFORMANCE MONITORING

The need for systematic and effective monitoring systems is an important lesson across evaluations. As the SDC evaluation notes: "if gender equality is not measured, then it will not change." The lack of effective and relevant monitoring constrains evaluation of results and, thus constrains accountability. Self-assessments, such as project completion reports, were also not effective in integrating gender dimensions in analysing results.

Several evaluations report the need to monitor that the emphasis on gender during the start-up of the programme cycle (programme and project frameworks and design) is sustained through project implementation (ADB, AfDB, CIDA, WB). Proposals for funding may require the inclusion of a social/gender analysis to meet quality-at-entry requirements, but there may be no further systematic requirements relating to implementing, monitoring or evaluating gender equality dimensions throughout the project/programme cycle.

The quality of the results orientation in general, and with regard to gender dimensions in particular, influenced the agency's ability to track progress and achieve intended results. The AfDB evaluation notes that one of the most common findings reported by the evaluations has been the lack of monitoring and evaluation (M&E) and supervision systems within donor organizations to track progress, allow for adaptive management, record gender equality results, and document good practices. A noteworthy exception in this regard is IFAD's Results and Impact Management System (RIMS), which includes a compendium of indicators at the output, outcome and impact level that serve as a starting point for developing results, and includes measures for progress in gender equality.

PARTNERSHIPS

Both gender-related advocacy and partnerships (with government institutions, with other development cooperation agencies, and with civil society organizations) need strengthening (AfDB,

FAO, IFAD, SDC, UN-HABITAT, WB). Related to this are concerns that newer aid modalities (SWAPs and budget support) could further weaken attention to gender equality. An exception seems to be in countries where the government itself is taking a strong position in favour of gender equality through its planning and finance Ministries and tracking donor inclusion of gender dimensions.

Some good practices are reported on how partnerships have led to generation of sex-disaggregated or gender-related data at the

global level. Some donors are promoting a global collaboration in producing and using data on women. These efforts have increased the availability of data during this evaluation period relative to the earlier evaluation period. Some partnerships are also noted among the multilaterals in relying on each other's gender diagnostics. Thus, in East Asia for example, there is an understanding between ADB and the WB on the countries where each would be responsible for gender diagnostics. In Nepal, DFID undertook a country gender assessment that has been widely used by other donors.

Box 1. Good practice in partnerships

UN-HABITAT's partnership with the Gender Water Alliance (GWA) is noted as a good practice in that it helped to provide specialist advice in water and sanitation in 17 sites across 14 countries. Through GWA, the UN-HABITAT program benefitted from a network of specialists across Africa. Recognizing that gender mainstreaming needed strengthened in all program cities, the GWA hired a water and sanitation specialist in each country to provide support and advice in programme implementation.

Another emerging pattern that could be adopted more widely is collaboration among agencies with different mandates. For example, in IFAD projects where illiteracy is a major constraint to women's economic empowerment, collaboration with education ministries or literacy training through NGO partnerships provides needed complementary support. The same collaboration is found among some SDC partners.

Lessons on results

These lessons on results are organized around the following framework: (i) strengthened human development; (ii) increased access to and ownership of economic resources; and (iii) improved voice and decision-making. In addition, some lessons that span areas of synergy between the three sets of outcomes are also provided. The evaluations show that (with the exception of IFAD and FAO, because of their specialization, and the bilaterals, whose contribution to human development tends to be through multilateral institutions), the gender focus is on human development, and that areas like infrastructure investment

and transport, which are also shown to be able to contribute to gender equality, receive much less attention. The IEG evaluation finds, however, that there is significant improvement both in quality and quantity in addressing gender issues in the infrastructure sectors during this decade when compared to that in the previous decade.

STRENGTHENED HUMAN DEVELOPMENT

The evaluations find the greatest success in diminishing the gap between women and men in access to education. The increased number of women in primary, secondary and tertiary education reflects a large social change in a relatively short period of time. (WB, ADB, WDR). This change, as noted in earlier research by the WB and others, leads to improved family wellbeing (health, nutrition, education of children, income). In some countries, the success of keeping girls in school has been such that boys are now falling behind in taking up tertiary education. However, several evaluations found that while access had increased, the quality of services was not keeping pace.

Completion and dropout rates were not improving in all cases.

A related success is that when countries decide to adopt family planning, it has resulted in a rapid decrease in family sizes: a decline that took over 100 years to achieve in the USA took 20 years or less in countries as varied as Morocco, Bangladesh, Zimbabwe and Iran (WDR). The IEG evaluation finds that narrowly focusing Bank support on interventions in the health sector has in many cases limited the achievement of desired health outcomes. As an example of the benefit of a broader focus, increased transportation resulted in larger numbers of women seeking medical help in Peru. Training local women as health attendants may help, owing to the high likelihood that they will go back to their communities to provide services (as in the Republic of Yemen). It also finds that costs are not always the only reason that health services are used too little. Lack of awareness of the need for women to use services can also be a problem in some countries, and there is some evidence that a door-to-door awareness-raising programme may be necessary, in combination with the provision of facilities, trained female personnel, and supplies.

INCREASED ACCESS TO AND OWNERSHIP OF ECONOMIC RESOURCES

The evaluations find that while there had been some gains in increasing women's agricultural productivity (IFAD, FAO) there remains significant room for improvement. FAO's flagship publication, *The State of Food and Agriculture*,¹² in 2011 focused on gender disparities and reported that women farmers' productivity would increase by 20% if they had equal access to agricultural resources, increasing food production by 2-4% - enough to free an additional 150 million people from hunger.

Many evaluations focus on measuring poverty reduction (the economic opportunities dimension), although there is little systematic evidence about the distribution of the additional income, partly due to the lack of data in the agency records. This focus is particularly prominent for agencies that did not give priority to human endowment investments. The evaluations

do not look systematically at evidence of workload reduction (or increase or transfer to other family members), although this is tracked in some instances by IFAD, especially in the provision of household water supply and rural transport.

Skills development and new technology for women is successful when aligned with economic opportunity and relevant for women. For example, women adopt new agricultural technology when they are in paid agricultural employment, and the development of processing skills to meet quality standards in value chain improvements have been successful (IFAD). On the contrary, the first IEG evaluation found that fewer women than men trained in aquaculture farming could benefit from the skills gained due to lack of land and access to water resources needed to establish such activities. Evaluations therefore find that it is important to design interventions for economic empowerment that take into account women's existing realities and perceptions, with regard to time, space in the house, health, other obligations and responsibilities, and constraints in movement, although the clear intent is to reduce these constraints. This approach to design would require the participation of beneficiaries in the planning of activities. (SDC, FAO, WB, IFAD).

The UNDP evaluation found an assumption in many activities that mere provision of microfinance for women would lead to gender equality; this was not true without additional technical and business support. The IEG evaluation found that while microfinance may increase the social empowerment of women, it is difficult to establish income-generating activities without other support, especially in remote areas where markets are inoperative. In a similar vein, the UN-HABITAT evaluation found that women from poor households do not have incomes for most housing finance schemes, but that despite their problems women tend to be more reliable borrowers than men.

IMPROVED VOICE AND DECISION-MAKING

All agencies attempt to enhance women's voice and decision-making in development activities, but with only modest levels of

¹² <http://www.fao.org/docrep/013/i2050e/i2050e00.htm>.

success. The evaluations provided limited evidence of good indicators or systematic monitoring of more gender-equitable decision-making, whether at the household level, the community level (where most of the data that is available is collected) or in formal government bodies. While UN agencies also focus on parliaments and legislatures, the development banks and bilaterals tend to focus on participation in decision making at the community level. However, the WDR does note that the influence of quotas for women in Indian local government structures has been to increase investment in long-term community improvements like water, schools and roads, since women give priority to such investments. IFAD measures whether local institutions like farmer organizations improve the balance between women and men in their leadership structure throughout the life of their programmes, but in many organizations, the information was collected only during the evaluation phase and therefore relies on beneficiary memory.

