

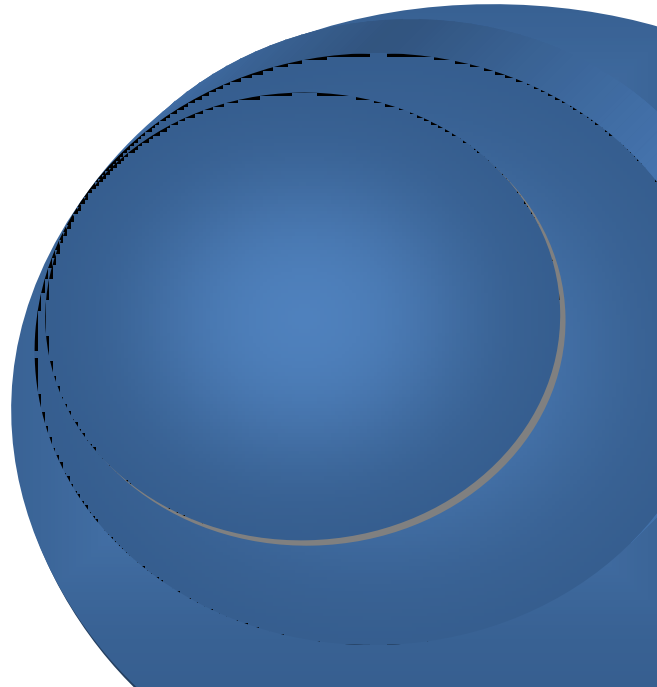


# Meta-Evaluation

## UNIFEM Evaluations 2004-2008

UNIFEM Meta-Evaluation covers the evaluations conducted in the MYFF period 2004/2007 and the first year of UNIFEM Strategic Plan 2008/2011. The analysis was separated in two different areas. The first part focuses on the quality, credibility and utility of UNIFEM's evaluations and how the organization can improve its evaluation efforts; the second part shows the results, findings, conclusions and recommendations presented in the different evaluation reports carried out by UNIFEM in the period 2004/2008.

**UNIFEM Evaluation Unit**  
**2009**





# **Evaluation Report**

## **Meta-Evaluation**

**UNIFEM Evaluations 2004-2008**

**2009**

This evaluation was conducted by  
Cecilia M. Ljungman and managed by  
UNIFEM Evaluation Unit

The analysis and recommendations of this report do not necessarily reflect the views of the United Nations Development Fund for Women, its Executive Board or the United Nations Member States. This is an independent publication prepared for UNIFEM and reflects the views of its authors.

## Foreword

During the last few years UNIFEM has made an increasing effort to strengthen its evaluation function and include it in its programmes and projects as a key accountability and learning tool to progress on the work on gender equality and women's empowerment.

In the Multi-year Funding Framework (MYFF) period (2004-2007), UNIFEM undertook a relevant number of evaluations of its programmes and made an effort to more systematically integrate the United Nations Evaluation Group's (UNEG) Norms and Standards. In the current Strategic Plan (SP) 2008-2011, the agency takes a step forward and makes a specific commitment towards evaluation, defining it as the cornerstone of UNIFEM's capacity to support continuous improvement and develop substantive effective strategies to achieve gender equality.

This meta-evaluation constitutes the first analysis of all UNIFEM evaluations conducted in the MYFF period and in the first year of the SP. It provides an assessment of the evaluation quality and consolidated findings of UNIFEM's programmes.

UNIFEM's Evaluation Unit contracted an independent evaluator, Cecilia M. Ljungman, to conduct this analysis with the double objective of contributing to learning through the identification of key findings, lessons learned and evidence of the key areas of UNIFEM's work; and to determine and uncover areas for improvement in the quality of the evaluations undertaken, identifying factors that may have weakened the validity, credibility and utility of evaluations conducted, based on UNEG Norms and Standards and UNIFEM guidelines.

The Meta-evaluation is divided in two different parts. The first section focuses on the quality of UNIFEM's evaluations and how the organization can improve its evaluation efforts. It is important to point out that, since we are considering two different programmatic periods, this first part contains an analysis of the MYFF period (2004-2007). The analysis on the quality, credibility and utility of UNIFEM Evaluations carried out in 2008, the first year of the implementation of the SP, can be found separately in "Annex 5 2008 Meta-Assessment".

The second part of the Meta-evaluation shows the results, findings, conclusions and recommendations presented in the different evaluation reports carried out by UNIFEM in the period 2004-2008. For this part, the analysis considers the complete period of the study (2004-2008)

We consider that this study presents important lessons learned on how to improve UNIFEM evaluations in the future. Among other findings, it assesses that from a total of thirty (30) reports finalized between 2004-2007 only eighteen (18) qualify as evaluations, and from those one (1) evaluation is rated excellent; eleven (11) are rated good; five (5) are rated average and one (1) is very weak. From the nine (9) evaluations reviewed for 2008, one (1) is rated as excellent; two (2) are rated as good; four (4) are rated average and two (2) are rated weak.

In addition, the meta-evaluation of UNIFEM's work in the period 2004-2008 provides relevant information about the agency's achievements and challenges in different thematic areas.

With regard to the main findings of the results of UNIFEM's work, we want to highlight that different aspects related to results-based management (RBM) were an important theme in most evaluations with several evaluations identifying an urgent need for UNIFEM to further enhance its use of RBM as a flexible and meaningful management tool within its projects and programmes. In contrast, generally, capacity development was recognized as UNIFEM's strength and many evaluations maintain that the achievements and other aspects of the projects and programmes are highly worth showcasing within UNIFEM and beyond because of their potential for learning and replication, including several instances in which UNIFEM acted as a catalyst so that modest investments led to broader changes.

Finally we would like to thank Cecilia M. Ljungman for having conducted such a systematic assessment and to all the informants that contributed to compile the information needed to carry out this evaluation and, of course, to all the people that through the years have contributed to strengthen evaluation function in UNIFEM.

We hope you find this document useful and that it will add to your evaluation knowledge. We intend to use it as a basis to improve the evaluation capacity within the organization and believe that it can represent a learning tool for UNIFEM staff in the different programmatic areas of the agency work. Overall, we hope that this exercise will contribute to UNIFEM's progress in the realization of women's human rights and gender equality.



**Belén Sanz**  
Evaluation Advisor, UNIFEM



# Table of Contents

<b>Foreword</b>	i
<b>List of abbreviations</b>	3
<b>Abbreviations used to denote UNIFEM evaluation Reports</b>	4
<b>Executive Summary</b>	7
<b>Evaluation Quality, Utility and Credibility</b>	15
<b>1 Introduction</b>	16
1.1 Methodology	16
1.2 Structure of the Report	18
<b>2. Mapping and Scoring of UNIFEM Assessments Undertaken between 2004 to 2007</b>	19
2.1 Findings related to Evaluation Characteristics	20
2.2 Findings related to Evaluation Quality	25
<b>3 Utility and Credibility</b>	32
3.1 Terms of reference	32
3.2 Evaluation teams	34
3.3 Stakeholders role in the process	35
<b>4 Analysis of Evaluation Quality</b>	36
4.1 Areas for Improvement	37
<b>5 Recommendations</b>	41
<b>Evaluation results</b>	44
<b>6 Introduction</b>	45
<b>7 Evaluation Themes</b>	46
7.1 Management for Results	46
7.2 Capacity Development	49
7.3 Partnerships	50
7.4 Regional Approaches	53
7.5 Human Rights and Gender Equality Perspectives	55



7.6	Institutional Learning & Knowledge-Sharing	56
7.7	Human Resources & Financial Management	58
<b>8</b>	<b>Evaluation Criteria</b>	<b>60</b>
8.1	Relevance	60
8.2	Effectiveness	62
8.3	Sustainability	67
8.4	Efficiency	68
8.5	Impact	69

## **Annexes**

Annex 1: Terms of reference	73
Annex 2: Quality Assessment and Scoring framework	81
Annex 3: Criteria for determining an Evaluation or Review	88
Annex 4: Profile of Assessments undertaken from 2004 to 2007	90
Annex 5: 2008 Meta-Assessment	103
Annex 6: List of documents consulted	120
Annex 7: List of persons interviewed	124

## List of Abbreviations

<b>CEDAW</b>	Convention on the Elimination of Discrimination Against Women
<b>CEE-CIS</b>	Central and East Europe and Commonwealth of Independent States region
<b>CIDA</b>	Canadian International Development Agency
<b>CSO</b>	Civil society organization
<b>EVAW</b>	Eliminating violence against women
<b>IAWG</b>	Inter-Agency Working Group
<b>LAC</b>	Latin America and Caribbean region
<b>MYFF</b>	Multi-Year Funding Framework
<b>NGO</b>	Non-governmental organization
<b>RFP</b>	Request for proposal
<b>SEAP</b>	South East Asia and the Pacific sub-region
<b>SMART</b>	Specific, measurable, attainable, realistic and timely
<b>TOC</b>	Table of content
<b>Tor</b>	Terms of reference
<b>UNEG</b>	United Nations Evaluation Group
<b>UNV</b>	United Nations Volunteers
<b>USAID</b>	United States Agency for International Development

## **Abbreviations used to denote UNIFEM evaluation reports**

<b>16 Days Campaign</b>	16 days Campaign against Violence against WOMEN in the Andean Region (2002)
<b>Afghanistan-WL</b>	Supporting Women's Leadership in Rebuilding Afghanistan (2007)
<b>AWP</b>	Arab Women Parliamentarian Project (2007 Review)
<b>Burundi Peace</b>	Engendering Peace Building and Mediation Processes in Burundi (2004)\
<b>Caucasus- Peace I</b>	Women for Conflict Prevention & Peace Building in the Southern Caucasus, Phase 1 (2004)
<b>Caucasus Peace II</b>	Women for Conflict Prevention & Peace Building in the Southern Caucasus Phase 2 (2007)
<b>CEDAW-Pacific</b>	Progressing the Implementation of CEDAW in the Pacific (2006)
<b>CEDAW-SEAP</b>	Progressing the Implementation of CEDAW in South East Asia (2008)
<b>CEE Economic Justice</b>	Gender and Economic Justice in European Accession and Integration (2004)
<b>DRC Election</b>	Mainstreaming Gender into the Electoral Process in DRC (2007)
<b>GRB-LAC</b>	Engendering Budgets: Making Visible Women's contribution to National Development in Latin America (2008)
<b>HIV/AIDS</b>	UNIFEM's Programme on Enhancing Human Security through Gender Equality in the context of HIV/AIDS (2006)
<b>IFAD</b>	Mainstreaming Gender in Poverty Reduction: IFAD UNIFEM Gender Mainstreaming Programme in Asia
<b>IGAD</b>	IGAD Gender Strategy in the Context of Globalization and the Feminisation of Poverty (2005)
<b>India Railways</b>	Stock-Tacking of the Railways-UNIFEM Project on Equalizing Gender Relationships in the Context of the HIV/AIDS Epidemic (2005 Review)
<b>India VAW</b>	Reinventing India: Actions for Empowerment to Address Violence Against Women

<b>Kenya G&amp;G</b>	Evaluating the Gender and Governance Programme Kenya (2008)
<b>Kenya GE&amp;WR</b>	Promoting Women's Human Rights and Enhancing Gender Equality in Democratic Governance in Kenya (2006)
<b>Kosovo GE&amp;WR</b>	Advancing Gender Equality and Women's Rights in Kosovo (2008)
<b>Kyrgyzstan-Land</b>	Women's rights to Land in Kyrgyzstan (2006)
<b>Local-Global</b>	From Global to Local: a Convention Monitoring and Implementation project (2004)
<b>MDG</b>	Gender and Millennium Development Goals (2007)
<b>Morocco Literacy</b>	Functional Literacy and Women's Empowerment in Morocco (2008)
<b>Morocco Political Participation</b>	Social Communication, Political Participation and Democratic Governance for Women's Empowerment in Morocco (2008)
<b>MYFF</b>	Evaluation on UNICEF's MYFF System (2007)
<b>Nigeria Empowerment</b>	Women's Economic Empowerment Project in Nigeria (2008)
<b>SA-Trafficking</b>	Prevention of Trafficking in Women and Children in South Asia
<b>SEELINE</b>	Evaluation of the South-eastern European Legal Initiative (SEELINE) (2004)
<b>SEE-WR</b>	CIDA-UNIFEM Accountability for Protection of Women's Human Rights (2008)



## Executive Summary

As part of the effort to improve evaluations and strengthen the learning from them, UNIFEM's Evaluation Unit has identified the need to analyse the evaluations conducted between 2004 and 2007 (MYFF period) and 2008 (first year of the Strategic Plan). The aim of this meta-evaluation is to aggregate and systematise the information generated from these evaluations and transform it into accessible knowledge for future evaluations and programming. The study has addressed two different questions: i) what is the quality, credibility and utility of UNIFEM's evaluations and how can the organisation improve its evaluation effort; and, ii) what are the results – findings, conclusions and recommendations – presented in the evaluation reports?

The methodological approach of the meta-evaluation consisted of the following. First, a database was established in which the characteristics of evaluation reports from between 2004 and 2008 were mapped. Second, each report was assessed to determine whether it met the criteria for an evaluation or review, or whether it constituted another type of assessment altogether. Third, the *quality* of each evaluation/review was analysed. Based on UNEG standards, a framework including 16 parameters for assessing the quality of the evaluations was devised to score each report. Fourth, interviews were conducted with UNIFEM staff and some evaluators involved in 9 evaluations from 2006 to 2008. The evaluations selected represented all regions, thematic areas, and evaluations with average, good or excellent quality scores. The interviews aimed to gather data about the respective evaluation processes, the credibility of the evaluation reports and the usefulness of the findings, conclusions and recommendations. Finally, the *results* of the 18 evaluation reports from 2004 to 2007 and 9 evaluations from 2008 were reviewed. Furthermore, the results of these evaluations were studied. Themes that were common were identified and analysed. Conclusions that relate to the evaluation criteria were also examined.

The meta-evaluation methodology faced a few limitations. To begin with, since most of the data underpinning this evaluation originates from the evaluation reports themselves and their respective annexes, the findings of the meta-evaluation are only valid to the extent that the UNIFEM evaluations present valid findings, conclusions and recommendations. In some cases, the meta-evaluation lacks data because an evaluation report also lacked this information. Furthermore, the evaluator has not always been able to fully triangulate data collected from the interviews to assess usefulness and credibility. The number of interviews was relatively limited because either stakeholders could not be contacted or did not respond to requests for interviews. Nevertheless, the data gathered sufficed to provide a general snapshot of usefulness and credibility and what factors were believed to have contributed these qualities.

### Mapping & Quality of UNIFEM's Evaluations

An average of five evaluations were undertaken every since 2004, but in 2005 and 2006, only three evaluations were conducted. Reasons for this are uncertain. The thematic areas in which most evaluations were undertaken relate to *achieving gender equality in times of peace and recovery from war*, followed by *reducing the feminisation of poverty*. Five evaluations have addressed cross-cutting thematic areas. Only one evaluation concerned HIV/AIDS while two focused on violence against women.

In terms of geographic spread, four global evaluations were undertaken since 2004, all during the MYFF period. Overall, the Africa, Asia/Pacific and CEE/CIS regions were all well represented by UNIFEM evaluations. However no evaluations were undertaken in the LAC region until 2008. Several sub-regions have not undertaken any evaluations in this period either. Among them are West Africa, Lusophone Africa, Southern Africa and the Caribbean.

The aggregate quality score for the 27 UNIFEM evaluations conducted during 2004 and 2008 is 30 out of 48 points, or “average”. Of the evaluations from the MYFF period, one scored “excellent”, eleven scored “good”, two scored “average”; one scored “weak” and three scored “very weak”. The scores earned by the 2008 evaluations were one “excellent”; two “good”; four “average”; two “weak”, and no “very weak” scores.

A respectable 60 per cent of the evaluations have scored “good” or above. This, nevertheless, needs to be seen in the light of the fact that as many as forty per cent of the 44 “evaluation” reports submitted to the Evaluation Unit actually turned out to be reviews or other types of reports and were therefore not included in the aggregate scoring exercise. Mistaking so many reports for being evaluations suggests that there is insufficient conceptual understanding within UNIFEM of what actually constitutes an evaluation.

Even among the evaluations of higher quality, there are **several areas where UNIFEM’s evaluations could improve**. Many of the reports have deficiencies in relation to content and structure. For instance, only four evaluations were complete when it came to basic report content. While reports seemed to have improved in this regard over the period evaluated, this needs to be addressed to ensure that the evaluations are credible and useful. Other content-related shortcomings were the often weak contextual backgrounds and descriptions of the evaluated intervention – including its goals, logic, history, organisation and stakeholders. Without these elements, the potential readership of the evaluation and the possibility to learn lessons is limited. The concluding sections of UNIFEM’s evaluations are another weak spot.

There are also opportunities to strengthen several aspects relating to the **methodological approaches** of UNIFEM’s evaluations. Only four evaluations scored full points for explaining the purpose, objectives and scope of the evaluation and providing a clear account of the methodology applied. A common shortcoming was a failure to consider the limitations of the methodology. In addition, more than half of the evaluations had no methodological tools to show for in the report or annexes.

Related to methodology but scored separately in this meta-evaluation are the extents to which the evaluations were **participatory**, applied a **gender equality and human rights** perspective in the methodology and included **ethical safeguards** in the evaluation design. In all three areas, there is considerable room for improvement.

The mapping and scoring of the MYFF period evaluations reveals some **patterns regarding evaluation characteristics and the quality score**. First, global and cross-cutting evaluations scored “good” or “excellent”. Second, evaluations that were managed by UNIFEM’s headquarters or co-managed with an external partner – either a donor or another UN agency – earned, on average, 35 points (a solidly “good” score). Third, evaluations of particularly high profile, also scored well. Fourth, evaluations that ensured stakeholder participation to verify and validate the evaluation’s key findings and conclusions before the report was drafted, scored an average of 38 points.

Finally, a clear advantage for evaluations was having a **team of evaluators** as opposed to a single consultant. Half of the evaluations were conducted by teams and all but one of these scored either “good” or “excellent”. On the other hand, among the nine one-person evaluations, only one-third scored “good” while the remainder scored average or below. This should come as no surprise. Successful consulting firms in most sectors have for decades realised the quality gains reaped by offering clients a team of experts with complementary skills. Through a co-ordinated effort, the team members generate synergy, which allows each member to maximise his or her strengths and minimise his or her weaknesses. The sum thus becomes greater than its parts.

The characteristics that seem to be associated with higher scores either signify higher costs (multi-member teams, longer evaluation processes and stakeholder validation and verification) or indicate that the evaluation had a relative robust budget (headquarters managed or managed jointly with a donor or UN agency). This seems to suggest that you get what you pay for.

## **Credibility and Usability**

Of the nine evaluations studied to assess credibility and usefulness, eight were used to a greater or lesser extent to prepare a second phase of the intervention. Furthermore, several of the UNIFEM staff interviewed felt that they had gained significant useful experience by being part of an evaluation. This implies that **evaluations are contributing in important ways to UNIFEM’s work and its learning processes**. In two cases, however, the evaluations were not as useful as expected by UNIFEM. While much of the evaluation content was taken to heart, UNIFEM had hoped for more evidence, analysis and ideas to guide its future work. In one of the cases, the terms of reference did not sufficiently cover UNIFEM’s need. In the other, the evaluators did not fully meet the demands of the terms of reference.

In terms of credibility, discussions revealed that there appeared to be a general **correlation with the quality score** earned by each evaluation. Thus, evaluations like the CEDAW-SEAP (score of 43) was regarded as highly credible while it was felt that GRB-LAC (score of 27) could have covered certain issues better (but was by no means considered incredulous). The stakeholders interviewed identified some aspects that contributed to usefulness and credibility. These included clear and detailed terms of reference from the start and ensuring that key stakeholders (within UNIFEM as well as external partners) were brought on board early in the evaluation process. Stakeholder participation was seen as enhancing ownership, strengthening commitment to the process and contributing to the evaluation’s usability. In particular, the importance of an activity before the report was drafted to verify findings and validate conclusions was highlighted.

The **expertise of the evaluators** was also regarded as important for the credibility and usefulness of the evaluation. However, this was not considered enough. Dialogue between UNIFEM and the evaluation team also had to be good. Evaluators and UNIFEM staff interviewed agreed that the rapport and understanding established between UNIFEM and the evaluation team was crucial. It allowed for a smooth evaluation process – which in itself contributed to quality and a more credible and useful evaluation.



## Looking Ahead

Without having had an evaluation unit or an evaluation strategy until recently, and still no evaluation policy of its own (a requirement of UNEG standards), UNIFEM is not yet where it should be in terms of evaluation practice. This undermines its learning, results-based management and accountability. The evidence uncovered by the meta-evaluation shows that while there has been a significant amount of quality evaluation work within UNIFEM, a few evaluations have been very weak.

Improving the credibility, usability and quality of UNIFEM's evaluations requires **informed and skilled evaluators**, combined with the **competent management** of the evaluators and evaluation process. The latter will entail increased capacity within UNIFEM and among its partners. Programme staff and managers need a sound understanding of evaluation concepts and processes. Staff with monitoring and evaluation functions requires more in-depth knowledge.

Significant efforts are already underway to address the need for greater capacity within UNIFEM by rolling out a training plan for staff and partners and preparing guidelines and tools. The newly established Evaluation Unit is leading this process and is already setting its mark. It has also engaged in (much appreciated) backstopping for UNIFEM's evaluations efforts in the last year; set up an intranet website, gradually developed internal systems and has drafted an evaluation strategy.

Ensuring that UNIFEM evaluations are conducted by informed and skilled evaluators is more nebulous than building in-house capacity. UNIFEM nevertheless has some influence over this. First, UNIFEM can make a greater effort to attract evaluators to bid on its evaluation assignments. Providing clear and detailed terms of references with briefing material, sufficient time to prepare proposals and longer timeframes for evaluation processes are means within UNIFEM's control that can encourage evaluators to consider submitting a proposal. Second, UNIFEM can aim to communicate better with the resource base of evaluators to both gain their interest and inform them of UNIFEM, its evaluation effort, standards and its expectations.

Below are **recommendations** for UNIFEM's Evaluation Effort:

1. UNIFEM's Evaluation Unit should continue and further reinforce its efforts at **raising the evaluation capacity** within the organisation.
2. When procuring an evaluation, UNIFEM's evaluation managers should ensure there is **sufficient time for tendering** companies/consultants to present proposals.
3. UNIFEM's management should promptly make its **evaluations available to the public**.
4. UNIFEM's Evaluation Unit should further **develop its intranet site** to support colleagues in their evaluation work.
5. UNIFEM's Evaluation Unit should consider **collecting analyses of good practice evaluation** processes.
6. UNIFEM's Evaluation Unit should consider ways to **broaden the resource base of qualified consultants** that are willing to bid on UNIFEM's evaluation assignments.

7. UNIFEM's management should **take into account the costs that improving evaluation quality** will entail.

### **Thematic Areas Addressed by Evaluations Conducted during 2004-2008**

The thematic areas that were common to several evaluation between the period 2004 and 2008 include management for results; capacity development; partnerships; regional approaches; gender equality & human rights; institutional learning & knowledge-sharing; and, human resources & financial management. The sections below summarise findings related to each area.

Most evaluations found fault with at least some aspect of the **management for results**:

1. A common problem with many of the projects and programmes evaluated was the setting of unrealistic, overly ambitious goals or scope, given the resources available, the time frame of the intervention and the capacity of UNIFEM.
2. While the programmes and projects were seen as designed in a relevant way given the needs, the evaluations found that research, baselines and/or analysis were missing. Above all, many of the evaluations found that an overall strategic perspective was missing from the design. Underlying assumptions and the corresponding theory of change were seldom made explicit in the project document, which weakened the design.
3. Monitoring and evaluation was often found to be inadequate and the prompt elaboration of monitoring instrument was seen as necessary.
4. Some saw an urgent need for UNIFEM to further enhance its use of RBM as a flexible and meaningful management tool within its projects and programmes.
5. UNIFEM was regarded as underselling itself by not reporting on achievements.

Lessons from the evaluations relating to **capacity development** include the following needs:

1. An explicit programme capacity development strategy
2. Careful targeting of suitable trainees and target groups
3. Undertaking robust needs assessments
4. Ensuring trainers have relevant knowledge and skills
5. Systematically tracking longer-term effects of the capacity development of different partners.

As **a partner**, UNIFEM was regarded as highly valuable:

- UNIFEM was described as dedicated, knowledgeable, exceptionally approachable and 'down to earth' compared to other international organisations;
- UNIFEM did not impose specific approaches or ideas, but was respectful of, responsive to and actually listened to the needs and suggestions of its local partners;
- UNIFEM status as a UN agency and its adeptness in acting as a neutral convener was praised;

- UNIFEM was seen as playing a unique role and adding value by linking civil society actors with government; and
- UNIFEM's status as a UN agency provided NGO/CSO partners with the needed legitimisation for their work, and helped individuals and groups in gaining access to high level forums that they would not otherwise be able to access.

However, in some cases, evaluations found that some partnerships appear to have more inherent potential than had been brought to bear to date. Also, while UNIFEM had been skilful in partnering with and networking among gender equality organisations, a few evaluations found that UNIFEM was less successful in forging links with agencies and organisations not working directly with women/gender equality such as line ministries, actors in the economic sector and regional organisations.

UNIFEM's relationship with other UN agencies varied. On the one hand, there was evidence of UNIFEM's "systematic and growing engagement in UNCTs and related/relevant working groups and processes" which was appreciated by many sister agencies. However, in some cases, relations were not completely harmonious.

A **regional approach to programming** can add value by i) coherently addressing issues that transgress national borders – such as human trafficking; ii) broaching subjects that may be too sensitive to approach at the national level; and, iii) enhancing capacity-building efforts by accessing regional resources and bringing together actors from different countries to exchange experience and lessons. Sadly, less than five of 10 regional evaluations reflect on the regional aspect of the programme in terms of strengths, weaknesses and opportunities and constraint that this may bring. This seems to in part to reflect the fact that UNIFEM itself has not always formulated strategies within its regional projects/programmes that optimise the regional aspects.

None of the projects/programmes evaluated appear to consist of a **full-fledged rights-based approach** in which the principles of equality and non-discrimination; participation and inclusion; and rule of law and accountability are systematically applied in analysis, design and implementation. Nevertheless, several projects/programmes applied relatively solid **human rights perspectives**. With a couple of exceptions, these projects/programmes constituted those that worked expressly with promoting women's rights (and CEDAW in particular) and/or achieving gender equality in democratic governance.

In terms of **institutional learning**, many evaluations maintained that the achievements and other aspects of the projects and programmes were highly worth showcasing within UNIFEM and beyond. These include several instances in which UNIFEM acted as catalyst so that modest investments led to broader changes. However, while some efforts have been made to improve documentation and dissemination of success stories, the considerable potential for learning within UNIFEM has only been partly tapped into to date. Evaluators did not find a lack of willingness to share experience within UNIFEM or among its partners; rather there was limited time made available for reflection, analysis and report writing. Opportunities for structured exchange, joint discussion of strategic and conceptual questions, and in-depth analysis of common lessons learnt have reportedly also been limited in many cases.

In terms of **human resources**, a number of evaluations mentioned the importance of having a dedicated programme manager and a national co-ordinator in the countr(ies) where UNIFEM's projects and programmes were implemented. The use of national staff was particularly praised.

Staff members in most countries were considered very committed and supportive of partners. Offices in several countries were praised for their knowledge and experience. A few of the evaluation reports point to staff being over-worked and/or working under great pressure. Also, staff turnover and vacancies adversely affected a number of projects and programmes. Furthermore, lack of clarity and shared understanding with regards to roles and responsibility within UNIFEM was raised in a few cases.

**Financial management** was a concern in some evaluations. Delays in disbursements were experienced. Financial management and monitoring was uneven or slow.

Few evaluations directly commented on **gender equality perspectives** in the programmes. In most cases, evaluations did not examine whether or not gender analyses had been undertaken or used as part of the project preparation process.

## **Overall Assessments by Evaluation Criteria**

**Relevance** – the extent to which a development intervention conforms to the priorities and policies of the target group, recipient and/or donor – was addressed by a little under two-thirds of the evaluations. On the whole, they deemed UNIFEM’s efforts as relevant.

**Effectiveness** – the extent to which planned objectives have been achieved – is particularly difficult to attain when the goals are overly ambitious in relation to the resources available. The transformational results that UNIFEM and its partners aim for typically take much more than a couple of years to achieve. Nevertheless, UNIFEM’s evaluations from this period report that at an overall level, effectiveness has ranged from fair to high. Outputs goals were largely fulfilled and a majority of interventions achieved significant outcomes as well. However, three otherwise well-performing governance projects/programmes all had comparatively limited results for its objectives in the economic sector.

Of the evaluations that addressed **sustainability**, most implied that sustainability in terms of the mere continuation or duration of results has not been achieved. Financial sustainability was also deemed low. It was pointed out that building sustainable capacity in governments –where staff turnover often is high – will require consistent and longer-term support. Several of the evaluations therefore also judged sustainability in terms of ownership, commitment, networking, institutionalisation and ability to attract other donors. In these terms, most of the projects/programmes had gained important ground. However, in all cases, the evaluators deemed that continued financial and technical support would be required to achieve permanent change.

One evaluation pointed out that the concept of sustainability underlying UNIFEM’s work in the area of implementing CEDAW had largely remained implicit. In terms of achieving sustainable results this has limited learning from experience of what works and what does not. Some of the evaluations brought up the need for exit strategies and “ethical review” in the context of UNIFEM’s engagement and subsequent disengagement in communities.

Relatively few evaluations assessed **efficiency**. All of those that did concurred with conclusion that UNIFEM has produced remarkable results with very limited resources. However, efficiency could

be enhanced by better financial monitoring, more effective results based management (in particular through improved monitoring and evaluation); less staff turnover; and enhanced communication practices.

In relation to **impact**, two evaluations made systematic efforts at assessing change. Both evaluations showed methodological rigour, worked from baselines and extensively applied participatory techniques to determine change. However, none of the other the evaluations between 2004 and 2008 thoroughly assessed the totality of effects brought about through the development intervention in question, or took into account the counterfactual. Moreover, the fact that some terms of references and reports use the term “impact” loosely or incorrectly suggests that there has been a lack of clarity within UNIFEM and among some of its evaluators concerning the meaning and implications of this assessment criterion.

A few evaluations justifiably maintain that assessing impact was not possible to assess due to the insufficient passage of time, lack of baselines and/or problems of inferring causality. Furthermore, assessing impact was made more difficult in that UNIFEM had generally not been systematic in reporting on longer-term effects of capacity-building efforts.

Nevertheless, some evaluations reveal indications of impact, although the methodological rigour applied to reach these conclusions may, in some cases, be questionable. The types of impact described can be loosely divided into micro (community level), meso (institutional level) and macro (policy and legislative level) levels.

At the macro-level, a handful of evaluations concluded that UNIFEM interventions influenced legislation and successfully contributed to providing formal frameworks for the realisation of women’s rights. Indications of impact at the meso-level were less reported on in the evaluations. They consisted of raised consciousness of gender equality issues within institutions resulting in subsequent actions.

Five evaluations offered indications of some impact at the community level. These include, for instance, that women became significantly more autonomous in their mobility in public spaces, enhanced their control over their financial, personal and/or health issues; and were better able assist their children with school. Women also gained the capacity to participate in society and were more sensitive to issues of the public domain. They became more active in their communities by attending meetings (women’s committees, young people’s committees, farmers’ groups, cooperatives, NGOs and health committees) and courts of the elders; participating in discussions about local development; and, contributing to conflict resolution. Furthermore, some of the women took on the role of being change agents by passing on knowledge to other women and family members. Increased women’s participation, in turn, influenced public perceptions of women’s role in the family and society and women’s rights. Finally, one evaluation went as far as to report that the support had impacted on income poverty in the community: by improving land use and management and supporting women’s mini-enterprise development, the project actually contributed to an increase the number of medium and rich households, while the number of poor families decreased.

# **Evaluation Quality, Utility and Credibility**

# 1 Introduction

Between 2004 and 2007, UNIFEM's work was guided by a corporate Multi-Year Funding Framework (MYFF). The MYFF included a strategic results framework that identified four goals that UNIFEM sought to support:

1. Reduction of feminised poverty and exclusion
2. Ending violence against women
3. Halting and reversing the spread of HIV/AIDS among women and girls
4. Achieving gender equality in democratic governance in times of war and peace

During this period, UNIFEM had no evaluation function. However, when UNIFEM's Strategic Plan 2008-2011 and Biennial Support Budget 2008-09 were formulated, they specified the aim to strengthen UNIFEM's results-based system. Both defined evaluation as the cornerstone of UNIFEM capacity to support continuous improvement of its effort to achieve gender equality. The Strategic Plan commits UNIFEM to strengthening and sharpening the focus of its evaluation effort. It obliges the establishment of systems and processes that are consistent with those of UNEG and that will facilitate a) independent, high quality evaluations; b) management responses to evaluations recommendations; and c) effective learning from evaluations.

In 2008 an independent Evaluation Unit with its own budget was established as an independent structure, with its own budget. It reports directly to the Executive Director. This constituted UNIFEM's first step in strengthening its evaluation function in line with UNEG norms and standards. By the end of 2008, an evaluation strategy was prepared. The Unit aims to manage strategic corporate evaluations; provide back-stopping to staff managing evaluations at the decentralised level; and, enhance the organisation's evaluation capacity by preparing tools, guidelines and rolling out an evaluation training programme.

As part of the effort to improve evaluations and strengthen the learning from them, UNIFEM's Evaluation Unit has identified the need to analyse the evaluations conducted between 2004 and 2007 (MYFF period) and 2008 (first year of the Strategic Plan). The aim of this study is to aggregate and systematise the information generated from these evaluations and transform it into accessible knowledge for future evaluations and programming. . The study has addresses two different questions: i) what is the quality, credibility and utility of UNIFEM's evaluations and how can the organisation improve its evaluation effort; and, ii) what are the results – findings, conclusions and recommendations – presented in the evaluation reports?