A few evaluations find that women's participation in community committees improved their participation as beneficiaries and provided them a space for participation outside their households. This is seen as a very important first step in more conservative communities such as in Afghanistan or Yemen or in part of India or Pakistan. The evaluation of SDC support also found that often women's participation in such fora is constrained by men especially when the public spaces for participation are far away from where women live and work and also result in high transportation and time costs. To sustainably influence gender relations and empower women, long-term support is needed, as well as carefully designed mechanisms that will address gender imbalances in participation, rather than strengthen existing gender stereotypes, for example by assigning toilet cleaning to women. There are two recent initiatives that provide systematic measures of changes in women's voice and decision-making. CARE has developed and tested measures of women's household-level empowerment in Bangladesh, and USAID/IFPRI/Oxford Poverty and Human Development Initiative have piloted measures of women's empowerment

in agriculture in several countries. These measures could usefully be considered by ECG members and their agencies.¹³ The UN-HABITAT evaluation found that making changes at the local level is challenging because this is the "level most associated with prejudices and conservative culture." Domination by local elites also makes it difficult for poor people. The first IEG evaluation found that this was an issue not only for women, but also for men from socially disadvantaged backgrounds. The UN-HABITAT evaluation, therefore, cautions against the assumption that a decentralization agenda is a positive step for gender equality or for women's empowerment without adequate safeguards in place.¹⁴

SYNERGIES BETWEEN THE THREE DIMENSIONS

The WDR shows that there are multiple linkages among these major dimensions. Increased education can influence economic success, and, conversely, the availability of jobs for young women can influence the decision to invest in education for girls, as can very modest levels of conditional educational or health care subsidy. Similarly, economic success can increase participation in community and family decision-making, but the converse is also true. Policy and programming decisions can impede or support more equitable distribution of opportunity among women and men.

Several evaluations, however, found that gender integration, even at a country level, is fragmented and sectoral at best. Additionally, the lack of a common results framework meant that synergies between sectors (which is very important to achieve gender equality) were totally lost, and interventions remained ad hoc at a project level. For example, the UNDP evaluation found that despite the devastating effect of HIV/AIDS, the UNDP policy and practice had not developed an inter-sectoral approach. The IEG evaluation found that without sufficient transport facilities and some all-weather roads, provision of health facilities in remote areas could not contribute to reducing maternal mortality.

13. See CARE Canada's Reaching New Heights, The Case for Measuring Women's Empowerment at care.ca/sites/default/files/files/publications/IWDreport-med.pdf and The Women's Empowerment in Agriculture Index published by USAID's Feed the Future.

14. The WDR, however, found that changes in conservative culture were possible with even very modest incentives, for example, in the case of Turkish subsidies for girls' education.

Several evaluations find that not taking gender into account undermines the effectiveness and sustainability of many projects and initiatives (FAO, WB). The evaluations and the WDR research increasingly show a correlation between prosperity and gender equality, although there is no evidence of a causal relationship. Similarly, several of the evaluations (WB, IFAD) have noted a correlation between overall project success and better results in promoting gender equality.

The WDR notes that progress in some areas meets greater resistance. Areas such as women's reproductive health, gender-based wage differentials, voice in decision-making, and the unequal sharing of household work remain 'sticky' – resistant to change - while in others, such as education and labour force participation, the gap between women and men is shrinking rapidly. Some of this resistance to change may be the result of neglect, or of institutional systems and procedures that do not give adequate attention to defining clear results or tracking progress. Issues relating to the latter are discussed in the following section.

Comparative review of the corporate gender evaluations

Background and context

This chapter sets out the findings of a rapid assessment of eleven thematic gender evaluations undertaken by evaluation units in the development agencies listed in table 1. Based on a rapid review of development literature (see references), the elements of a strong gender-aware evaluation are listed in table 2 below. This section assesses the evaluations against these dimensions and reports the findings.

Overall, the assessment finds scope for strengthening of evaluation methodology in three important areas. First, to be effective and transparent, evaluations need to assess against a clear results framework for gender equality, which is adapted to agency and country priorities. Second, there is a need for a consistent approach to gender mainstreaming, the main approach utilized to achieve gender equality. Finally, gender evaluation frameworks need strengthening to ensure that the evaluation focuses on results relevant for gender equality.

Table 2. Elements of a strong thematic gender evaluation

Dimension	What the assessment looked for
Clarity of evaluation objectives and associated results framework	<ul style="list-style-type: none"> • A clear results framework for gender equality or women's economic empowerment
Conceptual clarity and consistency in gender terminology	<ul style="list-style-type: none"> • Clarity in gender mainstreaming • Analytical frameworks for assessing results
Consistent and clear evaluation framework on how it will address or integrate gender issues	<ul style="list-style-type: none"> • Scope of the evaluation and focus on results • Conceptual clarity of evaluation questions • Use of gender-aware indicators with rating systems assessing changes to gender equality/equity and/or women's empowerment

CLARITY OF EVALUATION OBJECTIVES AND ASSOCIATED RESULTS FRAMEWORK

Strong evaluations evaluate against institutional or program objectives. However, gender equality is not a clear objective by itself (see box 2). It is, therefore, necessary to further define the term 'gender equality' by clearly explaining in generic terms the outcomes that an organization should contribute to in order to support gender equality, even when the agency has not defined the term. Support for gender equality is otherwise reminiscent of the famous quote by Lewis Carroll: "If you don't know where you are going, any road will take you there."

The lack of such elaboration of the term makes development of evaluation metrics very challenging for all 11 evaluations. Any outcome (related to women) is considered by the evaluation as contributing to gender equality. There is some evidence, for example, that successful microcredit schemes in Bangladesh are socially empowering women but as a result households had income to pay higher dowries allowing women to marry into "better" households. While the benefiting women appreciated this, it is likely that this strengthened gender stereotypes rather than contributed to greater gender equality. Without a clear results framework, it would be difficult for an evaluator to conclude whether the above-noted contribution supports gender equality.

Box 2. Results framework

Based on its benchmarking efforts, the IOE evaluation finds: “Neither IFAD nor the comparator agencies have developed a clear statement about their understanding of how to bring about gender equality and women’s empowerment in different country contexts.” There is need to further expand the outcomes and identify the intermediate outcomes, outputs, activities, and inputs that would contribute to the achievement of the desired outcomes. A monitoring framework needs to be nested in, and guided by, the results framework.

This does not mean that every evaluation has to adopt the same framework. The framework would only represent a broad and universal vision of how gender equality can be achieved. It would be applicable in all countries. The CIDA framework, for example, focuses on three dimensions: Equal participation of women with men as decision makers; Women and girls more able to realize their full human rights; and equal access to, and control over, the resources and benefits of development.

Broadly, the same outcomes would be needed in any country, whether it is Afghanistan, Canada or Tanzania. The areas and method of engagement would need to vary depending on country context. The intermediate outcomes sought and the outputs desired will depend on the country context. All dimensions may also not be important for each agency. Given that the MDBs do not typically follow a rights-based approach, they could instead look at reduced gender disparity in the human development dimension and review whether such support was leading to strengthened human development for both women and men.

IOE and FAO evaluations nested their evaluation frameworks around women’s economic empowerment, consistent with the agency’s approach. They then examined whether the desired outcomes were generated.¹⁵ Given their results framework, it would then be possible for the evaluation to confirm if the agency was contributing to gender equality. Without one, as the SDC evaluation noted: even where gender/sex disaggregated data are being collected and reported, the “question is: what are the desired outcomes of SDC’s gender equality mainstreaming? Until the Pakistan program is clear on its desired contribution to gender equality in its various sectors, it will be difficult to move beyond collection of disaggregated information at the input and output level.”