## 1.1 Methodology

The meta-evaluation consisted of the following activities:

- a) A database was established in which the characteristics of 30 reports from between 2004 and 2007 were mapped.<sup>1</sup> This included gathering information on the subject, year, timing, length,

---

<sup>1</sup> As a separate exercise, the same was done for 13 reports from 2008. The resulting assessment is included in Annex 5.

management, date and place of each evaluation. The content of the reports and their annexes were also examined to assess the extent they complied with UNEG standards. In all, 9 aspects of the evaluations were entered in the database.

- b) Each report was assessed to determine whether it met the criteria for an evaluation or review, or whether it constituted another type of assessment altogether. The criteria used in this assessment are included in Annex 3.
- c) The *quality* of each evaluation/review was analysed. Based on UNEG standards, a framework including 16 parameters for assessing the quality of the evaluations was devised to score each report. The framework and scoring system is included in Annex 2.
- d) The data collected through mapping and scoring the reports was aggregated and analysed.
- e) The *results* of the 18 evaluation reports from 2004 to 2007 were reviewed. Furthermore, the results of 9 evaluations from 2008 were also studied to enhance the utility of the analysis and illustrate the pervasiveness of certain themes. Findings common to several of the reports from the period 2004 to 2008 were analysed. Conclusions that relate to the evaluation criteria were examined.
- f) Interviews were conducted with UNIFEM staff and some evaluators involved in 9 evaluations from 2006 to 2008 (CEDAW-SEAP, Kyrgyzstan-Land, Kenya GE&HR, Kenya G&G, Afghanistan WL, GRB-LAC, Kosovo GE&WR, MYFF, HIV/AIDS)<sup>2</sup>. The evaluations selected represented all regions, thematic areas, and evaluations with average, good or excellent quality scores. Evaluations from 2004 and 2005 were not included because of i) the potential difficulty in identifying stakeholders from that period; and, ii) the likelihood of encountering less clear memory of the evaluation process among those stakeholders as a result of the passage of time. The interviews aimed to gather data about the respective evaluation processes, the credibility of the evaluation reports and the utility of the findings, conclusions and recommendations.

The meta-evaluation methodology faced a number of limitations. First, since most of the data underpinning this evaluation originates from the evaluation reports themselves and their respective annexes, the findings of the meta-evaluation are only valid to the extent that the UNIFEM evaluations present valid findings, conclusions and recommendations.

Second, in some cases, the meta-evaluation lacks data because an evaluation report also lacked this information. An example is the budget size of many projects evaluated. It has been beyond the scope of this evaluation to apply complementary data-gathering approaches to secure this data.

Third, some of the evaluation reports may not explicitly clarify, for instance, the management arrangements or the length of the evaluation process. In many cases, however, the evaluator has been able to deduce this information from bits of data contained in different parts of the report, the terms of reference and/or annexes.

Fourth, the evaluator has not always been able to fully triangulate data collected from the interviews to assess utility and credibility. The number of interviews was relatively limited because either stakeholders could not be contacted or did not respond to requests for interviews. Nevertheless, the

---

<sup>2</sup> The abbreviations used for the evaluations reports are clarified in the beginning of this document.



data gathered sufficed to provide a general snapshot of utility and credibility and what factors were believed to have contributed these qualities.

## 1.2 Structure of the Report

This report consists of two parts. The first part focuses on assessing the *quality, utility and credibility* of evaluations from the MYFF period of 2004 to 2007.<sup>3</sup> The first chapter maps out the characteristics of the evaluations. It also presents findings related to the quality of the reports. Chapter 2 maps the characteristics of the evaluations and scores the quality of each evaluation in relation to UNEG standards. Chapter three provides findings related to the credibility and utility of evaluations undertaken between 2006 and 2008. Chapter 4 consists of an analysis of the findings and concludes on areas for improvement. Chapter five provides recommendations for improving UNIFEM's evaluation effort.

Part 2 of the meta-evaluation looks at the *results* from the evaluations undertaken during the MYFF period (2004 to 2007) *and* 2008. Evaluation results from the past year (2008) are included in the analysis to ensure relevance and illustrate the pervasiveness and extent of certain issues. The first chapter looks at themes that have been raised. These include management for results; capacity development; partnerships; regional approaches; gender equality and human rights perspectives; institutional learning and knowledge-sharing; and, human resources and financial management. The second chapter summarises how the evaluators have assessed the projects/programmes in relation to relevance, effectiveness, efficiency, sustainability and impact.

There are seven annexes to the report. They include the Terms of Reference for the Meta-Evaluation; the Quality Assessment & Scoring Framework; criteria for determining an evaluation or review; profiles of the assessments undertaken during the MYFF period; the Meta Assessment of UNIFEM's 2008 Evaluations; the list of documents consulted; and, the list of persons consulted.

---

<sup>3</sup> Evaluations from 2008, the first year of the 2008-2011 Strategic Plan, are not included in Part 1 but are mapped and analysed in Annex 5.

## 2 Mapping and Scoring of UNIFEM Assessments Undertaken between 2004 to 2007

Upon applying the criteria to distinguish evaluations from reviews and other types of reports (included in Annex 3), the reports from the MYFF period have been classified as follows:

Table 1: UNIFEM: Reports from 2004 to 2008 Analysed in the Meta Evaluation

	Name of Report	Year	MYFF Goal
<b>OTHER TYPES OF REPORTS</b>			
1	Promotion of Women's Human Rights through the Elimination of <b>Violence against Women SE Asia</b>	2004	EVAW
2	Elimination of Violence Against Women in <b>Central Asia</b>	2004	EVAW
3	UNIFEM Peace & Security Programme in <b>Sierra Leone</b>	2005	Governance
4	<b>Savings Mobilisation</b> & Productive Investment Pilot in Iligan & Jagna: Regional Programme on Empowering Women Migrant Workers in Asia	2005	Poverty
5	Capacity Building <b>Workshop</b> for Mainstreaming Gender Analysis in HIV/AIDS Programming in the Caribbean	2006	HIV/AIDS Poverty
6	Empowering women in global production networks - NTFP India <b>Gum Karaya</b> Project	2007	
<b>REVIEWS</b>			
7	Strengthening the Network of <b>Home Based Workers SE Asia</b>	2004	Poverty
8	Strengthening the Network of <b>Home Based Workers South Asia</b>	2005	Poverty
9	Stock-Taking of the <b>Railways-UNIFEM Project</b> on Equalizing Gender Relationship in the Context of the HIV/AIDS Epidemic	2005	HIV/AIDS
10	Gender and Justice in <b>Afghanistan</b>	2005	Governance
11	<b>Arab Regional Trust Fund</b> for the Empowerment of Women, Phase I	2006	EVAW
12	Arab Women <b>Parliamentarians</b> Project	2007	Governance
	Name of Report	Year	MYFF Goal
<b>EVALUATIONS</b>			
13	<i>Gender and Economic Justice in <b>European Accession</b> and Integration</i>	2004	Poverty
14	Engendering Peace Building and Mediation Processes in <b>Burundi</b>	2004	Governance
15	Evaluation of the Southeastern European Legal Initiative ( <b>SEELINE</b> )	2004	Governance
16	Women for Conflict Prevention and Peace Building in the <b>Southern Caucasus</b>	2004	Governance
17	From <b>Global to Local</b> : A Convention Monitoring and Implementation Project	2004	Cross-cutting

18	<b>IGAD Gender Strategy</b> in the Context of Globalisation and the Feminisation of Poverty	2005	Poverty
19	Prevention of <b>Trafficking</b> in Women and Children in South Asia	2005	Poverty
20	Reinventing India: Action for Empowerment to Address <b>Violence against Women</b>	2005	EVAW
21	UNIFEM Programming in <b>HIV/AIDS</b>	2006	HIV/AIDS
22	Promoting Women's Human Rights and Enhancing Gender Equality in Democratic <b>Governance in Kenya</b>	2006	Governance
23	Progressing the Implementation of <b>CEDAW</b> in the Pacific	2006	Cross-cutting
24	Women's Rights to Land in <b>Kyrgyzstan</b>	2007	Poverty
25	Mainstreaming Gender in Poverty Reduction: <b>IFAD-UNIFEM</b> Gender Mainstreaming Programme in Asia	2007	Poverty
26	Mainstreaming Gender into the <b>Electoral Process in DRC</b>	2007	Governance
27	Supporting Women's Leadership in <b>Rebuilding Afghanistan</b>	2007	Governance
28	Women for Conflict Prevention and Peace Building in the <b>Southern Caucasus, Phase 2</b>	2007	Governance
29	Gender and the <b>Millennium Development</b> Goals	2007	Cross-cutting
30	Evaluation of UNIFEM's <b>MYFF</b> System	2007	Cross-cutting

## 2.1 Findings related to Evaluation Characteristics

The 18 evaluations listed above have been mapped according to the following variables:

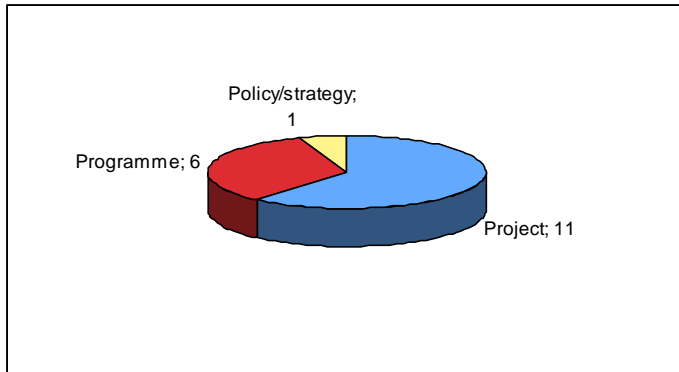
1. The **subject** of the evaluation (project, programme, thematic cluster or policy/strategy)
2. The **year** the evaluation was conducted
3. The **MYFF goal** area that the evaluation falls into
4. The **timing** of the evaluation
5. The **region** the evaluation was undertaken in
6. The **type** of evaluation team (internal, external or mixed)
7. Who **managed** the evaluation
8. The **length** of the evaluation process
9. The **report content** of the evaluation reports

The following sections provide a quantitative analyse of the mapping results.

### 2.1.1 Subject

More than half of the evaluations were project evaluations. Between 2004 and 2007, there were 11 project evaluations (nearly two-thirds) and 6 covering programmes. Only one policy/strategy evaluation was undertaken. There were no evaluations covering thematic clusters.

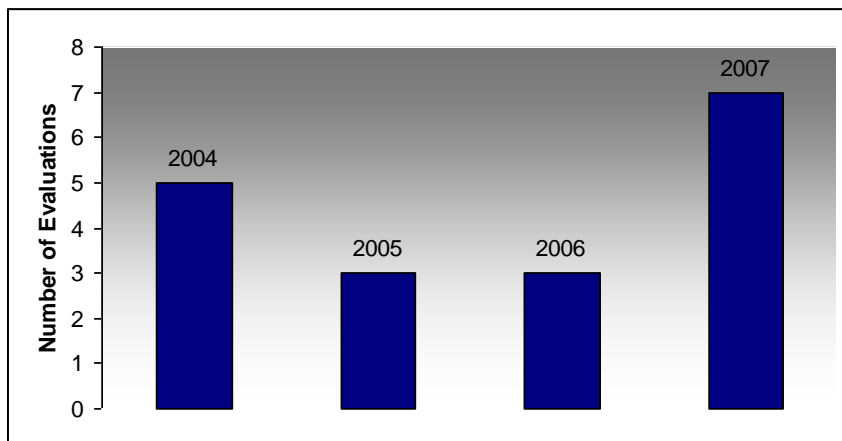
Figure 1: Evaluations by Subject



### 2.1.2 Year

An average of five evaluations were undertaken every since 2004. There was a dip in the number of evaluations undertaken during 2005 and 2006. In fact, in 2007, the number of evaluations conducted was more than double that of the two previous years. The meta-evaluation was not able to uncover reasons for the changes in the number of evaluations undertaken for the different years.

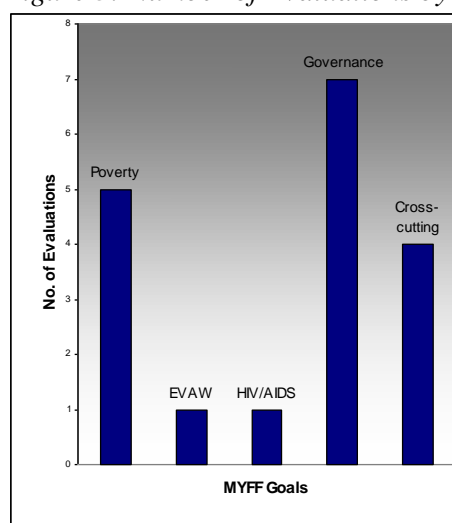
Figure 2: Number of Evaluations from 2004 to 2007



### 2.1.3 MYFF Goal

More than one-third of evaluations undertaken were related to the MYFF goal of achieving gender equality in governance in times of peace and recovery from war (hereinafter referred to as “governance”). Five evaluations relate to reducing feminised poverty (hereinafter referred to as “poverty”) and four cover cross-cutting concerns – such as the implementation of CEDAW, the Millennium Development Goals and the MYFF system itself. For the thematic areas of ending violence against women (hereinafter referred to as EAW) and halting the spread of HIV/AIDS among women and girls (hereinafter referred to as “HIV/AIDS”); only one assessment each qualified as evaluations.

Figure 3: Number of Evaluations by MYFF Goal



### 2.1.4 Timing

More than two-thirds of the evaluations were conducted at the end or towards the end of the programme/project period. However, in several cases, there was an intention to continue the project/programme by launching a second phase. In these cases, the evaluations served as form of mid-term evaluation. Three evaluations were actually undertaken at mid-term. It was not possible to determine the timing of three evaluations.

### 2.1.5 Region

The 18 evaluations are relatively evenly spread among the CEE-CIS, Asia-Pacific and Africa regions and the global level. However, there were no evaluations from the LAC region in this period.

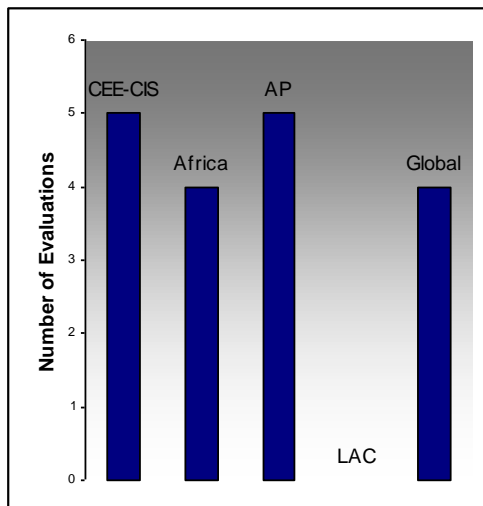
Of the four African evaluations, all concerned Horn/East & Central Africa and three were focused on a single country – Burundi, DRC and Kenya. The fourth, focused on the Horn & East (IGAD) sub-region. There were no evaluations from west or southern Africa.

Among the evaluations from the Asia-Pacific, three evaluations concerned South Asia, one the Pacific; and one covered different Asian sub-regions. There were no evaluations from the Arab states.

There were four global evaluations. They covered 2 to 4 countries as case studies. Morocco and Kenya featured as case studies in two evaluations each.

There were 6 evaluations that covered just one country. Four of these evaluations relate to the MYFF governance goal (Kenya, Burundi, DRC and Afghanistan). One focused on reducing feminised poverty (Kyrgyzstan) and one focused on eliminating violence against women (India).

Figure 4: Number of Evaluations by Region



### 2.1.6 Team

Most of the evaluations (17) were conducted by external evaluators. In the case of one evaluation, it is not clear whether the evaluation was external, internal or mixed. For half the evaluations, the team consisted of more than one person. Only one consulting company has undertaken more than one evaluation in this period – the Canadian consulting firm Universalia. It undertook two of the evaluations from 2004 and 2007.<sup>4</sup>

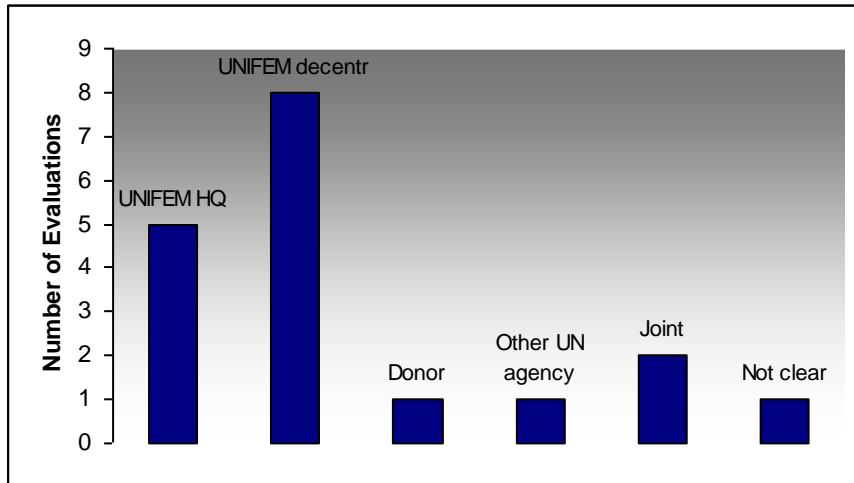
### 2.1.7 Management

Almost half of the evaluations were commissioned and managed by UNIFEM's sub-regional or country offices. Five evaluations were management by headquarters. During the period studied, UNIFEM did not have a specific institutional structure for evaluation. This function was managed by one person under the Directorate's supervision, and it did not have a specific budget. An Evaluation Advisor joined in 2007 and the Evaluation Unit was created in 2008.