Where there was no clearly enunciated results framework, as in the case of the WB, IEG formulated a results framework based on implicit statements in the WB Gender Strategy of 2001: (i) strengthened human development for men and women (reduced disparities without losing the gains for men or boys); (ii) equal access to economic

opportunities and assets; and (iii) equal voice and participation in development. Such a framework helped to better understand where the gaps lay. The evaluation found that WB support had contributed to the first dimension, and expanded its support for the second and third dimensions but without equally commensurate and sustainable results in several focus countries. Overall, it concluded (using the Gender-related Development Index and other indicators to measure progress in gender equality at the country level) that the WB had contributed substantially to gender equality in 4 of the 12 focus countries, modestly in another 6, and weakly in 2.

The results framework also helps the evaluation to test whether the institutional arrangements are helping to facilitate the achievement of results. The framework provides the parameters to analyse the kind of skills needed in any institution, prioritize the types of analytical work that is being undertaken, and assess whether gender strategies are likely to lead to desired outcomes. It also allows the evaluator to understand whether the appropriate

15. FAO outcomes included: (i) improved well-being and eased workloads by facilitating access to basic rural services and infrastructures; (ii) physical empowerment (full control over one’s own body, sexuality and fertility or ensuring enough food for a healthy body); (iii) economic empowerment leads to equal access to and control over means of production and economic independence in its projects dealing with production and/or management of natural resources (or increased income through more and more diversified production and related marketing as well as access to and control over means of production.) (iv) political empowerment (political self-determination and the creation of a power base in a self-determined direction) or involving the final users in the decision making on projects themselves or in the management thereof); and (v) socio-cultural empowerment leads to one person’s own identity, worth and self-respect (particularly related to re/integration of refugees in their old or new communities.

activities are being pursued that will likely lead to the desired results. For example, FAO projects are analysed along its four dimensions of empowerment, and the evaluation concludes that FAO support is mostly focused on physical and economic empowerment rather than the other two dimensions. Without a clear framework, anything to do with women (or with men) could be considered as contributing to gender equality. Additionally, such a framework allows evaluations to better identify risks to achievement of stated objectives and identify areas where there are gaps. Few of the evaluations, however, focused their institutional analysis around such outcomes.

objectives. Nevertheless, such an approach weakens evaluation findings. As early as in 2002, OECD documents noted that “problems will arise if evaluations are designed based on the assumption that mainstreaming automatically leads to gender equality outcomes. Institutional mainstreaming should not be evaluated without considering the extent to which this leads to changes in gender relations. Focus should be put on results as well as processes and institutional practices.”¹⁶ The AfDB synthesis report reiterates the point through its evocative title: *Mainstreaming Gender Equality: A Road To Results Or A Road To Nowhere?*

CONCEPTUAL CLARITY AND CONSISTENCY

Given that gender mainstreaming was the most common approach (whether it was at the country, sector, or project levels), each evaluation assessed the extent of gender mainstreaming. First, some evaluations treat gender mainstreaming as a goal, others as an instrument. Second, the evaluations define gender mainstreaming differently, decreasing the common lessons that can be gleaned. Third, such lack of clarity in the term may be one contributory cause to the resistance to gender mainstreaming found by the evaluations in all agencies. It also does not provide a uniform message to the client, given the apparent confusion between the evaluation agencies. The overall finding is clear – there is a need to strengthen consensus on these important conceptual terms. Finally, the section reiterates the finding that a results framework is needed for good gender mainstreaming.

While several evaluations focus on gender mainstreaming as an instrument to achieve gender equality, others treat gender mainstreaming as a goal by itself, the assumption being that effective gender mainstreaming will lead to gender equality. The wisdom of such an approach can be questioned. In some cases, this was driven by the fact that the evaluation units did not have either the time or the resources to explore causal links between gender mainstreaming and the desired

16. See Gender and Evaluations, Sheet 12 at <http://www.oecd.org/ataoecd/47/56/44896217.pdf>.

Table 3. Gender policies, objectives and approaches of agencies			
Organization	Latest policy	Policy goal or objective	Approach
AfDB	Y (2001)	Gender equality	Gender mainstreaming with focus sectors
ADB	Y (1998)	Gender equity	Gender mainstreaming
CIDA		Gender equality	Gender mainstreaming and women-specific programming
DFID	Y (TSP 2002)*	Empowerment of women	Gender mainstreaming
EBRD	N (GAP 2009)	Equity of opportunity and women's empowerment	Gender mainstreaming
IFAD	N (Framework 2008)	Gender equality and empowerment of women	Gender mainstreaming and women-specific programming
SDC	Y (2003)	Gender equality	Gender mainstreaming and women-specific programming
UN-HABITAT	Y (2002)	Women's empowerment	Gender mainstreaming
UNDP	N Strategy (2008-2011)	Gender equality	Gender mainstreaming
UNICEF	Y (2010)	Gender equality; equal rights of girls and boys to contribute to poverty reduction and achievement of MDGs	Gender mainstreaming
WB	Y (2003)	Addressing gender disparities and inequalities that are barriers to development	Mainstreaming at country level and selective integration at project level (gender equality is an instrument not a goal)
FAO	N (Strategic objective) 2009	Gender equality and the empowerment of women in support of FAO's mission	Gender mainstreaming
IDB	Y (2010)	Gender equality and women's economic empowerment	Gender mainstreaming do no harm

* Poverty Elimination and Empowerment of Women (2002) is considered to be the DFID gender strategy.

Second, the lack of consensus on what gender mainstreaming involved (see table 4) meant that the metrics for the evaluations varied. Some evaluations did not formulate indicators to monitor the extent of gender mainstreaming, whether it was at the policy, country, or project levels. Even when gender mainstreaming was defined similarly in a handful of evaluations around the ECOSOC definition, the utilization of different sets of indicators (or lack of any indicators) meant that the terms were interpreted differently. Some defined it in terms of process similar to that in the ECOSOC definition. On the other hand, some employed a 'results focus' in defining gender mainstreaming to "avoid limiting the evaluation to the overly general formulation of whether or not 'gender issues' have been integrated into policy and program work." The latter approach has some benefits in that if applied well it can provide greater clarity in terms of organizational effectiveness.

Third, there is need for greater consensus on the approach. The majority of evaluations found that gender mainstreaming was a sound approach, and that the lack of a results framework, weak accountability, inadequate skills, limited resources, and ineffective monitoring that reduced the effectiveness of gender mainstreaming as a strategy. One evaluation did not wish to over emphasize the "dichotomy of 'women-focused initiatives' and 'gender mainstreaming' which the gender mainstreaming strategy has falsely contributed to and which, in many cases has created more confusion than clarification." A second evaluation noted that some perceive gender mainstreaming as "an alienating concept to talk of rights of women and men, boys and girls." Finally, one evaluation noted that an exclusive focus on gender mainstreaming is inadequate to support gender equality and that a twin-track approach was needed, one that mainstreams gender issues and the second that complements the mainstreaming strategy with targeted interventions to promote gender equality and women's empowerment.¹⁷

agencies, responsible for assessing gender mainstreaming, are unclear on the best options for implementing gender mainstreaming. It will only help to strengthen their resistance to gender mainstreaming in countries where such resistance exists. Second, it is not that all staff within the donor community support gender mainstreaming. The lack of clarity only strengthens their assumptions that gender mainstreaming is all about women and that it will adversely impact men and take away their assets and resources, reiterating their resistance to such approaches.

17. As noted above, most activities to help women 'catch up' or reduce structural inequality – for example in targeting – could be defined as either 'gender mainstreaming' or 'women-targeted interventions'.

The lack of consensus is also not beneficial for the client if the key evaluation

Table 4. Eleven evaluations and six interpretations of gender mainstreaming

Definition 1: The process of assessing the implications for women and men of any planned action, including legislation, policies or programs, in all areas and at all levels. It is a strategy for making women's as well as men's concerns and experiences an integral dimension of the design, implementation, monitoring and evaluation of all policies and programs in all political, economic and societal spheres so that women and men benefit equally and inequality is not perpetuated. The ultimate goal is to achieve gender equality.