<sup>4</sup> Universalia also undertook two evaluations for UNIFEM in 2008 – *Progressing the Implementation of CEDAW in Southeast Asia* and *Advancing Gender Equality and Women's Rights in Kosovo* – and are undertaking two more in 2009.

Two evaluations were undertaken jointly with donors – Belgian Co-operation and DfID respectively. Two evaluations were externally managed – one by IFAD and one by Sida. It is unclear from the report who managed the evaluation of *Gender and the Millennium Development Goals*, but it appears like it is UNDP.

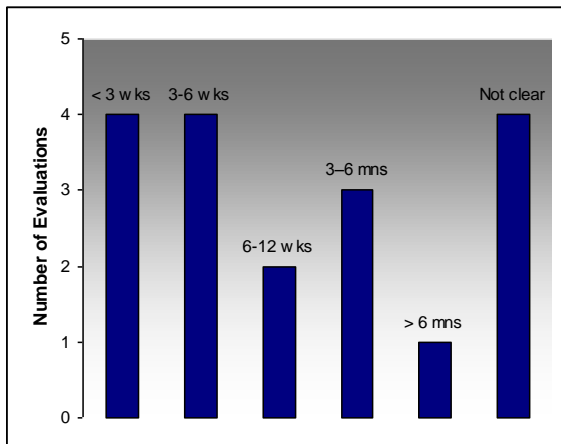
Figure 5: Number of Evaluations by Manager



### 2.1.8 Length of evaluation process

The length of the evaluation process varied considerably. A little under half of the evaluations were less than 6 weeks long (with 4 being less than 3 weeks) and one evaluation lasted more than 6 months. Three evaluations took 6 weeks to 3 months. Three evaluations lasted for 3 to 6 months. It was unclear how long the process lasted for 3 evaluations.

Figure 6: Number of Evaluations by Length of Evaluation Process



### 2.1.9 Report content

According to the UNEG standards, an evaluation report should include a table of contents, an executive summary and a list of abbreviations. The annexes should include the terms of reference,

the list of persons consulted/interviewed, the list of documents reviewed and examples of the methodology and/or data collection tools used.

Less than one-quarter (4) of the reports managed to include all of these in the report. In the table below, the grey boxes represent missing content. The reports from 2004 and 2005 lack more content than the reports from the two subsequent years – resulting in more grey boxes at the top of the table. The evaluations relating to the MYFF goal of governance and the cross-cutting evaluations tend to conform better to the UNEG standards in relation to report content.

Table 2: Contents of the Evaluation Reports

	ToC	Exesum	Acronyms	ToR	Interviewees	Documents	Tools
Gender and Economic Justice in <b>European Accession</b> and Integration	2004	Yes	No	No	No	No	No
Engendering Peace Building and Mediation Processes in <b>Burundi</b>	2004	Yes	No	Yes	No	Yes	Yes
Evaluation of the Southeastern European Legal Initiative ( <b>SEELINE</b> )	2004	Yes	No	No	No	No	No
Women for Conflict Prevention and Peace Building in the <b>Southern Caucasus</b>	2004	No	Yes	No	No	No	No
From <b>Global to Local</b> : A Convention Monitoring and Implementation Project	2004	Yes	Yes	No	No	Yes	Yes
Enhancing Women’s Economic Security: <b>IGAD Gender Strategy</b>	2005	Yes	No	Yes	Yes	Yes	Yes
Prevention of <b>Trafficking</b> in Women and Children in South Asia	2005	No	No	No	No	Yes	Yes
Reinventing India: Action for Empowerment to Address <b>Violence against Women</b>	2005	Yes	No	No	Yes	No	No
UNIFEM Programming in <b>HIV/AIDS</b>	2006	Yes	Yes	Yes	No	Yes	Yes
Promoting Women’s Human Rights & Enhancing Gender Equality in Democratic <b>Governance in Kenya</b>	2006	Yes	Yes	Yes	Yes	Yes	Yes
Progressing the Implementation of <b>CEDAW</b> in the Pacific	2006	Yes	Yes	Yes	Yes	Yes	No
Women’s Rights to Land in <b>Kyrgyzstan</b>	2007	Yes	No	No	Yes	No	No
<b>IFAD-UNIFEM</b> Gender Mainstreaming Programme in Asia	2007	Yes	Yes	Yes	Yes	Yes	Yes
Mainstreaming Gender into the <b>Electoral Process in DRC</b>	2007	Yes	Yes	Yes	Yes	Yes	Yes
Supporting Women’s Leadership in <b>Rebuilding Afghanistan</b>	2007	Yes	Yes	Yes	Yes	Yes	Yes
Women for Conflict Prevention and Peace Building in the <b>Southern Caucasus, Phase 2</b>	2007	Yes	No	Yes	Yes	Yes	No
Gender and the <b>Millennium Development</b> Goals	2007	Yes	Yes	Yes	No	Yes	Yes
Evaluation of UNIFEM’s <b>MYFF</b> System	2007	Yes	Yes	Yes	Yes	Yes	Yes

## 2.2 Findings related to Evaluation Quality

The quality of the evaluations has been assessed using 16 different parameters that are largely derived from the UNEG standards. For 13 of the parameters, each evaluation was awarded a score between 0 and 3. For the remaining 3 parameters – i) the inclusion of safeguards, ii) debriefing of stakeholders and/or ensuring a validation process at the end of the evaluation; and, iii) the existence of a management response to the evaluation – a score of 0 or 1 was given. The table in Annex 2 provided the framework for the assessment. It specifies criteria for each parameter and score. The highest attainable score is 48 points. The evaluations between 2004 and 2007 received a total average score of 30 points.



Using multiples of 8, a grade range was established as follows:

**Very weak:** 16 points or under (1 evaluation)

**Weak:** 17 to 24 points (0 evaluations)

**Average:** 25 to 32 points (5 evaluations)

**Good:** 33 to 40 points (1 evaluation)

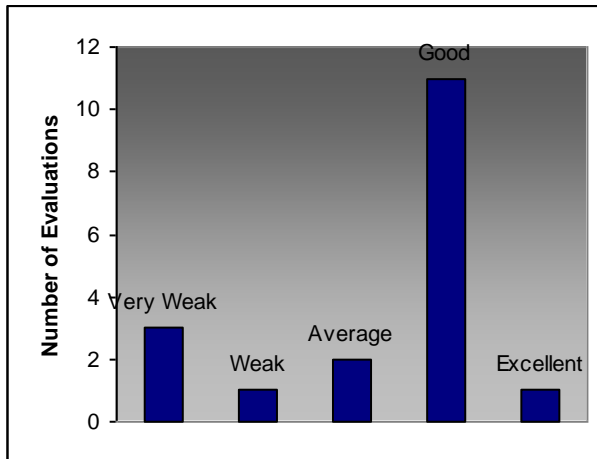
**Excellent:** 41 or more (1 evaluation)

*Table 3: List of Evaluations Studied from 2004 to 2007*

	<b>Evaluation 2004-2007</b>	<b>Score</b>		<b>Year</b>
1	Gender and Economic Justice in <b>European Accession</b> and Integration	<b>33</b>	Good	2004
2	Engendering Peace Building and Mediation Processes in <b>Burundi</b>	<b>30</b>	Average	2004
3	Evaluation of the Southeastern European Legal Initiative ( <b>SEELINE</b> )	<b>17</b>	Weak	2004
4	Women for Conflict Prevention and Peace Building in the <b>Southern Caucasus</b>	<b>25</b>	Average	2004
5	From <b>Global to Local</b> : A Convention Monitoring and Implementation Project	<b>33</b>	Good	2004
6	Enhancing Women's Economic Security: <b>IGAD Gender Strategy</b>	<b>15</b>	V. Weak	2005
7	Prevention of <b>Trafficking</b> in Women and Children in South Asia	<b>14</b>	V. Weak	2005
8	Reinventing India: Action for Empowerment to Address <b>Violence against Women</b>	<b>16</b>	V. Weak	2005
9	UNIFEM Programming in <b>HIV/AIDS</b>	<b>43</b>	Excellent	2006
10	Promoting Women's Human Rights & Enhancing Gender Equality in Democratic <b>Governance in Kenya</b>	<b>36</b>	Good	2006
11	Progressing the Implementation of <b>CEDAW in the Pacific</b>	<b>33</b>	Good	2006
12	Women's Rights to Land in <b>Kyrgyzstan</b>	<b>33</b>	Good	2007
13	Mainstreaming Gender in Poverty Reduction: <b>IFAD-UNIFEM Gender Mainstreaming Programme in Asia</b>	<b>37</b>	Good	2007
14	Mainstreaming Gender into the <b>Electoral Process in DRC</b>	<b>35</b>	Good	2007
15	Women for Conflict Prevention and Peace Building in the <b>Southern Caucasus, Phase 2</b>	<b>34</b>	Good	2007
16	Supporting Women's Leadership in <b>Rebuilding Afghanistan</b>	<b>35</b>	Good	2007
17	Gender and the <b>Millennium</b> Development Goals	<b>37</b>	Good	2007
18	Evaluation of UNIFEM's <b>MYFF System</b>	<b>40</b>	Good	2007

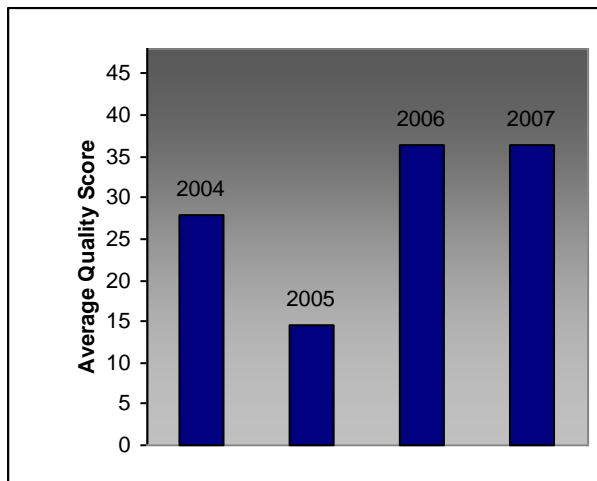
The graph below illustrates how the scored evaluations fared when given the corresponding grade:

Figure 7: Distribution of Quality Scores for Evaluation Conducted between 2004 and 2007



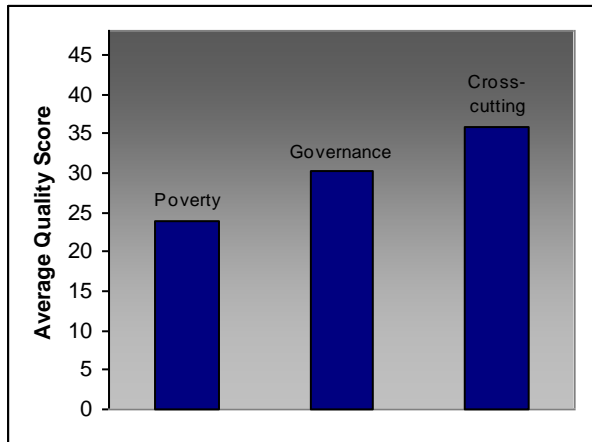
The quality of evaluations was higher for those conducted in 2006 and 2007. There was a sharp fall in quality between 2004 and 2005, as can be seen in figure 8:

Figure 8: Average Quality Score by Year



For the MYFF goal areas of EAW and HIV/AIDS, only one evaluation each qualified as an evaluation. This makes comparisons with the other strategic areas skewed. Therefore, in the table below, these goal areas have been omitted, leaving only 3 thematic areas in the table. The average score related to the goal area of poverty was comparatively lower and earned a “weak” mark. The evaluations in the area of governance scored “average”; while all the cross-cutting evaluations were assessed as “good”.

Figure 9: Average Quality Score by Thematic Area (HIV/AIDS and EAW excluded)



Within each region, there is a spread of different scores. There is at least one weak or very weak evaluation for each region and at least two good evaluations. However, among the global evaluations, the scores were higher and none scored under “good”. Together they averaged 36 points.

### 2.2.1 Basic key information

The basic key information that the title page (or the first few pages) of an evaluation report should include are the subject (i.e. activity, programme, policy etc.) being evaluated; name of the evaluators; name of the organisation(s) that commissioned the evaluation; and the date. In most cases, the title pages of evaluation reports contain this information. In some cases it became clear who commissioned the report only after opening the document. For a couple of evaluations, it was not clearly evident who had written the report or who had commissioned without reading the report. Nevertheless, the evaluations generally scored very highly in relation to this parameter – resulting in an average of **2.9** out of 3.

### 2.2.2 Executive summary

Over one-third of the evaluations had no executive summary. On the other hand, of those that did, nearly half scored full points. The average score for the quality of the executive summary was **1.7**.

### 2.2.3 Evaluation purpose, objectives, scope and methodology

Only 4 evaluations scored full points for explaining the purpose, objectives and scope of the evaluation and providing a clear account of the methodology applied. Typically, evaluations failed to explain the limitations of the evaluation process. The average score for the presentation of the purpose, objectives, scope and methodology was **2.1**.

### 2.2.4 Context of subject

Six evaluations scored full points for providing an analytical account of contextual background (institutional, political, socio-economic, cultural, geographical or any other relevant context information) that contributed to the utility and accuracy of the evaluation. This included providing relevant statistics to support the contextual analysis. One evaluation provided no contextual information at all, and five evaluations did not provide adequate accounts of the context. The average score for the contextual analysis was **1.9**.

### **2.2.5 Description of the subject**

One-third (or 6) evaluations clearly described the subject being evaluated. Two evaluations provided no information at all. The majority of the evaluations were not able to provide budget figures for the project/programme being assessed. Furthermore, they tended not to provide detailed information on the project/programme's management structure, stakeholder roles and their contributions. They also often failed to explain the expected results chain and intended impact of the programme/project being evaluated. The average score for describing the evaluation subject was **2.0**.

### **2.2.6 Findings**

Findings are factual evidence, data and observations that are relevant to the specific questions asked by the evaluation. Because of their central importance in assessing the quality of an evaluation, findings were scored on a scale of 0 to 5. A third of the evaluations provided quality findings that earned them full points. And another third provided findings that were of above average standard. The final third presented findings that scored "3" (average) or below. Among these evaluations, problems included the failure to sufficiently measure the extent outputs and outcomes have been achieved and/or presenting findings without sufficient evidence and triangulation. Sometimes a problem may have been that the project document had unclear objectives, outputs and outcomes. In one case, the evaluation team addressed this problem by reconstituting the objectives together with the project staff to facilitate the evaluation process, resulting in a report with a solid and logical findings section. The average score for the presentation of findings was **3.6**.

### **2.2.7 Analysis**

In an evaluation report, analysis should be structured as a logical flow so that data and information are presented, analysed and interpreted systematically. Reasons for accomplishments, difficulties, constraints and enabling factors are identified and analysed to the extent possible. Because of its importance in assessing the quality of an evaluation, the quality of the analysis undertaken in each report was scored on a scale of 0 to 5. Just under one third of the evaluations provided quality findings that earned them full points. Half of the evaluations included analyses that were of average standard or below. A common problem was the failure to structure the analysis logically and systematically. The average score for analysis was **3.6**.

### **2.2.8 Conclusions**

Conclusions are the assessment of the intervention and its results against given evaluation criteria, standards of performance and policy issues. Because of their importance in assessing the quality of an evaluation, the quality of the conclusions in each report was scored on a scale of 0 to 5. Less than one-fifth of the reports scored full points in this area. The conclusions of 56 per cent of the evaluations were of average standard or below, and half of these were weak or very weak. Even among the otherwise good evaluations, the overall conclusions were weak because the evaluators failed to summarise and at the same time lift the analysis to a higher level to add value. In several cases, evaluators were not successful in providing well-argued insightful conclusions for each evaluation criterion. The conclusions sections were often mixed with lessons learnt, without distinguishing between them. The average score for the quality of the conclusions was **3.2**.

### **2.2.9 Recommendations**

While over a third of the evaluations provided good enough recommendations to earn them full points, even among these, the presentation of the recommendations could be improved by structuring

them as a *numbered statement in bold type* with a *subject*. An example of a well-present, clear and concise recommendation is provided in the box below.

#### Example of a Well-Presented Recommendation

1. **UNIFEM should speedily proceed with, and provide regional and sub-regional offices with, required support and advice on the implementation of its decentralisation strategy.** The decentralisation process is eagerly awaited by several if not all of UNIFEM's field offices, and carries considerable potential for improving the speed and efficiency of various administrative tasks, and of decision-making processes.