Definition 2: (1) Going beyond activities for women alone to ensuring that women participate in and benefit from "mainstream" modalities; (2) Focusing not on women in isolation, but instead within social units like families and communities; (3) Recognizing that it is often not possible to improve women's lives without involving men; (4) Acknowledging that the principles underpinning concerns for women's practical needs may similarly call for attention to others who are marginalized or overlooked, such as young men; and (5) Working for gender equality, but in ways that do not necessarily seek sameness for men and women, but promote equal opportunities to be healthy, to develop capabilities and to be treated with dignity and respect.

Definition 3: The term *mainstreaming* has its origins in approaches to education that include a diversity of learning abilities, rather than segregating or excluding those with disabilities from normal education. Mainstreaming suggests changes in established procedures and organizational cultures as new concepts and ways of doing business are accepted. The process of mainstreaming implies flexibility, innovation, learning, and acceptance of new cultural norms—which, however, may not be readily embraced by the existing mainstream. Therefore, the process is often subject to resistance.

Definition 4: A 'results focus' in defining gender mainstreaming to "avoid limiting the evaluation to the overly general formulation of whether or not 'gender issues' have been integrated into policy and program work." This evaluation focused on four areas or intermediate outcomes: (i) responding to the different roles, needs, priorities of women and men, with sensitivity to their practical needs; (ii) promoting equitable access to basic service provision; (iii) promoting equality of opportunity influence and benefit by responding to women's strategic needs; and (iv) attempting to challenge the balance of power between men and women in the public sphere.

Definition 5: Gender mainstreaming in poverty reduction should involve: (i) collection of sex-disaggregated data; (ii) social impact and gender analysis prior to any field based initiative; (iii) gender consideration integrated into MDG programs and monitoring; (iv) gender aware macroeconomic policies; (v) support for gender analysis at the regional or national level; (vi) Focus on gender inequality as a factor at the household level in impeding food security and well-being; (vii) consideration of gender roles in the workforce; (viii) consideration of gender-differentiated impacts of privatization, and social service outlays on women. (ix) recognition of gender relations as factors affecting the long-term impacts of women-targeted projects such as micro-enterprise and savings.

Definition 6: A strategy to ensure that women's and men's concerns and experiences are integral to the design, implementation, monitoring and evaluation of all legislation, policies and programmes.

Finally, it is important that the desired outcomes must guide evaluations in assessing the appropriateness of the gender mainstreaming strategies and the selection of indicators. For example, where the agency supports gender equality, the evaluation must assess whether the agency has undertaken relevant analysis that will lead to an understanding of both male and female issues and whether the agency has designed activities to reduce barriers to gender equality, which can vary depending upon the context.¹⁸ Yet another approach is to address power relationships between men and women – when couched in terms of complementarity and equal value of roles, this approach has demonstrated positive results in Guatemala, Peru, Cuba and Uganda. This would require evaluation to focus on whether the agency has used the right lens in mainstreaming gender issues, because the outcomes are aiming for different set of institutional rules and behaviours. Women- or men-focused indicators would be appropriate in this case. A third approach would focus on women's

empowerment. Effective evaluations cannot assess such an objective with indicators similar to that for gender equality. For example, in the case discussed earlier in Bangladesh, the social empowerment of women would be considered an appropriate outcome, although this may not lead to gender equality (a different objective). A fourth approach is that taken by the IDB, which is to ensure in addition to gender mainstreaming that the activities will mitigate to the extent possible any harm cause by the activity to women. This would require IDB to first assess whether gender mainstreaming is contributing to gender equality, and second to whether it is helping to identify and mitigate any unintended harm. A fifth approach – the human rights approach—would require the evaluation to focus on the rights of individual women and the obligations of the state to respect individual rights. Evaluations need to assess whether gender mainstreaming is crafted to achieve the different objectives. Such distinction is not evident, however, in the evaluations.

Box 3. Indicators for decision making in the evaluation of UNICEF support

Indicators to define decision making. Any result that contributes to: (a) More equal participation of women with men as decision makers in shaping sustainable development of their societies and/or of female and male children within the household, community and in a school setting in a development programme context; and (b) More equal participation of women with men as decision makers in shaping sustainable development of their societies and/or of female and male children within the household, community and in a school setting in an emergency or humanitarian response context.

Scale for rating indicators. Assign ratings of 0 to 3 based on the criteria below. 0 = No relevant activities planned. 1 = Assumption that regular project activities will have a positive impact on gender equality since the target beneficiaries are girls or women. 2 = Some explicit gender equality activities planned that are likely to achieve gender equality results anticipated. 3 = Relevant gender equality activities planned and these are well integrated into overall programme/project and are likely to achieve gender equality results anticipated.

18. In countries such as Mongolia, boys are being taken out of schools to support their fathers in caring for livestock. A gender equality approach would require activities to encourage parents to send boys to school.

Quality of evaluation frameworks

Evaluation frameworks for assessing gender equality must be right for the agency, flexible enough to be adapted to various contexts and types of gender policies, and tailored to the availability of resources, skills, and time. In the evaluations assessed for this report, the scope, quality of framework, and terminologies are widely different, and some of these aspects tend to weaken the transparency of the evaluation frameworks, thereby reducing the credibility of the evaluations. These can mainly be attributed to the lack of a clear results framework as well as a lack of focus on outcomes.

About half of the evaluations craft questions around the standard Organisation for Economic Co-operation and Development/Development Assistance Committee (OECD/DAC) criteria of relevance (strategic focus), efficiency, or effectiveness, (program results). Those evaluations that did not cover results, focused on the first two criteria. While most evaluations were comfortable in dealing with the first and third criteria, this was not so with the concept of efficiency. IEG preferred to replace ‘efficiency’ with quality of policy implementation; and ADB with ‘responsiveness’ to institutional, internal, and external factors. Three other evaluations attempted to discuss the efficiency of objectives. The SDC evaluation noted that efficiency questions would be addressed in the context of project evaluation and monitoring and would not be treated in-depth in the overarching evaluation. IOE aimed to assess efficiency but found that the team could not rate the efficiency even at the project level because the lack of relevant information, coupled with methodological complications, made it impossible to gain a proper appreciation of performance in this evaluation criterion. The OED evaluation also attempted to use efficiency as one of the criteria but did not carry the intention forward through the evaluation. The UNICEF evaluation found similarly and concluded that efficiency can be assessed only if UNICEF decides to start tracking its gender equality expenditures more systematically and links them directly to explicit gender equality

results. It would be timely for OECD to update guidance in this area.

Only half of the evaluations crafted their criteria or questions around gender-related outcomes. CIDA did this explicitly. Its set of criteria was explicitly around outcomes: (i) CIDA’s commitment (to outcomes); (ii) enabling outcomes and effectiveness; (iii) development outcomes; and (iv) relevance (continuing need for CIDA’s investments). This was possible, however, because CIDA has enunciated a clear outcome-based framework for gender equality. The discussion of results then becomes more meaningful because all the 4 criteria focus on gender-related outcomes. For example, in assessing its comparative advantage, CIDA asks: “What are the comparative advantages (areas of strength and experience) of CIDA in realizing gender equality outcomes, compared to other donors?” The reflection of outcomes in the evaluation questions is appealing because the evaluation explicitly focuses on the extent to which the agency is contributing to desired outcomes. Table 5 presents a comparative perspective of the different questions for relevance.