(From the *Evaluation of UNIFEM's MYFF System*, 2007)

The most common fault of the evaluations that did not receive full points for the recommendations was not being formulated in a clear and concise manner. Recommendations were in some cases hidden in the text and priorities for action were not clear. For example, UNIFEM is one of two implementing partners in IFAD's Programme for Mainstreaming Gender in Poverty Reduction in Asia. Thus the recommendations in the evaluation of this programme do not pertain to UNIFEM. The average score for the quality of the recommendations provided was **2.2**.

#### 2.2.10 Lessons learnt

Lessons learnt are general conclusions that are likely to have a potential for wider application or use. Not all evaluations generate lessons. However, if it is the case that no lessons were generated, it should be stated in the report. Almost half of the evaluations did not include lessons learnt at all, or provided ones of poor quality (typically conclusions that were not contributions to general knowledge). Nevertheless, 4 evaluations provided valuable lessons. The average score for the quality of the lessons provided was **1.6**.

#### 2.2.11 Gender equality and human rights

Since virtually all evaluations concern gender equality<sup>5</sup> and therefore discuss gender issues in one way or another, almost all evaluations scored at least one point in this area. Few evaluations, however, examined if and to what extent a quality gender analysis underpinned the project/programme's assumptions and theory of change. With regard to human rights, the majority of evaluations made at best only a passing reference to them. Nevertheless, five evaluations used rights-based language to some extent and/or analysed results in relation to promoting rights. Most of these evaluations concerned projects/programmes with a pronounced rights focus, e.g. *Women's Rights to Land in Kyrgyzstan*, (land and women's human rights); *Reinventing India: Action for Empowerment to Address Violence against Women* (the rights of women affected by violence), the *Global to Local* evaluation, ( monitoring the implementation of CEDAW); and, *UNIFEM's Programme on Enhancing Human Security through Gender Equality in the Context of HIV/AIDS*, which aimed to integrate gender equality and human rights into key policies, plans, and activities that address HIV/AIDS at the national level. The evaluation *Women for Conflict Prevention & Peace Building in the Southern Caucasus* also scored higher because it systematically analysed the lack of a rights-based approach in the programme and looked into opportunities to do so. However, none of the evaluations fully lived up to UNEG's current standards in this area, let alone applied innovative rights-based methodologies. The average score for considering gender equality and human rights issues was low: **1.3**.

---

<sup>5</sup> An exception is the MYFF evaluation

### **2.2.12 Stakeholder participation**

The extent to which stakeholders participated varied. Only one evaluation scored a “3” – it was the assessment of *Women’s Rights to Land in Kyrgyzstan* which applied an innovative, well-described and fully participatory methodology. Otherwise, evaluations tended to be more or less consultative. The evaluations that scored “2”, explained in the methodology section how they used focus groups, questionnaires and involved a large number of stakeholders. Two evaluations involved virtually no consultations. The average score was **1.3**.

### **2.2.13 Clear communication**

Fourteen evaluations used clear, precise and professional language and earned full points in this area. However, the language of 1 evaluation was sub-standard and 3 reports were very poorly written. Although UNIFEM has sizeable projects in Francophone and Spanish-speaking countries, there are almost no evaluations from these areas and all reports are in English.<sup>6</sup> The use of graphs, illustrations and tables was, if at all, limited for most evaluations. The average score was **2.6**.

### **2.2.14 Ethical safeguards**

None of the evaluations included a discussion of the extent to which the evaluation design included ethical safeguards where appropriate. This would have included protection of the confidentiality, dignity, rights and welfare of human subjects – including children – and respect for the values of the beneficiary communities. By listing beneficiaries who returned questionnaires by name and identifying them in the report, one evaluation is ethically unsound.

### **2.2.15 Validation process**

An evaluation, particularly one that purports to be consultative, should include a verification and validation process with key stakeholders. This provides an opportunity to correct factual errors and confirm the validity of conclusions. Such a process is likely to augment the quality of the workshop and its future utility. It can take the form of workshop(s), seminar(s), e-conference(s) etc. before the evaluation team drafts the evaluation report. From what can be gathered from the methodology sections of the reports, about a third of the evaluations undertook some form of validation process with stakeholders.

### **2.2.16 Management response**

Within UNIFEM, concerned management entities are required to produce an explicit response to an evaluation’s recommendations. There are management responses to 4 evaluations, all undertaken after 2007. Three of these evaluations were managed by headquarters.

---

<sup>6</sup> The one evaluation undertaken in a Francophone country – Gender Mainstreaming in the Electoral Process in DRC – but the report was in English.

### 3. Utility and Credibility

An evaluation's *utility* depends on how useful the evaluation has been/will be for its stakeholders for internal learning, accountability and/or management purposes. *Credibility* is the extent to which stakeholders deem the evaluation process and results to be believable. It is determined by the extent that i) the evaluation is perceived as being independent, impartial and/or lacking conflict of interest; and ii) the evaluator is perceived as being familiar with the context and has expertise in evaluation and the subject matter under investigation. An evaluation of high quality is more likely to be useful and credible than one of poor quality. However, high quality does not necessarily guarantee credibility and utility.

To gain an understanding of the utility and credibility of UNIFEM's past evaluations, interviews were undertaken with some stakeholders (UNIFEM staff and evaluators) involved in nine evaluations from between 2006 to 2008<sup>7</sup> that together represent a geographical and thematic spread. Two of these evaluations scored "excellent", two scored "average" and the rest scored "good". While the number of interviews was relatively limited (see methodology section), they sufficed to provide a general snapshot of utility and credibility and what factors were believed to have contributed to these qualities.

Eight<sup>8</sup> of the evaluations were used to prepare a second phase of the intervention (CEDAW-SEAP, Kyrgyzstan-Land, Kenya GE&HR, Kenya G&G, Afghanistan WL, GRB-LAC) or a new strategy (MYFF, HIV/AIDS). In the case of CEDAW-SEAP, the evaluation was also used to discuss UNIFEM's work with CEDAW at a recent internal workshop for 50 staff members from around the world. This suggests that evaluations are contributing in important ways to UNIFEM's work and its learning processes. In two cases, however, the evaluations were not as useful as expected by UNIFEM. While much of the evaluation content was taken to heart, UNIFEM had hoped for more evidence, analysis and ideas to guide its future work. In one of the cases, the terms of reference did not sufficiently cover UNIFEM's need. In the other, the evaluators did not fully meet the demands of the terms of reference.

In terms of credibility, discussions revealed that there appeared to be a general correlation with the quality score earned by each evaluation. Thus, evaluations like the CEDAW-SEAP (score of 43) was regarded as highly credible while it was felt that GRB-LAC (score of 27) could have covered certain issues better (but was by no means considered incredulous).

The interviewees considered aspects of the evaluation and its process that they believed contributed to the utility, credibility and quality of the evaluations. These are discussed below.

#### 3.1. Terms of reference

Most of the stakeholders interviewed believed that having clear and detailed terms of reference from the start led to better quality and more useful evaluations. The process of developing the terms of reference was also considered important.

---

<sup>7</sup> Since UNIFEM's evaluation work started to change in 2008 with the existence of the Evaluation Unit, 2008 evaluations were included to make the data collected more up-to-date and therefore more useful.

<sup>8</sup> Data on the utility of Kosovo GE&WR was not gathered.

The interviews revealed that the processes of drafting the respective terms of references were relatively consultative. Terms of reference were shared with relevant staff within the organisation and sometimes with external partners. Some of the terms of references were short on specifics; others were detailed – even outlining the exact methodology to be used by the evaluators.

For the short terms of references, the consultants ended up spending considerable time with UNIFEM at the beginning of the process jointly determining the key evaluation questions to be addressed and what methodologies to use. An advantage of this approach was that it allowed the evaluation team to help shape the assignment and bring in an outsider's ideas and perspectives to the table. While in the cases examined, this approach did result in either good or excellent quality scores, this is potentially a risky approach: the evaluators and UNIFEM can start off the process with very different ideas and expectations of what is required. If the preparatory phase is drawn out and/or contentious, goodwill can be eroded, time can be wasted and costs can rise.

An innovative approach to preparing the terms of reference for an evaluation was tested by the managers of the Kyrgyzstan-Land evaluation. In keeping with the participatory nature of the project, the managers decided to maximise participation throughout the evaluation process. A suggested participatory approach for the evaluation was drawn up, drawing on good practices of other UN agencies. A broad range of stakeholders – government officials from all provinces, NGOs, community organisations, etc. – were invited to participate in a full day's workshop to discuss the evaluation. Detailed discussions were held on the methodology, organisational issues and how to best design the evaluation to ensure that the data collected was aligned with the project's indicators, but at the same time avoided imposing ideas on the beneficiaries. This process was instrumental in gaining greater ownership – an important factor given the logistical challenge the evaluation would face by applying participatory techniques in seven provinces (nine sites) and involving 161 beneficiaries. The terms of reference that resulted from the workshop were clear, specific, methodical and include guidance notes.

In one instance, the evaluation was co-managed with a donor and the terms of reference were established jointly. While this joint management contributed to the budget of the evaluation, a drawback of the set-up was the resulting emphasis in the terms of reference which favoured the donor's agenda. In this case, the evaluation team spent considerable effort establishing the added value of UNIFEM in the programme, which was not a concern of UNIFEM, but of the donor. Meanwhile, UNIFEM stakeholders felt that the evaluation did not provide sufficient insight to help determine what components to scale-up and include in the second phase of the programme. Thus the evaluation was less useful than hoped.

The stakeholders of the CEDAW-SEAP evaluation (which scored highest on quality) deemed that the clarity of the terms of reference contributed significantly to a smooth evaluation process as well as the quality, utility and credibility of the evaluation report. The ToR provided a concise background and clearly outlined the purpose; scope; evaluation questions; methodological approach; evaluation management set-up; evaluation products and time line; composition, skills and experience of the evaluation team; and, the ethical code of conduct for the evaluation. The level of detail was considered suitable. (The format appears to have been since adopted as a standard format by the Evaluation Unit.)



### 3.2. Evaluation teams

It was recognised that the expertise of the evaluators was important for the credibility and utility of the evaluation. However, this was not considered enough. Dialogue between UNIFEM and the evaluation team also had to be good. In a couple of cases, the dialogue was not optimal.<sup>9</sup> In fact, both evaluators and UNIFEM staff interviewed agreed that the rapport and understanding established between UNIFEM and the evaluation team was crucial. It allowed for a smooth evaluation process – which in itself contributed to quality and a more credible and useful evaluation. In one case, evaluators undertaking an assignment never travelled to New York at the start of the assignment to meet the manager. In hindsight, it was believed that this was an error which made the evaluation process more challenging.

Evidence suggests that UNIFEM's request for proposals (RFP) usually left consultants with two-to-three weeks to submit a proposal. Some of the UNIFEM interviewees had encountered difficulties in identifying suitable candidates for evaluation tasks after issuing an RFP. Different solutions were sought. For instance, in one case, the managers paired together consultants who had previously not worked together. In a second case, an NGO was hired to assist with collecting data at the field level, while a national research institute was given the task to analyse the data. While these approaches resulted in decent evaluations, the evaluation managers did not feel that the process and/or evaluation achieved optimal quality, credibility and/or utility. In the second case, the manager had to step in to help strengthen the analysis of the final report.

Unsolicited, the issue of local/national/regional consultants was raised in a number of interviews. All agreed that national consultants, if well qualified, could contribute to the quality, credibility and utility of an evaluation. UNIFEM field staff mentioned being faced with the dilemma of choosing between national consultants who knew the subject area and languages well, but often lacked sufficient evaluation and/or analytical skills and experience; and international consultants, who were considered insufficiently knowledgeable of the local context to provide the services required.

Evaluators, on the other hand, raised the difficulty of identifying skilled national consultants, particularly at short notice. Delays in decisions often meant that country case studies were chosen at a fairly late stage in the evaluation. Contracting a national consultant with relevant qualifications and no conflict of interest was often a challenge. When UNIFEM stipulated the need to include national consultants on the evaluation team, it was suggested that clear guidelines were provided for what role the national consultants should play – interpreter, support to data collection effort, workshop facilitator, evaluator etc.

---

<sup>9</sup> When reading some of the evaluation reports, the meta-evaluation detected signs that suggested friction during the evaluation process. This was always coupled with evaluation reports of lesser quality. Subsequent interviews confirmed that these evaluation processes had not panned out as well as had hoped. On the other hand, evaluations of high quality all enjoyed fairly smooth evaluation processes. This highlights the importance of good communication between the evaluators and UNIFEM throughout the evaluation process for a high quality product.

### 3.3. Stakeholders role in the process

Most interviewees commented on the importance of ensuring that key stakeholders within UNIFEM as well as partners were brought on board early in the evaluation process. Participation enhanced ownership, strengthened commitment to the process and contributed to the evaluation's utility. The CEDAW-SEAP evaluation was the first to **establish an intranet site** for the evaluation to encourage participation and the interviewees deemed it contributed to the success of the evaluation.

Evaluators valued involving stakeholders within the organisation in the process because it mobilised the relevant parts of UNIFEM to provide logistical support; facilitate access to documents and data; identify informants; and, organise meetings. (In fact, UNIFEM was considered particularly collaborative in this regard.).

The evaluators and UNIFEM staff interviewed highlighted the importance of holding an activity before the report was drafted to verify findings and validate conclusions. They strongly believed that this contributed to the evaluation's quality, utility and credibility. Generally, this took the form of a workshop, meeting or teleconference during which the evaluation team presented key findings, conclusion and recommendations.

A consultative evaluation process lengthens the evaluation process. It appears that in several instances, the evaluation timeline had to be adjusted several times. Some interviewees held that this was because the timelines were not realistic to begin with, given UNIFEM's relatively slow turnaround time. It was deemed that working a realistic timeline from the beginning would strengthen the evaluation process.

Several of the UNIFEM staff interviewed felt that they had gained significant useful experience by being part of an evaluation. Staff involved in more recent evaluations expressed (unsolicited) that they held the support provided by the Evaluation Unit in very high regard. The support included comments to terms of reference, assistance in procurement of evaluators and advice and backstopping along the way.

## 4. Analysis of Evaluation Quality

This section analyses important findings from the previous chapters.

The aggregate quality score for the 18 UNIFEM evaluations conducted during 2004 and 2007 is 30 or “average”. However, a respectable 60 per cent of the evaluations have scored “good” or above. This, nevertheless, needs to be seen in the light of the fact that as many as forty per cent of the 30 “evaluation” reports submitted to the Evaluation Unit actually turned out to be reviews or other types of reports and were therefore not included in the aggregate scoring exercise. Mistaking so many reports for being evaluations suggests that there is insufficient conceptual understanding within UNIFEM of what actually constitutes an evaluation. This potentially undermines UNIFEM’s accountability and ability to learn from its work. Since evaluation is among other functions a management tool, improving the understanding of evaluation – purposes, processes and uses – will not only enhance the afore mentioned aspects, it is also likely to enhance UNIFEM’s results-based management.<sup>10</sup>

The mapping and scoring exercise reveals some patterns regarding evaluation characteristics and the quality score. First, **global and cross-cutting** evaluations scored “good” or “excellent”. Second, evaluations that were managed by UNIFEM’s **headquarters or co-managed with an external partner** – either a donor or another UN agency – earned, on average, 35 points (a solidly “good” score). Third, evaluations of particularly **high profile**, also scored well. Fourth, evaluations that included an activity to ensure stakeholder participation in the **verification and validation** of the evaluation’s key findings and conclusions before the report was drafted, scored an average of 38 points.

When examining the seven top-scoring evaluations (most of which can be considered high-profile evaluations) all seven fall into one or more of these categories:

- Managed by UNIFEM headquarters and of especially strategic importance to UNIFEM (MYFF, HIV/AIDS, MDG, Afghanistan-WL).
- Managed by or together with other UN agencies (by IFAD for the Gender Mainstreaming Programme in Asia and with or by UNDP for Gender and the Millennium Goals)
- Managed with or by donors (the Kenya G&G funded by Sida; DRC-Election funded by DfID).
- Included a validation activity with stakeholders as part of the evaluation process (MYFF, HIV/AIDS, MDG, Afghanistan WL, DRC-Election, Kenya G&G).

Another factor that seemed favourable for higher scores was the length of the evaluation process. While this does not necessarily reflect the amount of time used by the evaluators for the work (for instance, a multi-member team can dedicate more hours to the evaluation in less calendar time or evaluators may only work halftime or less during the evaluation period), it gives some indication of the input of time. Predictably, there is a correlation between short evaluation processes and low quality scores. Of the four evaluations with 3-week processes, one scored very weak, and two scored

---

<sup>10</sup> This is discussed further in Chapter 2 of Part 2.

average. Meanwhile, three of the four evaluations that had a process over 3 months at least scored “good”.

Finally, a clear advantage for evaluations was having a **team of evaluators** as opposed to a single consultant. Half of the evaluations were conducted by teams and all but one of these scored either “good” or “excellent”. On the other hand, among the nine one-person evaluations, only one-third scored “good” while the remainder scored average or below. This should come as no surprise. Successful consulting firms in most sectors have for decades realised the quality gains reaped by offering clients a team of experts with complementary skills. Through a co-ordinated effort, the team members generate synergy, which allows each member to maximise his or her strengths and minimise his or her weaknesses. The sum thus becomes greater than its parts. Furthermore, if the evaluators come from an established firm, chances are that there are support staff in the background and quality assurance mechanisms that the evaluation benefits from.

The characteristics that seem to be associated with higher scores either signify higher costs (multi-member teams, longer evaluation processes and stakeholder validation and verification) or indicate that the evaluation had a relative robust budget (headquarters managed or managed jointly with a donor or UN agency).<sup>11</sup> This seems to suggest that you get what you pay for.<sup>12</sup>

## 4.1. Areas for Improvement

Even among the evaluations of higher quality, there are several areas where UNIFEM’s evaluations could improve. These could be divided into reporting and methodological shortcomings. The sections that follow analyse these two aspects.

### 4.1.1. Report Content & Structure

Many of the reports have deficiencies in relation to content and structure. For instance, only four evaluations were complete when it came to basic report content. While reports seemed to have improved in this regard over the period evaluated, this needs to be addressed to ensure that the evaluations are credible and useful.

Other content-related shortcomings were the often weak contextual backgrounds and descriptions of the evaluated intervention – including its goals, logic, history, organisation and stakeholders. Without these elements, the potential readership of the evaluation and the possibility to learn lessons is limited.

The concluding sections of UNIFEM’s evaluations are another weak spot, even among some of the high scoring evaluations. As discussed above, evaluators failed to summarise the analysis and lift it to a higher level to add value. Furthermore, some evaluations either did not provide conclusions for each evaluation criterion or did so without accompanying them with well-argued and insightful text. There were misunderstandings about conclusions and lessons learnt, in some cases probably stemming from the way the term “lessons learnt” was originally used in the terms of reference for the evaluation.

---

<sup>11</sup> As discussed in the introduction, it was beyond the scope of this evaluation to study the evaluation budgets.

<sup>12</sup> Some of the interviewees – both evaluators and UNIFEM staff – found the evaluation budgets they worked with to be too small for the tasks at hand, requiring evaluators and sometimes UNIFEM staff to contribute labour beyond what was included in the budget.

### 4.1.2. Methodology

There are opportunities to strengthen several aspects relating to the methodological approaches of UNIFEM's evaluations. Only four evaluations scored full points for explaining the purpose, objectives and scope of the evaluation and providing a clear account of the methodology applied. A common shortcoming was a failure to consider the limitations of the methodology. In addition, more than half of the evaluations had no methodological tools to show for in the report or annexes.

Related to methodology but scored separately in this meta-evaluation are the extents to which the evaluations were **participatory**, applied a **gender equality and human rights** perspective in the methodology and included **ethical safeguards** in the evaluation design. In all three areas, there is considerable room for improvement.

Only one UNIFEM evaluation – the Participatory Assessment of Women's Right to Land in Kyrgyzstan – applied **participatory techniques** throughout the evaluation process. A handful showed a solid attempt at participation, but most evaluations were consultative at best.

Ensuring high levels of participation can be potentially costly and time-consuming. An exclusively participatory approach may also not be the best way to analyse certain management and performance dynamics. In addition, participatory techniques that, for instance, allow people to identify their own indicators of success, may not be compatible with a project that has a pre-defined logical framework.<sup>13</sup> Nevertheless, there is widespread acknowledgment in the evaluation community that participation by key stakeholders significantly improves relevance and outcomes of the intervention and enhances the utility of the evaluation. Furthermore, the process itself can empower local people to initiate, control and take corrective action.

#### Participatory Techniques: Kyrgyzstan Good Practice

The participatory assessment in Kyrgyzstan – including its terms of reference, the methodology used and how the results have been presented – represents a good practice evaluation that UNIFEM can learn from. The approach resulted in the evaluation being able to concretely provide extensive evidence of outcomes and impact in a more convincing way than most of the other evaluations.<sup>14</sup> Contributing to this evaluation's successful outcome were i) a very clear set of terms of reference (themselves derived from a participatory process) that were very specific in providing details on what form participation should take; ii) buy-in from a range of stakeholders who supported the evaluation process; iii) the existence of baseline data; and iv) the evaluation being undertaken by a team who had sound knowledge of the local context.

When it comes to **gender equality and human rights**, only five evaluations made some (though not enough) effort to integrate these perspectives in its analysis. This included, for instance, using some rights language and identifying results in terms of rights. Often the extent of the human rights perspective was dependent on the degree that the project/programme focused on human rights. None of the evaluations actually analysed UNIFEM's attempts at applying a rights-based approach – although one evaluation included a discussion on the opportunities for human rights approaches

<sup>13</sup> The same may be true of appreciative inquiry, a participatory approach that focuses on successes and may not address all expected outputs and outcomes in a logframe.

<sup>14</sup> The Morocco Literacy evaluation from 2008 provided convincing evidence of outcome, impact and lack thereof. It also engaged actively with beneficiary populations and, like the Kyrgyzstan evaluation, had a baseline to refer to but was not designed in an equality participatory way.

within the context of the project. Nor did the evaluations examine the gender equality and/human rights analysis underpinning the interventions – in some cases this was because there was one. Part of the problem is that it has taken time for the UN’s rights-based approach to be fully internalised. Arguably, the approach has yet to permeate all aspects of UNIFEM’s work. Consultants, too, often have insufficient capacity to understand the implications of this approach on evaluations. The new UNEG guidelines of integrating gender equality and human rights in to evaluations is a welcomed tool that can potentially raise performance in this area. However, to promote understanding of the extent to which gender equality and human rights permeate all aspects of its interventions, UNIFEM will need to stipulate that evaluations specifically examine and assess this.

**Ethical safeguards** were not addressed by a single evaluation methodology. Given that UNIFEM’s work spans several sensitive areas (e.g. violence against women and children, women at risk/affected by/living with HIV/AIDS), thought needs to be given to safeguards. Meanwhile, one evaluation breached UNEG standards in relation to privacy and confidentiality. In this case, the onus rests not only on the evaluation team, but also on the evaluation manager within UNIFEM to deny approval of a report that violates confidentiality.

Overall, there are not many examples of **innovative methodological** approaches. Document reviews, interviews and possibly a couple of site visits tend to be the norm. Surveys were used in some instances but without any significant successes. In a handful of evaluations, focus groups were used to gather data. There are a number of ways in which data collection could be made more innovative. Approaches and techniques that could be useful in UNIFEM evaluation contexts include, for instance, participatory Strength, Weaknesses, Opportunities and Threats analysis (SWOT) workshops, appreciative inquiry, outcome mapping, and the most significant change approach.

#### **Surveys:**

Data gathered by survey can be very useful to assess the merit or worth of a development intervention. Before launching into a survey, thought needs to be given to whether it is the most appropriate way to collect reliable and valid information that contributes to answering the evaluation questions. A survey for the sake of it is a waste of resources and, at worst, can jeopardise an evaluation. A survey should only be stipulated by a terms of reference if there already is a very clear idea of **what** type of information should be collected, **who** should be canvassed and **how** the data should be gathered.

Conducting surveys using a web-based applications (such as survey monkey) has facilitated the process enormously. It allows for a neat design, efficient launch, instant computations and substantial possibilities and ease in analysis. However, unless all stakeholders to be targeted by the questionnaire have internet access, the survey will face a coverage bias.

The importance of wording questions suitably so that they are answerable, quantifiable and do not encourage biases cannot be overstated. To encourage a high response rate, it is important to minimise the burden on the respondent by designing questionnaires that are attractive in appearance and easy to complete. Every word and phrase in the draft questionnaire needs to be carefully weighed and then translated into the relevant languages. Preferably the questionnaire should be piloted to identify potential problems. Surveys that take ten minutes or less to complete have more successful response rates, avoiding a non-responsive bias.

One optional open-ended evaluation question can be included in the questionnaire to provide qualitative data. However, a questionnaire containing several open-ended questions may result in fewer responses and unmanageable data. Another way to gather qualitative views using the internet is to establish an evaluation site to which all stakeholders can contribute with views, stories, ideas, good practices etc. The evaluator can establish the boundaries and rules and screen contributions. The set-up may not work for all evaluations, but when it does, it produces rich data.

### 4.1.3. Way Forward

Without having had an evaluation unit or an evaluation strategy until recently, and still no evaluation policy<sup>15</sup> of its own (a requirement of UNEG standards), UNIFEM is not yet where it should be in terms of evaluation practice. This undermines its learning, results-based management and accountability. The evidence uncovered by the meta-evaluation shows that while there has been a significant amount of quality evaluation work within UNIFEM, a few evaluations have been very weak. The fact that 40% of the assessments received by the evaluation unit do not meet the criteria for what constitutes an evaluation reflects knowledge gaps within the organisation. Furthermore, certain sub-regions have not produced a single evaluation during this period (e.g. Caribbean, West Africa).

Improving the credibility, utility and quality of UNIFEM's evaluations requires **informed and skilled evaluators**, combined with **competent management** of the evaluators and evaluation process. The latter will entail increased capacity within UNIFEM and among its partners. Programme staff and managers need a sound understanding of evaluation concepts and processes. Staff with monitoring and evaluation functions requires more in-depth knowledge.

Significant efforts are already underway to address the need for greater capacity within UNIFEM by rolling out a training plan for staff and partners and preparing guidelines and tools. The newly established Evaluation Unit is leading this process and is already setting its mark. It has also engaged in (much appreciated) backstopping for UNIFEM's evaluations efforts in the last year; set up an intranet website, gradually developed internal systems and has drafted an evaluation strategy.

Central to successful evaluation management is the existence of a clear and sufficiently detailed set of terms of reference. The importance of this instrument for the evaluation process and resulting report cannot be over-stated. UNIFEM has often applied consultative approaches to developing ToRs, which is a good practice, as long as competing interests do not reduce the ToR to being a smorgårbord of activities, instead of a focused document with clear purpose and aims. The weaknesses of UNIFEM's evaluations that have been identified by the meta-assessment (content of report, gender & human rights perspectives, ethical safeguards, etc.) can be addressed by stipulating appropriate actions in the terms of reference.

Ensuring that UNIFEM evaluations are conducted by informed and skilled evaluators is more nebulous than building in-house capacity. UNIFEM nevertheless has some influence over this. First, UNIFEM can make a greater effort to attract evaluators to bid on its evaluation assignments. Providing clear and detailed terms of references with briefing material, sufficient time to prepare proposals and longer timeframes for evaluation processes are means within UNIFEM's control that can encourage evaluators to consider submitting a proposal. Naturally, a wide range of channels must be used to disseminate the RFP and a reasonable evaluation budget is needed. Second, UNIFEM can aim to communicate better with the resource base of evaluators to both attract their interest and inform them of UNIFEM, its evaluation effort, standards and its expectations.

---

<sup>15</sup> UNDP has an evaluation policy to which UNIFEM could in theory turn to, but it is being updated.

## 5 Recommendations: Improving UNIFEM's Evaluation Effort

UNIFEM is in the process of scaling up its evaluation effort. A number of processes are starting to be put into place. The following recommendations, presented roughly in order of priority, are based on the analysis of the quality, utility and credibility of the evaluations undertaken in the recent years; and this consultant's experience as an evaluator and manager of evaluations.

### 1. UNIFEM's Evaluation Unit should continue and further reinforce its efforts at raising the evaluation capacity within the organisation.

Staff need to have a solid basic understanding of key concepts. This includes, but is not necessarily limited to:

- The difference between monitoring, a review and an evaluation
- The meaning and implications of the five OECD/DAC evaluation criteria (relevance, effectiveness, efficiency, sustainability and, in particular, impact)
- The general content of the UNEG norms, standards and guidelines
- Different types of evaluations
- The basic content and structure of a terms of reference for an evaluation
- The basic content of an evaluation report, as per UNEG standards
- The ethical considerations in evaluations

UNIFEM's Evaluation Unit has planned to roll out a training programme to address the organisation's evaluation capacity and has already providing backstopping to staff members managing evaluations. A template for terms of references with an accompanying guide has already been prepared and the Unit aims to prepare additional guidelines and tools. The Unit should implement its current strategy to augment the organisation's evaluation capacity, taking into consideration the relevant recommendations below.

### 2. When procuring an evaluation, UNIFEM's evaluation managers should ensure there is sufficient time for tendering companies/consultants to present proposals.

UNIFEM has found it difficult to attract enough highly experienced consultants to bid on its evaluations. Evidence suggests that UNIFEM typically gives consultants about two-three weeks to prepare proposals. However, the longer consultants are given to prepare a proposal, the greater the likelihood that UNIFEM will receive quality proposals – both in terms of content and proposed candidates for the evaluation team. It is also likely to receive more proposals to choose from. Writing a proposal and compiling the necessary documents may take two weeks, but assembling a team of qualified and available consultants will often take much longer. Companies seldom have consultants sitting around waiting to be contracted. Top consultants are often booked many months in advance. One month should be the *minimal* time that is allocated for consultants to respond to a proposal request.

Giving consultants a longer time range than needed during which the evaluation has to be completed will also attract better proposals. It will allow for the participation of consultants who may be booked in the most immediate weeks but who are free in the subsequent months. Thus, if



a report is required by August and the evaluation is deemed to take two months to complete, the consultants would be given the period from April to August to undertake the evaluation.

This will of course mean that UNIFEM has **to plan its evaluations well in advance**. A good practice by some donors is to publish an annual evaluation plan with basic information on the internet so that aspiring consulting firms in turn can plan for proposal writing. This means that when the request for tenders is launched, keen consulting firms will have mobilised themselves to prepare a proposal.

### **3. UNIFEM's management should promptly make its evaluations available to the public.**

To date, UNIFEM has not published any evaluation reports or made them publicly available on the internet. UNEG's standards require that UN agencies have a transparent dissemination policy and make their evaluations publicly available, except in those cases where the reasonable protection and confidentiality of some stakeholders is required. Not only is such a policy important for transparency, learning and public accountability, UNIFEM should not expect staff and evaluators to live up to UNEG standards if the organisation's own policies are not in line with these standards.

### **4. UNIFEM's Evaluation Unit should further develop its intranet site to support colleagues in their evaluation work.**

Establishing an intranet site for evaluation within UNIFEM is a welcomed move. It currently contains evaluations reports from 2004 to 2008 and various reference material. However, the lay-out is not ideal, some reports are included more than once and others are missing. At least five of the reference links to evaluation tools do not work. **More effort is needed to make the website useful.** Until UNIFEM has its own guidelines and tools prepared, it is important that the website contains functioning and useful links. For instance, Sida's evaluation manual, *Looking Back, Moving Forward* (2<sup>nd</sup> revised edition) would be a valuable addition.

UNIFEM's Evaluation Unit should also **identify good practice evaluations** (such as CEDAW-SEAP and Kyrgyzstan-Land) and place them on the site for staff to learn from. This meta-evaluation could also become a tool in enhancing staff evaluation capacity to be placed on the intranet.

The Evaluation's Unit's plan to prepare a newsletter is an excellent idea. The newsletter should be placed on the intranet site and could even be used to inform on new items placed on the site.

### **5. UNIFEM's Evaluation Unit should consider collecting analyses of good practice evaluation processes.**

When a useful, credible and high quality evaluation has been completed, some evaluation units of development co-operation agencies request the consultant to prepare a synopsis of the evaluation methodology and process – including strengths, weaknesses and lessons learnt (a template for the synopsis is usually provided). Typically, the evaluator's contract is extended to include such a task after the draft or final report has been completed. UNIFEM could consider taking up this practice when evaluation processes have been successful, yielded useful results and/or involved applying innovative approaches.

**6. UNIFEM's Evaluation Unit should consider ways to broaden the resource base of qualified consultants that are willing to bid on UNIFEM's evaluation assignments.**

As mentioned above, UNIFEM has sometimes found it difficult to attract enough experienced consultants with relevant expertise to bid on its evaluations. Four of its best evaluations have been undertaken by the same consulting firm (procured through competitive bidding). Following recommendations 2, 3 and 4 above can contribute to receiving more and better proposals:

- Clearer and more detailed terms of references with accompanying briefing material make it easier for consulting firms to i) decide whether to bid; ii) prepare a proposal; and iii) identify appropriate consultants;
- More time for preparing tenders and a greater time span to undertake the evaluation allows for better quality proposals and more options in terms of evaluation team members;
- Access to UNIFEM's evaluations will allow prospective consultants to learn about UNIFEM and its evaluation effort.

The Evaluation Unit should consider developing a communications strategy towards the evaluation resource base that is based on active communication, openness and transparency. It could include:

- providing information on upcoming evaluations and examples of good practices on the UNIFEM website;
- preparing periodic newsletters (e.g. 2-4 times a year) that are posted on the internet or sent to potential evaluators directly;
- ensuring that short-listed consultants who do not win an evaluation assignment be given prompt feedback on the strengths and weaknesses of their proposals;
- inviting prospective evaluators to an annual information meeting and/or encouraging consultants to visit the Unit when in town to present themselves and to learn more about UNIFEM and its evaluation needs;
- networking with global and regional entities with evaluation expertise (e.g. development co-operation evaluation units, institutes, consulting firms, NGOs) to expand the database of known evaluators with gender equality expertise.