Table 5. Differences in evaluation questions

Agency	Sample questions on policy relevance or strategic focusing
CIDA	Is there a continuing need for CIDA investments to contribute to achieving gender equality results? How has CIDA positioned gender equality as part of its goals and commitments, in response to perceived needs?
DFID	What evidence is there of what works and what doesn't in integrating gender in newer aid modalities and in different country contexts?
IED (ADB)	Is ADB's approach to implement Gender and Development Policy adequately differentiated to have country buy-in especially for category I and category II projects?
IEG (World Bank)	To what extent is the World Bank's policy resulting in the "right" gender issues (that is, issues relevant for poverty reduction and economic growth, as noted in OP/ BP 4.20) being adequately addressed at the country level?
IOE (IFAD)	Is IFAD's strategic guidance on gender relevant for IFAD's rural poverty reduction mandate?
OED (FAO)	Extent to which has gender been mainstreamed in FAO National Medium Term Priority Framework and what are the likely prospects it will be in the next round of Country Programming Frameworks
SDC (E+CD)	What is the mix of strategies for addressing gender equality and how does this affect the quality and impact of the SDC contribution?
UNDP (IEO)	How effective are the approaches used by UNDP in promoting gender equality?
UN-HABITAT	To what extent has the Gender Equality Action Plan made UN-Habitat's work on gender more strategic and coherent?
UNICEF EO	Does the current UNICEF Gender Policy properly interpret the framework for gender equality as per the relevant UN and other conventions, institutions and reporting mechanisms?

Establishing benchmarks for gender mainstreaming provides some challenges. The lack of data is a severe constraint handled differently in the evaluations. For example, in one case, the selection of countries was motivated by the availability of data. Thus, only countries with Gender-related Development Index and Human Development Index ratings in 2005 were selected. Most evaluations were consistent in the view that the extent of gender mainstreaming is a percentage of all projects that integrated gender considerations over the entire portfolio of projects. In order to be consistent with the WB gender policy that requires only selective gender integration into sectors and activities identified by Country Assistance Strategies, IEG distinguished between projects that had gender-relevant objectives (about 70 percent) and those which had not. It then calculated the extent of gender mainstreaming as a percent of projects with gender integration overall the entire portfolio as well as a percentage of the 70 percent.

In assessing results of gender mainstreaming at the institutional and project levels, IED had a clear advantage in that its benchmarks for gender mainstreaming were institutionally set. In 2001, ADB introduced a four-category classification of gender orientations in its operation. ADB's framework for results introduced targets that at least 40 percent of all projects (numbers) would have mainstreamed gender by 2012. Although the evaluation finds that benchmarks were necessary in all four categories, given ADB's core sectors, and there was a need to shift from inputs (gender mainstreaming) to results, nevertheless, it had a well-defined benchmark for gender mainstreaming at the project level.

IEG and the OED evaluations tackle the difficult question of whether gender integration is required in all projects and programs. In the case of IEG, it was necessitated by the agency's more selective approach to gender integration. The evaluations found that in roughly 20-25 percent of projects and activities were gender neutral and that even if gender was not integrated, these activities were likely to benefit both men and women equally or

equitably. Such an assessment also helped to set a more refined benchmark for levels of gender integration.

The IOE evaluation includes a detailed benchmarking review to compare integration with that of other agencies as well as to identify commonalities and differences in order to derive good practices and pointers for learning. The benchmarking review included ADB, CIDA, OECD, SDC, UNDP, WB, and WFP and is based on documentary review of existing literature as well as discussions with key informants in the organizations covered.

The 2009 IEG gender evaluation compares integration of gender considerations against its 2001 evaluation. This was helpful in that it provides a baseline and a comparator to assess progress in integrating the Bank's policy, thereby allowing the IEG evaluation to conclude that the scope and nature of gender integration has improved since the nineties both in terms of quantity and scope. However, the 2009 evaluation also finds that attention to implementation of policy decreased during the in the period between the 2001 and 2009 evaluations. Prior evaluation reports of Habitat and IED were also available, but given the new processes introduced during the interim period, the prior evaluations were not fully usable. IFAD applied a before-and-after approach by dividing the projects into three roughly equal groups covering three time periods: 2003-2005 (six projects); 2006-2007 (seven projects); and 2008-2009 (nine projects). This permitted better understanding of how the approach has evolved over the years.

Table 6. A basket of mixed methods								
Evaluation unit	Doc. review	Inter-views	Participatory method	Beneficiary satisfaction	Staff survey	Field visits	Web-based or VC sessions	Lit. rev. or meta evaluations
OVEP				Not assessed				
ADB	Yes	Yes	Yes	Yes	No	Yes	No	Yes
CIDA	Yes	Yes	Yes	Not known	Yes	Yes	Not known	Not known
DFID	Yes	Yes	Yes	Yes	No	Yes	Not known	Not known
IFAD	Yes	Yes	Yes	Yes	No	Yes	No	Yes
SDC	Yes	Yes	Focus group discussions	Yes	Yes	Yes	No	No
UN-HABITAT	Yes	Yes	Not known	Not known	Yes	Yes	Not known	Yes
UNICEF	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not known
WB	Yes	Yes	Focus group discussion	Yes	Yes	Yes	Yes	Yes
UNDP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not known
FAO	Yes	Yes	Yes	Yes	No	Yes	Not known	Not known

Given the paucity of quantitative data to assess women's economic empowerment or changes in gender equality, each evaluation relied on a basket of methods to triangulate qualitative information from different sources (see table in annex I). This constraint is even more acute at the local or community levels for while there may be sex-disaggregated data at the national levels, such data at the local or subnational levels is most often absent. Meta evaluations to understand past gender results are possible only in limited cases where past evaluative work had considered gender issues. Perhaps for this reason, only the evaluation of IFAD support included such a meta evaluation, because previous IOE evaluations had integrated gender based on primary data collection from the field. Thus, most evaluations utilized a blend of methods ranging from desk review to field assessments, and then validated information from the various sources.

In this regard, DFID takes an interesting stand – the evaluation assesses without structured indicators and instead, uses semi-structured interviews, focus group discussions utilizing the Change Assessment and Scoring Tool (CAST), documentary

analysis and triangulation of information from various sources to assess gender impacts. This is appealing in that the assessment found that such participatory assessments (which involved both female and male beneficiaries) to be most successful method in understanding gender-related results. However, availability of time and resources are essential for participatory methodologies.

Finally, although the term gender equality was used regularly, most evaluations took a gender and development approach (defined by DFID as “an approach that bases interventions on analysis of men's and women's roles and needs in an effort to empower women to improve their position relative to men in ways that will benefit and transform society as a whole”). Thus evaluations assessed women's empowerment as a first step and only then, whether such empowerment led to gender equality. For example, the evaluation of CIDA support assessed whether CIDA had contributed to the realization of human rights for women; given women access to and control over resources and benefits of development; and advanced women's equal participation with men as decision-

makers. It then takes the analysis further to discuss whether such results, especially in the first two dimensions led to gender equality. The assessment does, however, find that evaluations have gone beyond a purely women in development approach – although there may be some confusion most evaluations aimed to assess women’s empowerment or gender equality. The lack of adequate data meant that some resorted to reflecting results for women, particularly outside the health and education sectors.

Discussion of male gender issues appears only in a minority of evaluations. The UNICEF evaluation is one of the few that deals with issues related to boys throughout the evaluation. The IEG evaluation focused on gender disparity in education, pointing out that in many cases, boys were a disadvantaged group but in the discussion of equal access to economic resources, the focus is on women. The CIDA and IEG evaluations note that it is important to involve men in the effort to achieve gender equality. CIDA, IEG, and UNICEF evaluations note that both men and women had been involved in the consultation process. It is possible that this happened in all evaluations but was not particularly highlighted. One of four IEG criteria to assess gender mainstreaming at the project level was whether feedback was sought from both men and women at the design stage. The DFID evaluation was specifically interesting in that it defined gender equality as “Women having the same rights and opportunities as men, including the ability to participate in public life”.