**7. UNIFEM's management should take into account the costs that improving evaluation quality will entail.**

As seen above, there is an (unsurprising) correlation between the size of the evaluation budget and the quality of the evaluation. Contracting evaluation teams (as opposed to single consultants); securing competent and knowledgeable evaluators; ensuring a consultative evaluation process throughout; including a verification and validation exercise; and, translating, publishing and disseminating evaluation reports will all contribute to raising evaluation costs. These costs are not only the fees and expenses of external consultants but also the cost of the time (and maybe travel) of other stakeholders – including the evaluation manager, programme staff, partner organisations, etc.

# Evaluation Results

## 6. Introduction

This document constitutes the second part of the UNIFEM Meta Evaluation and focuses on evaluation results. The background to and methodology used is provided at the beginning of Part1 of the evaluation.

The UNIFEM evaluations undertaken between 2004 and 2008<sup>16</sup> have provided some important findings, conclusions and recommendations for UNIFEM that deserve attention for institutional learning. These have been analysed and are presented in this document.

There are, however, several challenges involved in trying to draw overall conclusions about UNIFEM's work from the evaluations of this period. First, the terms of reference for the evaluations have differed in focus. Thus the evaluations have not always assessed comparable aspects of the programmes/projects. Second, the nature of the projects evaluated is in some respects specific to the context in which they are implemented. For example, literacy as a channel to teach women about their rights is relevant in the Moroccan context, but not in the Balkans. Third, in some cases, up to five years have passed since the findings and conclusions were made. It is quite likely that in several cases UNIFEM has taken steps to address the issues raised. It has, however, been beyond the scope of this evaluation to verify whether the findings and conclusions still hold true.

The first chapter looks at themes that have been raised. This includes management for results; capacity development; partnerships; regional approaches; institutional learning and knowledge-sharing; and, human resources and financial management. Upon request of the Evaluation Unit, it also includes a section on gender equality and human rights perspectives. The second chapter summarises how the evaluators have assessed the projects/programmes in relation to relevance, effectiveness, efficiency, sustainability and impact. While caution need to be applied to any generalisations about UNIFEM based on these evaluations, the following chapters may serve as a useful input to UNIFEM's work.

---

<sup>16</sup> To ensure relevance and illustrate the pervasiveness and extent of certain issues, this section not only analyses the results of the evaluations undertaken between 2004 and 2007, but also includes the evaluations from 2008 and, in a couple of instances, refers to reviews undertaken during this period. For the meta-assessment of the quality of evaluations conducted in 2008, please Annex 5.

## 7 Evaluation Themes

This chapter analyses how management for results; capacity development; partnerships; regional approaches; institutional learning & knowledge-sharing; human resources & financial management; and, gender equality & human rights are addressed and discussed in the evaluations conducted between 2004 and 2008.

### 7.1 Management for Results

Aspects related to results-based management (RBM) was an important theme in most evaluations. While in most cases difficulties and insufficiencies were raised, there were a few positive findings:

- The SEE-WR project was considered by its evaluators as having a solid project document with a logical framework of very high quality. It was prepared in a participatory manner with a range of stakeholders, who all praised UNIFEM's consultative planning approach. The logical framework permitted the elaboration of a clear work plan and effective monitoring of progress towards goals.
- Although it found several key challenges related to effective monitoring and reporting, the MYFF evaluation held that UNIFEM's internal and donor reporting demonstrated the organisation's capacity to monitor and track results. It regarded UNIFEM's effort to collect data at country, sub-regional and regional levels as impressive.
- The evaluators of the Morocco Political Participation programme found that the system and tools established to monitor progress were well-functioning and provided a wealth of useful data. UNIFEM's partners put in place a reporting process that provided baseline information on women targeted for the different activities undertaken. This made it possible to assess change.
- The Morocco Literacy project's baseline data allowed the evaluation to track qualitative changes using eight focus groups with beneficiary women.<sup>17</sup>

The other evaluations found fault with at least some aspect of the management for results. To begin with, a common problem with many of the projects and programmes evaluated was the setting of unrealistic, **overly ambitious goals** or scope, given the resources available, the time frame of the intervention and the capacity of UNIFEM (CEDAW-SEAP, India-VAW, HIV/AIDS, Kosovo GE&WR, Kenya G&G, DRC-Election, SEELINE, Caucasus-Peace II GRB-LAC, SA-Trafficking). Meanwhile, the Caucasus-Peace I and Afghanistan-WL evaluations reported that the goals of the programmes were not sufficiently clear. The Kenya GE&WR evaluation found that while the objectives were clear, outcomes and activities were disjointed from them. In the case of the DRC-Election evaluation, the project goals were designed based on a certain level of resources, but when the anticipated donations were not received, the project goals were not adjusted to take into account reduced resources.

Many evaluations also commented on **the design of projects** and programmes. While the programmes and projects were seen as designed in a relevant way given the needs, the evaluations

---

<sup>17</sup> Neither of the Moroccan reports provide any information on how the projects were managed. It is thus not possible to tell whether UNIFEM, its partners or both are to be credited for the successful approach to management for results.

found that research, baselines and/or analysis were missing. In most cases, no risk analyses were undertaken – a particularly important exercise in conflict-affected countries (DRC-Election, Afghanistan-WL, Kosovo GE&WR, Kenya G&G, Caucasus-Peace I).

Above all, many of the evaluations found that **an overall strategic perspective** was missing from the design. **Underlying assumptions and the corresponding theory of change** were seldom made explicit in the project document, which weakened the design. Some areas where assumptions needed to be made explicit include, for instance, those related to sustainability, regional programming; rights-based approaches; promotion of peace-building by working with gender equality;<sup>18</sup> and, changes in men’s and women’s lives through change processes that stem from capacity development. Both the MYFF and CEDAW-SEAP evaluations held that if underlying assumptions and theories of change were made more explicit, then the logical framework and monitoring of and reporting on results stood a better chance of being of good quality.

Given the above finding, it is not surprising that many of the evaluations found the **logic of the projects** was insufficiently consistent (e.g. MYFF, CEDAW-SEAP, Kosovo GE&WR, Kenya G&G, Caucasus-Peace II, CEE-Economic Justice, GRB-LAC). Several of these evaluations are highly critical and go into some detail about incoherence between outputs and outcome; multiple conceptualisations of the project structure; overly complicated logframes; and, insufficient, not SMART or irrelevant indicators.

**Monitoring and evaluation** was often found to be insufficient and the prompt elaboration of monitoring instrument was seen as necessary (e.g. India-VAW, 16 Day Campaign, HIV/AIDS, Burundi-Peace, Afghanistan-WL, Kenya G&G, DRC-Election, Caucasus-Peace II, GRB-LAC). The MYFF evaluation found that there was a perception among programme staff that funds for monitoring and evaluation could be “traded” for more programme activities. Another challenge was that partner organisations often had very limited knowledge of RBM and therefore found it “difficult or even impossible to report by results”. The CEDAW-SEAP evaluation maintained that staff were aware of the gaps in systematic tracking of longer-term progress.

Some of the evaluations found insufficient, poor or lacking **baseline data** (MYFF, Kenya G&G, Kosovo GE&WR, Afghanistan-WL, DRC-Election, GRB-LAC). This not only affected the quality of the project/programme design, but also, crucially, the ability to monitor change. Depending on the level of ambition, baseline data can be as costly to gather as many of the projects themselves. However, as suggested in the Kosovo evaluation, efforts could be made to gather modest data such as self-assessments by individuals and organisations targeted in capacity-building efforts:

UNIFEM may wish to choose selected parameters that are key for its work, such as ‘capacity development’, and establish user-oriented tools (e.g. rubrics or scorecard-type tools) to determine what types of capacities the project is trying to strengthen, what existing capacities of partners are at project onset, and what improved capacities are expected to look like.

---

<sup>18</sup> Peace-building by promoting gender equality and women’s rights is discussed to some length in Caucasus-Peace I evaluation.

Meanwhile, the CEDAW-SEAP evaluation established that national and regional baseline data were collected. Although these contained some deficiencies, the information was reported useful for assessing relevance and effectiveness.

**Reporting** on projects and programmes was raised by several evaluations as needing improvement in terms of quality (e.g. MYFF, Afghanistan-WL, South Caucasus-Peace II, GRB-LAC). Many reports tended not to make links between outputs and outcomes and many were oriented toward completion of activities rather than results:

UNIFEM is underselling itself by not reporting on achievements. (Afghanistan-WL)

On the other hand, the CEDAW-SEAP evaluation concluded that UNIFEM had compiled “concise, informative and reader friendly programme progress reports” that made “visible efforts to focus upon results rather than activities”.

Several evaluations saw an urgent need for UNIFEM to further enhance its use of **RBM as a flexible and meaningful management tool** within its projects and programmes. Indeed, the CEDAW-SEAP evaluation highlighted a number of areas where work had been adversely affected by RBM tools and related processes: it often hindered or placed additional burdens on the programme team rather than serving its management needs.

The results were mixed when it came to **strategic coherence** and **synergies** between components and projects. At the overall level within a country or region, some evaluations found that more could be done to link UNIFEM interventions (Local-Global, India-VAW, HIV/AIDS, DRC-Election). For instance, both the DRC-Election and India-VAW evaluations stated that opportunities had been lost by not strategically linking and integrating project components.

Meanwhile, CEDAW-Pacific found that UNIFEM was able to relate its other programmes in the region – Women in Politics; Women, Peace and Security; Elimination of Violence against Women and Women and HIV/AIDS – to CEDAW as an overarching framework for women’s rights. The Local-Global evaluation also mentioned how synergies were successfully established between the Global to Local project and projects such as Women’s Land Rights project in Kyrgyzstan and the Moroccan project to reform family law. However, the evaluation found this was an exception rather than a rule.

#### Coherence

UNIFEM has helped the Roma, Ashkali, and Egyptian (RAE) women’s network to strengthen their capacity to effectively advocate for their rights. The evaluation found that compared to the other outcomes, (‘Harmful and discriminatory attitudes and practices change to promote and protect the rights of women and girls’) appears to be a bit of a ‘mixed bag’ approach, i.e. a combination of stand-alone initiatives that have in common that they did not fully fit under one of the other outcomes. Also, there are a number of other project achievements that have not been captured under (this outcome), but could be as that they illustrate that and how UNIFEM works towards changing harmful and discriminatory attitudes and practices.

(From Kosovo GE&WR Evaluation)

## 7.2 Capacity Development

Generally, capacity development was recognised as UNIFEM's strength. It has been effective at the output level (e.g. CEDAW-Pacific, MYFF, CEDAW-SEAP, India-VAW, Kenya G&G, Afghanistan-WL, Kosovo GE&WR, Morocco Political Participation, Kyrgyzstan-Land, DRC-Election, Caucasus-Peace evaluations, GRB-LAC). In some cases, there was also some evidence of outcomes (e.g. CEDAW-Pacific, Local-Global, MYFF, CEDAW-SEAP, HIV/AIDS, Kosovo GE&WR, Kyrgyzstan-Land, Caucasus-Peace II). Lessons from the evaluations relating to capacity development include the following:

1. An explicit programme **capacity development strategy** is helpful in view of:
  - Outlining the programme/projects core understanding of 'capacity' and 'capacity development';
  - Specifying the programme's intended approach to addressing specific aspects of individual and institutional capacities and defining the limits of its intended approach;
  - Providing outlines of a framework for gathering information to further clarify what specific types of knowledge and skills are relevant for working with different partners and/or on different thematic areas;
  - Providing guidance on how to systematically track capacity-building results over time;
  - Basing programme strategies on a **clear and explicit conceptualisation of capacity development** – including the individual, organisational and institutional levels. This relates to the theory of change discussed above.<sup>19</sup>
2. Careful **targeting of suitable trainees** and target groups is essential. The CEDAW-Pacific evaluation held that this was an important aspect contributing to the success of the programme. The review of the HIV/AIDS capacity building effort in the Caribbean (2008) found that selected trainees were often not in positions to influence policy. Furthermore, this project and the Arab Women Parliamentarian project (reviewed in 2007) both overlooked the need to involve *men* in training efforts. Similarly, the Morocco Literacy evaluation saw the need to include husbands and other male relatives as part of the sensitisation work. The evaluators of GRB-LAC concluded that some of the programme's objectives may have been easier to achieve if government entities and officials had been included in the capacity-building effort. Meanwhile, the CEDAW-Pacific evaluation identified the training of less traditional groups (e.g. Muslim women counsellors, women with disabilities, a nurses association and teachers' association, male teacher unionists) as examples of good practice.
3. **Undertaking robust needs assessments** is crucial. This was particularly raised by the evaluation of Kenya G&G. It was also brought up in other evaluations in the context of M&E and the need for baselines. As discussed above, ex-ante self-evaluations of trainees could provide some of this data.
4. **Trainers need to have relevant knowledge** and skills. In the Morocco literacy project, the evaluators believe that the lack of change in the trained women's gender perspectives and knowledge of women's rights was a likely result of the insufficient capacity (and/or conviction) of the trainers to broach these subjects. The CEDAW-SEAP evaluation found that some national trainers did not have sufficient capacity to monitor/track the mid- to long-term effects of training.

---

<sup>19</sup> Taken from CEDAW-SEAP evaluation page 57.



The CEDAW-Pacific evaluation emphasised the importance of teaching actual training skills in training of trainers workshops (see box below).

#### **Training of trainers**

TOT workshops are often not as successful as hoped. While participants might increase their own knowledge, they are not always in a position and/or have the training/facilitation skills to pass on this knowledge. Training and practice in how to transfer information is rarely included in TOT workshops. Most participants in this workshop were already working in the area of human/gender rights as legal rights training officers, community paralegals or women's government or NGO officers. There was therefore considerable existing knowledge and training experience on which to build. Participants were in many cases in a position to use the new knowledge immediately. This was a positive outcome and a useful lesson for UNIFEM and other organisations considering TOT workshops.

It would also be helpful to provide some level of on-going support and motivation for these key CEDAW trainers, perhaps through a specific electronic newsletter (or facilitating their participation in the listserve). This would enable access to new information and sharing of experiences ... linking the trainers into providing support for country sectoral plans on gender mainstreaming/CEDAW implementation, where they have the appropriate background and skills, would also capitalise on these valuable resource people.

(from CEDAW-PACIFIC Evaluation, p 35.)

5. **Longer-term effects of the capacity development** of different partners need to be **systematically tracked**. The evaluation reports reveal that UNIFEM's monitoring effort of training activities is often weak (see section 5.1 above). There seems to be insufficient follow-up after workshops. To adequately determine the effects of capacity-building, regular ex-post surveying of participants is an informative approach, but with the exception of the Moroccan Literacy project, there is scant evidence from the reports that this is undertaken.<sup>20</sup> The CEDAW-SEAP evaluation explained that this has "somewhat limited UNIFEM's ability to demonstrate concrete results at individual or organisational levels emerging from its interventions, which is not merely an issue in view of programme accountability, but also in relation to UNIFEM's and its partners' abilities to learn from implementation experiences, and adjust programming approaches accordingly."

### **7.3 Partnerships**

UNIFEM's partners can be broadly categorised into governments, regional organisations, civil society organisations and UN partners. Several of the evaluations discuss the nature of UNIFEM's partnerships. Many pointed how valued UNIFEM is among its partners (CEDAW-SEAP, India-VAW, HIV/AIDS, Afghanistan-WL, Burundi-Peace, Kosovo GE&WR, SEE-WR, Kenya G&G, Caucasus-Peace II). In most cases UNIFEM is described as dedicated, knowledgeable, exceptionally approachable and '**down to earth**' compared to other international organisations:

UNIFEM is special – they never make us come and beg for support, but they actively reach out to us and offer support where it is needed. (Ecuadorian NGO, MYFF)

---

<sup>20</sup> One positive example from the CEDAW-SEAP evaluation is that UNIFEM in Thailand collaborated with the NGO Way Lampang to track changes among former elected female trainees.

UNIFEM staff members were seen as being readily available to provide advice or hands on support:

Rather than a donor, UNIFEM is perceived as a partner and ally who is genuinely interested and passionate about the issues at stake, and who treats those it works with as partners rather than as mere grant recipients. (Kosovo GE&WR)

The importance of UNIFEM is much more than it's funding (Partner, Caucasus-Peace II)

Several reports highlighted that UNIFEM did **not impose** specific approaches or ideas, but was respectful of, responsive to and actually listened to the needs and suggestions of its local partners (CEDAW-SEAP, India-VAW, HIV/AIDS, Afghanistan-WL, Kosovo GE&WR, SEE-WR, Kenya G&G, Caucasus-Peace II). The stakeholders of the project SEE-WR ascribed the project's success to UNIFEM's participatory approach towards its partners and flexibility in the face of rapidly changing situations. The CEDAW-SEAP evaluation, however, found that there was also a down-side to UNIFEM's "soft approach" which was that it sometimes hindered the delivery of timely results.

UNIFEM status as a UN agency and its adeptness in acting as a **neutral convener** is praised in several evaluations (16 Day Campaign, Kosovo GE&WR, Kenya G&G, SEE-WR, Caucasus-Peace II). The Kosovo GE&WR evaluation believed that UNIFEM's soft approach strengthened its credibility and high reputation in eyes of local partners, which thereby allowed it to retain a position of neutrality and position to mediate between different stakeholders. In the highly politicised and ethicised societies of Kosovo and Kenya, UNIFEM's neutral way of conducting itself was particularly appreciated.