Mainstreaming gender issues in evaluations of evaluation units

To further assess attention to gender equality dimensions in evaluations, a sample of recent evaluations from six independent evaluation units were assessed for their attention to gender equality. To the extent possible, more recent evaluations were selected (those prepared in the last five years). Where there were similar evaluations across units, such as on the agency's response to global crisis or climate change within this period, these were selected. 'Special studies' with a date after 2009 were picked for EBRD given that its gender action plan was only initiated in 2009, and there were not adequate number of other evaluations. In the case of IDB, language was a constraint; only evaluations available in English were picked for this rapid assessment, and some of these were more research pieces and were classified under "other". The 59 evaluations reviewed for this section are provided in annex I. Table 7 below provides the types of evaluations that were evaluated. Given that the purpose of this evaluation was not "accountability", such a method was considered reasonable. While corporate gender evaluations are necessary in order to assess policy relevance, institutional arrangements for gender integration, and understand gender dimensions of corporate wide results, it is

perhaps equally, and if not more, important to integrate gender into the regular project, sector, country, and global evaluations. This would provide a steady reminder to agencies to increase attention to gender issues. In order to understand the extent of gender integration in regular evaluations, a rapid assessment was undertaken on the extent to which independent evaluation units in six multilateral agencies integrated gender considerations into their regular (non-gender) evaluations. The assessment aims to suggest ways to strengthen gender-related integration into evaluative work.

Fifty-nine documents from the six independent evaluation units were selected for the assessment. Where available, evaluations (corporate-, country-, sector-, and project-levels) were selected from sectors where gender issues appeared to be a relevant issue (but excluding the health and education sectors).¹⁹

Table 7. Sample evaluations by type of assessment

Type	ADB	AfDB	EBRD	IDB	IFAD	WB	Total
Corporate	1	3	1	1	1	1	8
Country	3	2		4	3	3	15
Other				5			5
Project	3	1			5	3	12
Sector	3	4	3		1	3	14
Special			5				5
Grand total	10	10	9	10	10	10	59

19. Health and education sectors were omitted for the following reasons: (i) two of the six agencies (IFAD and IBRD) do not work in these sectors; and (ii) several evaluations point out that there is significant progress in gender mainstreaming in these sectors.

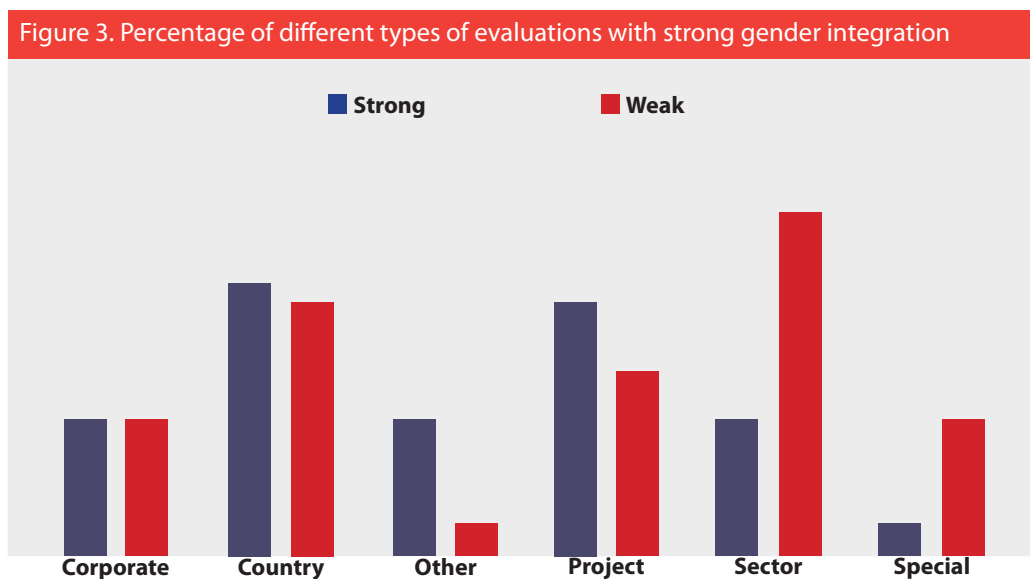
Each evaluation is assessed against three indicators: (i) extent of relevant discussion of the gender dimensions in substantiating findings; (ii) integration of gender into the framework; and (iii) use of some data in supporting results. In addition, the assessment examines the extent to which evaluations included at least one recommendation or lesson on gender in the report. Each of the three indicators was assessed using a two-scale rating – strong and weak—and each evaluation that received two ‘strong’ ratings was considered a strong evaluation.

This assessment also reviews whether any of the selected ECG participants provided guidelines to their staff on integrating gender considerations into corporate, country, project, or sector evaluations. Table 6 summarizes the findings. ADB and IFAD appear to be the only two evaluation units, which have substantively instructed their staff to look at gender issues.

ADB and IFAD instructions require that evaluations consider the gender-related results. ADB’s instructions to staff undertaking Country Assistance Program Evaluations (CAPE) were comprehensive and required mainstreaming gender considerations into its evaluation framework including providing sample gender-aware indicators. Similar instructions at the project level were not of equal quality, but these are reportedly under review. The IFAD Evaluation Manual provides detailed instructions. It notes that the promotion of gender equity and poverty are to be considered as integral dimensions within the various evaluation criteria adopted by the evaluation. It also notes that relevance is assessed inter alia with regard to project design features such as gender mainstreaming. It stresses the need to integrate gender into assessing who benefitted, and notes that it needs to be considered explicitly under each of the five impact domains, as appropriate. Based on the findings of the 2010 CLE on gender, IOE also introduced a new criterion on gender equality and women’s empowerment, to be applied in all evaluations undertaken. From 2011, gender considerations are to be separately evaluated and reported in a dedicated section of the main annual

evaluative review. The AfDB Guidelines include one reference to providing guidance to staff to assist the Board in monitoring the strategic objectives of the Bank, such as local participation, gender mainstreaming, regional integration and private sector promotion, which have an important bearing on the relevance, efficiency and effectiveness of the Bank’s operations.

The review finds that gender issues are addressed weakly or not at all in less than half of the 59 evaluations. More than 20 percent of these 59 relevant evaluations were totally gender blind. Given that gender-relevant evaluations and other research products were considered, this number clearly indicates the needs for strengthening of gender considerations in the majority of evaluations undertaken by Evaluation Agencies. Specifically, about 46 percent of the evaluations included gender integration that was considered strong. The majority of IDB’s research pieces (other in figure 3 and table 7) integrated gender into their research frameworks very well. In the other types of evaluations, project and country-level evaluations integrated gender better than in other types (figure 3).



IOE and IED evaluations had the highest levels of gender integration. While the sample is small, and the evaluation did not focus on before and after the guidelines, nevertheless it is worth noting that the two organizations that had issued clear guidance to staff to integrate gender considerations into their evaluation had the highest percentage of evaluations that integrated gender. IOE evaluations particularly focused on gender issues because a separate gender rating was required in all evaluations. IED had taken the approach that there should not be a separate gender rating and this left gender issues not being treated as explicitly in the ADB evaluations. While it can be debated whether there should be a separate rating for gender, it appears that this step by IFAD has encouraged evaluators to pay greater attention to gender.

Three evaluations of the recent crises conducted by ADB, EBRD, and WB were assessed to compare the extent of gender integration in each. While the EBRD evaluation unit report is gender blind, the ADB and WB evaluations make a few references to women's issues. The ADB evaluation expresses an aim to discuss the consequences of the crisis on unemployment (of men and women), but does not fully follow through on this intention. Both reports point to evidence showing the loss of jobs in export-oriented industries, some of which employ a large number of women. The ADB evaluation

notes that the Bangladeshi government announced package to stimulate exports, agriculture production, and the energy sector, and the WB evaluation notes that IFC has invested in 17 microfinance institutions in 9 countries in 2 Regions—Latin America and the Caribbean (35 percent) and Europe and Central Asia (65 percent)—reaching over 1.6 million people, most of whom are either women or rural inhabitants (or both). Both evaluations fail to take the issue up further and this means missed opportunities in assessing the gender dimensions of social and poverty impact.

A similar comparison was undertaken of corporate evaluations of support provided by ADB, AfDB, EBRD, IFAD, and WB. All evaluations, except the one of EBRD, make some reference to women's or gender issues. The IED evaluation integrated the results of its corporate gender evaluation. IEG and IOE evaluations were more extensive in their treatment of gender issues providing some scope for comparison. The IEG report integrated gender extensively in separate sectors, but in the absence of a clear results framework and an overall summary, the findings were not pulled together. While IOE extensively mainstreamed gender issues, the evaluation unit has recently required a separate rating for gender under the revised guidelines – this drew greater attention to gender issues.

The overall findings of low gender evaluations are consistent with self-assessments undertaken by IEG and OED of the extent of integration into their respective evaluations. These report similar weaknesses. IEG's Gender Evaluation reports that project Implementation Completion Reports (ICRs) were silent in about one-fourth (22) of the closed projects that integrated or discussed gender issues at appraisal and had an IEG Implementation Completion Report Review available. In 41 percent (9) of these cases, IEG ICR Reviews addressed gender issues; in the other 59 percent (13), IEG did not note the ICR's failure to discuss gender issues or consider this in rating outcome or ICR quality.

OED's Gender Evaluation undertook much more extensive self-assessment. It reports similar weaknesses: "Between 2002 and 2010, OED completed 39 major evaluations of various sorts: thematic, strategic, country

and large emergency operations. There were references to women and/or gender issues in 29 of these, even if at times this was only to state that no attention had been paid to these issues or that no meaningful work had been done. Of the 10 evaluations that contained no reference to women or gender, six appeared to be missed Opportunities in terms of evaluative evidence on gender issues."

In most cases, gender considerations are not woven into the evaluation framework but are provided in a separate paragraph on gender or women's issues, or just include some passing references. Only a handful of evaluations integrate gender considerations into the evaluation framework. Although the sample is very small and not very representative, there seems to be some improvement in gender integration in the latter years. Nevertheless, there is a considerable scope for improvement.

Box 4. Good practice integration

The Ghana Rural Services project evaluation (IFAD project-level evaluation) is a good practice in that it: (i) reviewed quality of gender analysis and integration into design at entry, during implementation, and in the design and monitoring framework; (ii) utilized a gender-aware stratified sampling methodology, ensuring that both male and female beneficiaries were consulted in understanding results; (iii) developed indicators for assessing results that were sex-disaggregated to better understand how the resources were allocated and benefits generated; and (iv) focused on women's economic empowerment in analysing the information and understanding results.

Less than 40 percent of the evaluations use some quantitative data to support their findings. Some of the research work undertaken by IDB integrated gender-related or sex-disaggregated indicators explicitly into the overall frameworks. Most evaluations, however, continue to provide data related to women, and mostly in terms of outputs. For example, the focus is on women's labour participation rate without any attention to gender disparities in this area. Or again, although quantitative data is provided on increased business opportunities for women, the focus is entirely on women without any inquiry into how it affects their gender relations. Similarly, data is provided on loans received by women, although there is little inquiry

on whether such loans are leading to empowerment or reduction of inequalities. About 40 percent suggests relevant lessons or recommendations. These are those evaluations where there is a strong treatment of gender. A few recommendations stress the need to integrate gender into evaluation frameworks. Some of the lessons and recommendations are provided in box 5.

Box 5. Summary of some gender-related lessons and recommendations in regular evaluations

AfDB: Undertake, as a priority, a review of constraints contributing to weak quality at entry with regard to poverty, gender and environmental aspects in operations and country strategies, and take appropriate measures to promote improvements.

IED: Evidence further demonstrates the need to increase the number of gender-related indicators in design and monitoring frameworks and strengthen tracking of gender results in ADB's project performance monitoring systems.

IFAD: (1) Ensuring that gender-disaggregated monitoring data are identified in the logical framework, and then collected, is an important and necessary step to making gender issues visible. (2) Training, including literacy training, can be particularly important for empowering women. Before training, women's participation in decision-making committees was limited by their lack of literacy and numeracy, and their inability to express themselves in the Lao language.

EBRD: The third, financial instruments – chiefly loans – are the primary vehicle through which EBRD serves development outcomes whilst pursuing transition impacts. However, these transition–development linkages can be strengthened through loan conditions, and related technical assistance, that in addition to promoting financial sustainability also promote specific patterns of sector growth – targeting women, poor regions and so on.

IEG: The Bank did not have an adequate understanding of the political dynamics in the country, nor did it have the mandate or experience for this difficult endeavour. Mainstreaming of gender and youth issues in the Bank's program was not effective. While some measures to this end were included in some Bank projects, there was no serious follow up.

Conclusion

What stands out sharply when these evaluations are reviewed for results and through the lens of good evaluation methodology are two aspects. The first is the weakness in results frameworks (namely, lack of clarity in gender equality objectives). The second is the weakness in evaluation frameworks (confusion about the theory of change at the level of outcomes and the understanding of gender mainstreaming), as well as lack of targets and clear indicators with transparent rating scales. The weak evaluation frameworks mirror the lack of clarity in gender equality policies and results frameworks. Reviewing the set of evaluations brings both the differences among agencies and this overall weakness into sharper perspective.

As well as a lack of clear gender equality objectives, the results frameworks (and evaluation frameworks) are incoherent in articulating a theory of change or programme theory, and are variable in their ability to articulate the logical links between impact, outcomes and outputs. Indicators are inconsistent or absent, and there are few targets. This weakness then carries over into the evaluation frameworks and contributes to the elusive nature of results cited in the opening DFID quotation, or in the 'invisible results' CIDA evaluators could not pin down.

There is currently considerable work being done to sharpen the policy and results frameworks for gender equality (ADB, USAID, UN Women, WB) as well as in the agency response to the corporate gender evaluations. This should provide a more solid basis assessment of contributions to gender equality in future evaluations, whether they are specifically about gender equality or in other evaluation work carried out by ECG members.

ECG members can build on this work to harmonize their own evaluative practice and to encourage agency management to adopt more robust results frameworks and indicators in their ongoing portfolio management.

Annex I. Methodology for assessment of gender integration in evaluations of evaluation units

The table below provides a list of the evaluations selected for review. Although each evaluation has been rated to obtain aggregate figures, individual ratings are not disclosed in this paper. Each of these 59 evaluations evaluation was assessed against three indicators: (i) extent of relevant discussion of the gender dimensions in substantiating findings; (ii) integration of gender into the framework; and (iii) use of some data in supporting results. In addition, the assessment examined the extent to which evaluations included at least one recommendation or lesson on gender in the report. Each of the three indicators was assessed using a two-scale rating – strong and weak—and each evaluation that received two ‘strong’ ratings was considered a strong evaluation.

Evaluations may be generally divided into different types:

- a. Corporate-level evaluation – an evaluation of the entire or selected parts of the portfolio of the agency.
- b. Project-level evaluation – an evaluation of an individual project or in some cases a cluster of 2-3 projects.
- c. Sector-level evaluation – an evaluation of support in a particular sector .
- d. Country-level evaluation – where the country is the unit of account and results are aggregated from the project and sector levels are aggregated to the country level.
- e. Others – impact or outcome related research undertaken by IDB.
- f. Special – some evaluations undertaken by EBRD.

List of evaluations assessed

#	Org	Type	Date of report	Name
1	EBRD	Special	2010	Project completion report/assessment report
2	EBRD	Special	2010	Small Business Finance Operations
3	EBRD	Special	2010	The EBRD's Response to the 2008-2009 crisis
4	EBRD	Special	2011	Sustainable Energy Initiative
5	EBRD	Special	2011	Early Transition Countries Non-bank Microfinance Institution (MFI) Framework - February 2011
6	EBRD	Special	2011	Technical Cooperation Programs to Support the Tajik Agricultural Finance Framework
7	EBRD	Corporate	2011	Annual Evaluation Report 2010
8	EBRD	Sector	2011	Forestry and Forest-Industry Sector Study in Romania
9	EBRD	Sector	2010	Technical Cooperation Program to Support the Tajik Agricultural Finance Facility
10	EBRD	Sector	2010	European Union/EBRD SME Facility
11	WB	Sector	2010	The World Bank Group's Response to the Global Economic Crisis
12	WB	Country	2011	Timor Leste
13	WB	Country	2010	Nepal
14	WB	Sector	2010	Earnings Growth and Employment Creation
15	WB	Sector	2011	Climate change and the World Bank Group. Phase II: The Challenge of Low-Carbon Development

Gender equality and development evaluation units

#	Org	Type	Date of report	Name
16	WB	Country	2011	Peru
17	WB	Project	2011	Paraguay - Fourth Rural Water Supply and Sanitation Project; Ecuador - Rural and Small Towns Water Supply and Sanitation Project
18	WB	Project	2011	Colombia's Social Safety Net Project (PPAR)
19	WB	Project	2011	Ethiopia - Productive Safety Nets Project (PPAR)
20	WB	Corporate	2012	Results and Performance of the World Bank Group 2011
21	AfDB	Country	2011	Gabon Country Assistance Evaluations, 1996-2008
22	AfDB	Sector	2010	Evaluation of Project Supervision at the AfDB, 2001-2008
23	AfDB	Country	2008	Cameroon: Review of Bank Group Assistance to the Multi-sector, 1996-2004
24	AfDB	Sector	2011	Review of Bank Group Assistance to the Multi-sector
25	AfDB	Sector	2010	Agricultural Water Management: An Evaluation of the African Development Bank's Assistance in Ghana and Mali, 1990-2010
26	AfDB	Sector	2010	Evaluation of Policy Based Operations in the African Development Bank, 1999-2009
27	AfDB	Project	2011	Project for the Electrification of 17 Rural Centres
28	AfDB	Corporate	2011	Independent Assessment of the Quality at Entry of ADF 2005-2008 Operations and Strategies
29	AfDB	Corporate	2011	Evaluation of Project Supervision at the AfDB, 2001-2008
30	AfDB	Corporate	2009	2008 Annual Review of Evaluation Results What Have We Learned?
31	ADB	Country	2011	Maldives
32	ADB	Country	2011	Uzbekistan
33	ADB	Country	2010	Lao PDR
34	ADB	Sector	2010	Viet Nam: Rural Infrastructure Sector Project
35	ADB	Sector	2010	Real Time Evaluation of ADB's Response to the Global Crisis
36	ADB	Sector	2009	Sector Assistance Program Evaluation for the Agriculture and Rural Development Sector in Cambodia
37	ADB	Project	2011	Samoa: Small Business Development Project
38	ADB	Project	2011	Technological and Professional Skills Development Sector Project
39	ADB	Project	2011	The Asian Development Fund Operations: A Decade of Supporting Poverty Reduction in Asia and the Pacific Region
40	ADB	Corporate	2011	2011 Annual Evaluation Review
41	IFAD	Country	2011	Socialist Republic of Viet Nam
42	IFAD	Country	2011	Republic of Ghana
43	IFAD	Corporate	2011	Annual Report on Results and Impact of IFAD Operations (ARRI)
44	IFAD	Project	2011	Jordan: Yarmouk Agricultural Resources Development Project

#	Org	Type	Date of report	Name
45	IFAD	Project	2011	Socialist Republic of Viet Nam: Rural Income Diversification Project in Tuyen Quang Province
46	IFAD	Project	2011	Ghana: Rural Enterprises Project, Phase II
47	IFAD	Sector	2011	IFAD's Private-Sector Development and Partnership Strategy
48	IFAD	Project	2011	Armenia: Rural Areas Economic Development Programme - Project Completion Report Validation
49	IFAD	Project	2011	United Republic of Tanzania: Rural Financial Services Programme and Agricultural Marketing Systems Development Programme
50	IDB	Country	2011	Bolivia Country Evaluation
51	IDB	Country	2010	Chile Country Evaluation
52	IDB	Country	2010	Dominican Republic Country Evaluation
53	IDB	Other	2008	Evaluation of the Quality of Economic Analysis for Projects Approved 1997-2006
54	IDB	Sector	2011	Same Bureaucracy, Different Outcomes in Human Capital?
55	IDB	Sector	2011	The Welfare Impacts of Social Housing Programs in Latin America
56	IDB	Sector	2010	The Impact of Improving Access to Justice on Conflict Resolution
57	IDB	Sector	2009	Executive Summary of the Second Independent Evaluation of the Expanded Project Supervision Report Exercise
58	IDB	Sector	2010	Panama's Housing Programs - An Evaluation
59	IDB	Sector	2010	Country Program Evaluation Jamaica 2003-2008

Annex II. References

- 1998 Evaluating Gender Equality - Policy and Practice, An Assessment of SIDA's Evaluations in 1997-1998. Published by SIDA Evaluation.
- 2002 Office of the Special Adviser on Gender Issues and Advancement of Women, Gender Mainstreaming An Overview, United Nations.
- 2003 Gender Mainstreaming: Productive Tensions in Theory and Practice, Sylvia Walby University of Leeds (S.Walby@leeds.ac.uk). Contribution to ESRC Gender Mainstreaming Seminars, 2003-4.
- 2003a, DAC Evaluation Series, Review of Gender and Evaluation, Final Report to DAC Network on Development Evaluation, J. Hunt, R. Brouwers. Published by the DAC Network on Development Evaluation.
- 2005 Gender and Development – A Review of Evaluation Reports 1997-2004, by Nordic Consulting Group, A.S. Janne Lexow and Stein Hansen. Published by NORAD.
- 2004 Some Reflections on Gender Mainstreaming and Intersectionality, Jenny Riley, International Women's Development Agency, Melbourne.
- 2009 Guidance Note on Gender Mainstreaming and Social Exclusion in Research, DFID Research.
- 2011 How to Perform Evaluations – Gender Equality, Canadian International Development Agency.
- 2011 Writing Terms of Reference For An Evaluation: A How-To Guide, IEG, The World Bank.
- 2011 Mainstreaming Gender Equality: A Road to Results or a Road to Nowhere? An Evaluation Synthesis, May 2011, OPEV, African Development Bank.
- 2011 Handbook on Human Rights and Gender Equality in Evaluations in the UN, UNEG, http://www.unevaluation.org/HRGE_Guidance.

Annex III. United Nations definitions relating to gender equality

Equality between women and men (gender equality): refers to the equal rights, responsibilities and opportunities of women and men and girls and boys. Equality does not mean that women and men will become the same but that women's and men's rights, responsibilities and opportunities will not depend on whether they are born male or female. Gender equality implies that the interests, needs and priorities of both women and men are taken into consideration, recognizing the diversity of different groups of women and men. Gender equality is not a women's issue but should concern and fully engage men as well as women. Equality between women and men is seen both as a human rights issue and as a precondition for, and indicator of, sustainable people-centred development.

activities - policy development, research, advocacy/ dialogue, legislation, resource allocation, and planning, implementation and monitoring of programmes and projects.

Gender: refers to the social attributes and opportunities associated with being male and female and the relationships between women and men and girls and boys, as well as the relations between women and those between men. These attributes, opportunities and relationships are socially constructed and are learned through socialization processes. They are context/time-specific and changeable. Gender determines what is expected, allowed and valued in a woman or a man in a given context. In most societies there are differences and inequalities between women and men in responsibilities assigned, activities undertaken, access to and control over resources, as well as decision-making opportunities. Gender is part of the broader socio-cultural context. Other important criteria for socio-cultural analysis include class, race, poverty level, ethnic group and age.

Gender mainstreaming is a globally accepted strategy for promoting gender equality. Mainstreaming is not an end in itself but a strategy, an approach, a means to achieve the goal of gender equality. Mainstreaming involves ensuring that gender perspectives and attention to the goal of gender equality are central to all