The reports from 2008, but also from 2004 to 2007, often mention the unique role UNIFEM plays and the added value it brings by **linking civil society actors with government**. UNIFEM's status as a UN agency provides NGO/CSO partners with the needed legitimisation for their work, and helps individuals and groups in gaining access to high level forums that they would not otherwise be able to access.

"UNIFEM grants are important to us, but most important is its support in networking. Under UNIFEM's umbrella we have been able to successfully network with women from the region, and jointly influenced international decision- makers." NGO partner in SEE

The GRB-LAC evaluation reported that by receiving support from UNIFEM, the local departments of gender affairs were given a more central role and increased standing in the eyes of the local governments. The CEDAW-SEAP evaluation held that UNIFEM's facilitation role and support helped to improve the frequency and quality of interactions between government agencies and NGOs, in particular in relation to the CEDAW national report preparation process and review. In Nigeria, UNIFEM facilitated contacts between the women in the project area and other potential partners such as micro-credit institutions, commercial banks, UNDP and the International Institute for Tropical Agriculture.

The CEDAW-SEAP, India-VAW and HIV/AIDS evaluations found that UNIFEM's **choices of partnerships** generally have been effective and appropriate. The Caucasus-Peace I evaluation praised UNIFEM for working with women who remain rooted in their communities, as opposed to travelling to international conferences and thereby avoided the creation of an elite international group. However, in some cases, evaluations found that individual partnerships appear to have more

inherent potential than had been brought to bear to date. For instance, the Kenya G&G evaluation concluded that UNIFEM had high legitimacy with the government of Kenya but it did not develop this partnership within programme. It recommended that UNIFEM use its strategic position to greater effect.

Other **missed opportunity** regarding partnerships was raised in some evaluations. The SEE-WR evaluation reported that while UNIFEM had been skilful in partnering with and networking among gender equality organisations, there was insufficient effort to forge links with other agencies relevant within the project's context, namely those responsible for privatisation or international organisations that are active in promoting economic reform and European integration. The evaluator further conveyed that while UNIFEM had a close relationship with the Ministry for Women's Affairs, it failed to establish partnerships with other line ministries. Similarly, the review of the Arab Women's Parliamentarians Project (2007) highlighted that the project had not to date partnered with relevant regional organisations or brought male parliamentarians on board. Furthermore, the GRB-LAC evaluation reported that the project omitted government agencies in the capacity building effort, even though indicators were set for measuring changes in public sector plans, processes, programmes and budgets.<sup>21</sup>

The Kosovo GE&WR, SEE-WR, GRB-LAC and CEDAW-SEAP evaluations raised the importance of UNIFEM **working with governments**:

“By supporting and engaging with government institutions, UNIFEM publicly models the very core of a rights-based approach, namely the normative expectation that women's human rights are to be addressed not only by civil society, but also by government. Its function of connecting right holders and duty bearers makes UNIFEM's engagement with government institutions highly valuable, even if short-term successes within individual government institutions may be limited.” (Kosovo GE&WR)

A few evaluations commented on UNIFEM's relationship with **other UN agencies**. The MYFF evaluation found that there was evidence of UNIFEM's “systematic and growing engagement in UNCTs and related/relevant working groups and processes”. It mentioned successes in shaping policies, plans, and projects of UNCTs in relation to gender equality issues, and that UNIFEM carried out “highly relevant work analysing the efforts of UNCTs to close existing gaps in gender equality”. However, it maintained that UNIFEM suffered from limited authority and status within the UN, meagre core funds and minimal access to high-level decision-making within the UN.

The other evaluations that mention UN partners all concerned projects/programmes in volatile countries and revealed mixed results. For instance, UNIFEM played an important leadership role in relation to gender issues within the UN Kosovo Team's (UNKT) work. Likewise, Burundi-Peace mentioned that a half-dozen UN/multilateral organisations had benefited from UNIFEM's advice, technical support and “timely nuanced analysis” on matters relating to the Burundi peace process and its gender dimensions. It also mentioned administrative problems between UNIFEM and UNDP, but provided few details.

The DRC-Election evaluation found that relations between office personnel from UNIFEM, MONUC and UNDP had deteriorated to the detriment of the UNIFEM project. In fact, UNIFEM's

---

<sup>21</sup> Interviews revealed that UNIFEM had worked with government actors but had not reported on it, leaving no evidence for the evaluators to analyse.

project was supposed to have been guided by a consultative committee with UN partners but the mechanism was never realised. The Afghanistan-WL evaluation detected tension caused by a misunderstanding about UNIFEM's capacity to assist gender mainstreaming within the UN, because partners did not “distinguish between UNIFEM being a ‘motor’ for gender mainstreaming and it being ‘responsible’ for gender mainstreaming”.

Caucasus-Peace II noted that UNIFEM was actively engaged and co-ordinated with other UN partners – UNDP, UNFPA and UNICEF – in Georgia and Armenia in joint programmes/projects, a survey and assessments. UNIFEM was a member of the UNCT in both of these countries. Meanwhile, in Azerbaijan UNIFEM's interaction with other UN agencies was much more limited, possibly partly because it was not invited to be part of the UNCT. The evaluator deemed that the extent of UNIFEM's co-operation with the rest of the UN system depended mainly on “whether these agencies were interested in cooperation or not”.

In three evaluations, **communication with partners** was brought up. Clarifying UNIFEM's aims and strategies to avoid a mismatch of expectations seemed to be the underlying issue. The 16 Days Campaign evaluation saw the need for better communication with partners to clarify UNIFEM's role in the campaign and increase burden-sharing. In the DRC-Election project, stakeholders – including women's organisations, the government, bilateral and multilateral partners – criticised UNIFEM for not having a clear and effective communication strategy to make the project, its objectives, resources and operation known. Afghanistan-WL noted that relations had deteriorated between UNIFEM and some of its Afghan partners. Similarly, as mentioned above, relations with its UN partners were not optimal. The evaluation established that communication failures (lack of communication and well as manner and mode) were the culprit.

## 7.4 Regional Approaches

A regional approach to programming can add value to development processes in a number of ways, for instance by: i) coherently addressing issues that transgress national borders – such as human trafficking; ii) broaching subjects that may be too sensitive to approach at the national level; and, iii) enhancing capacity-building efforts by accessing regional resources and bringing together actors from different countries to exchange experience and lessons.

Around 10 of the evaluations covered regional projects/programmes. Sadly, less than half reflect on the regional aspect of the programme in terms of strengths, weaknesses and opportunities and constraint that this may bring. In some cases, this is because the terms of reference did not raise the issue, other times the evaluators failed to address the issue.<sup>22</sup>

The CEDAW-SEAP evaluation, nevertheless, provides significant insight into the regional dimension. It established that “with regards to the (potential) value added by the programme's regional approach, there would have been considerably more room for both UNIFEM and CIDA to make their respective underlying assumptions more explicit”. For a start, UNIFEM and its donor, CIDA, discovered after nearly four years of implementation that their understanding of the term “regional” differed, where the latter saw a ‘regional programme’ as one that aimed for distinct regional results that are different in quality than the sum of individual results achieved at country

---

<sup>22</sup> The ToR of the GRB-LAC evaluation did specifically stipulate that the regional dimension be analysed. However, the report provides only limited reflection on this area.

level. This discovery implied that only a limited amount of discussion and reflection took place with regard to one of the core concepts that underlay CEDAW- SEAP's overall design. The evaluation provided the following findings and reflection:

- “While both partners appear to have acted on the assumption that working at a regional (or multi country) level ‘*naturally*’ provided specific benefits in view of the scope and/or quality of development results, these benefits have not, or only to limited extent, been made explicit.”
- “One dimension that does not appear to have been explored (explicitly) by either UNIFEM or CIDA is the seemingly trivial question of what constitutes a ‘region’.”
- Since legal commitments to CEDAW exist exclusively at the *national* level, implementing CEDAW means ‘translating’ the international Convention into concrete action at the national level. The importance of the regional dimension lay in its *strategic potential* to enhance the extent, quality, or speed of CEDAW implementation at the national level, i.e. through the use of regional resources for national level capacity development, or by creating political pressure among countries in the region.

The evaluators recommended UNIFEM to focus on the same priorities in all the countries in the future to make the best use of the programme's regional nature – both in terms of creating synergies for programming, and in view of UNIFEM's ability to systematically capture and analyse experiences.

The SEE-WR evaluation was specifically requested to examine whether the regional dimension was adequately taken into account and the advantages of regional exchange were made the most of. It found that central to project design of SEE-WR was the concept that in view of shared history and challenges, a regional approach to promoting gender equality was appropriate. However, in light of lessons learnt from previous work in the region, considerable flexibility was built into project design, so that the project could tailor its activities in each of the four countries – Serbia, Kosovo, Bosnia-Herzegovina, and Montenegro. The evaluators deemed that the regional dimension was effectively integrated and lay at the heart of the project design and implementation process. Further, the evaluation found that stakeholders placed premier value on learning from regional and international experience, whether through the sharing of draft reports or regional workshops.

The Caucasus-Peace II evaluation highlights the importance of the regional dimension. The evaluator established that the Women for Conflict Prevention and Peace-Building Project was unique in its focus on fostering reconciliation and peace-building through building national women's networks and building alliances between the women's movements in the region. Furthermore, it was one of very few projects in the South Caucasus with a regional scope and it is one of the few regional attempts to link civil society with the peace process. The report also provides evidence on how this regional approach has resulted in propensities for peace.

SA-Trafficking reported that there had been “limited learning from the regional character of the project, in terms of deeper understanding of the subject, which could promote the formulation policies and mechanism to provide support services, awareness-raising and legal reforms” to end trafficking.

The SEELINE evaluation, on the other hand, suggested that the regional approach to a legal advocacy network was ill-advised because the countries involved were not sufficiently homogenous

legally and politically to benefit from common reform initiatives. What was appropriate in one economic or political situation might be irrelevant or even counter-productive in another. However, the project's methodology required consensus recommendations. The evaluation furthermore deemed that the number of countries involved in the project was not a manageable size. The evaluation concluded that a regional approach would be more appropriate for proposing reform to a multinational legislative body where compromise across borders is required, but was not optimal as the basis for national reform across a diverse region.

## 7.5 Human Rights and Gender Equality Perspectives

Few evaluations directly commented on **gender equality perspectives** in the programmes. However, in cases where it was reported that an overall strategic perspective was missing from the design (see section 7.1), it was sometimes underlying assumptions and theories of change relating to gender equality that were lacking. In most cases, evaluations did not examine whether or not gender analyses had been undertaken or used as part of the project preparation process. In a couple of instances, however, evaluators did. For instance, the SA-Trafficking evaluation, found the analytical foundation of the project to be weak. It pointed out that the close link between reduction of gender inequality and poverty had not been analysed and nor had issues of women's empowerment, sexuality, right to and control of resources, or social consequences of migration been adequately examined or addressed within the project's conceptual approach. Likewise, the Caucasus I evaluation concluded that project activities were "not based on gender or human rights analysis".

None of the projects/programmes evaluated appear to consist of a **full-fledged rights-based approach** in which the principles of equality and non-discrimination; participation and inclusion; and rule of law and accountability are systematically applied in analysis, design and implementation. By virtue of its highly participatory approach (and thus making process an outcome in itself) combined with a two-pronged approach of building capacities of duty-bearers and rights holders, Kyrgyzstan Land may be, in practice, closer towards achieving a rights based approach. Nevertheless, several projects/programmes applied relatively solid **human rights perspectives**.<sup>23</sup> With a couple of exceptions, these projects/programmes constituted those that worked expressly with promoting women's rights (and CEDAW in particular) and/or achieving gender equality in democratic governance. Examples include Morocco Literacy, Kyrgyzstan Land, Morocco Political Participation, Kenya G&G, Global to Local; Kosovo GE&WR, SEE-WR and CEDAW-SEAP.

While not falling into the above mentioned category, it appears that the UNIFEM's *Programme on Enhancing Human Security through Gender Equality in the Context of HIV/AIDS* also maintained a relatively strong human rights perspective throughout. The evaluation reported that UNIFEM was successful in grounding HIV/AIDS programming in a gender *and* human rights framework in a number of contexts and countries. It provides several examples of how this has been done. Furthermore, it concludes that UNIFEM's initiatives to empower (HIV) Positive Women's networks around the globe strengthened these networks in their ability to organise and to become more vocal advocates for their rights in national and global forums. In India, for instance, this contributed to the drafting of the HIV/AIDS legislation, and the formulation of the Domestic Violence Bill.

---

<sup>23</sup> For a discussion on rights-based approaches versus rights based perspectives, please see Cecilia Ljungman "A Rights-Based Approach to Development", in Mikkelsen *Methods for Development Work and Research A New Guide for Practitioners*, Second Edition, Sage Publications 2005.

None of the evaluations assessed the quality of UNIFEM's efforts to apply a rights-based approach. Nevertheless, in some cases there were comments on the approach. For instance, the CEDAW-SEAP evaluation stated that although UNIFEM has already undertaken "considerable work in defining its corporate understanding of a HRBA, it may be helpful to elaborate on and specify this general understanding in more detail in the context of a specific programme such as CEDAW-SEAP." Meanwhile, the SEE-WR evaluation reported that the "the project oriented itself strongly on CEDAW, which reflects a rights-based approach to gender equality" but that its approach to women's economic rights was less impressive and there was a need to "better exploit synergies between the human rights and economics perspectives".

While the Caucasus I evaluation (2004) did not assess UNIFEM's rights-based approach (because one was not applied), it is the only evaluation that offered a more in-depth discussion of the topic. It argued that "while (UNIFEM) is working towards developing the theory of being a women's human rights organisation into practice, there remains a lack of in-depth understanding of what it means either to practice rights-based programming or practice programming for women's human rights". It examined the sub-region in the global human rights context, assessed the human rights capacity of the staff in the sub-region and the considered the opportunities and relevance of applying a rights based approach. These are included in the box below.

#### **Peace, Security and Women Human Rights in Southern Caucasus**

"The Capacity of local partners (in southern Caucasus) to embrace a human rights approach is limited; many groups, instead, prefer to talk in terms of women being 'naturally' peaceful and taking on this role because they are mothers, rather than rights-holders. Addressing the Project more in terms of gender and of women's human rights should make a difference to this situation, with appropriate technical support and advice.

It is suggested, therefore, that the Project adopt a more imaginative approach; this is to say that, by working on women and peace-building, the Project is promoting women's human rights through a gender perspective AND is promoting the protection of women's human rights as determined by women themselves. All governments and many non-state actors (particularly those who have aspirations of government themselves) wish to be seen to be good on human rights and so it is an entry point for the Project to invite interlocutors to engage with peace-building in order to further women's human rights. This will involve a strategy of discussion and advocacy with a range of actors at decision-making levels within each country, including the business community, the judiciary and senior civil service, as well as parliamentarians.

Where it is, genuinely, too controversial to even use the phrase 'human rights', the concept of 'human security' creates an opportunity to talk about the same issues, but in a slightly less sensitive way. Further, non-state actors particularly may be more receptive to this concept as it gets away from debates about responsibility (i.e. whether non-state actors have any responsibility for human rights compliance) and encourages them to think about security from a non-nation state perspective. This approach should not be used, however, as a 'get out' from talking about human rights on the basis that the government or NSE *may* be reluctant to engage."

## **7.6 Institutional Learning & Knowledge-Sharing**

Many evaluations maintain that the achievements and other aspects of the projects and programmes are highly worth showcasing within UNIFEM and beyond because of their potential for learning and replication (Local-Global, CEDAW-SEAP, India-VAW, HIV/AIDS, Kosovo GE&WR, Burundi-Peace, Kenya DG, Kenya G&G, SEE-WR, DRC-Election, Nigeria Empowerment, Morocco

Literacy, GRB-LAC, IGAD). These include several instances in which UNIFEM acted as catalyst so that modest investments led to broader changes. However, while according to the MYFF evaluation, some efforts have been made to improve documentation and dissemination of success stories,<sup>24</sup> as stated in the CEDAW-SEAP evaluation:

“The vast potential for learning inherent in CEDAW-SEAP has only been partly tapped into to date.”

Evaluators did not find a lack of willingness to share experience within UNIFEM or among its partners, for instance:

“Partners and implementers in Burundi expressed the wish to have more communication with UNIFEM on the outcome of activities and on the sharing of results and lessons learnt. They felt that engendering of peace processes and conflict resolution is still new and there is a great need for sharing and learning.”  
(Burundi-Peace)

Rather, there was limited time made available for reflection, analysis and report writing. An effective management strategy to systematise institutional learning was also found absent. The Kenya-GE&WR evaluation maintained that among UNIFEM’s partners, useful information was lost because women were overwhelmed and intimidated by the need to convert from an oral to a writing culture. It recommended that audio and video technology be used to help women groups document ‘lessons learnt’ and best practices.

#### **Box 5.1: Institutional Learning**

CEDAW-SEAP in Timor Leste has worked with a local NGO to develop and apply dance and drama performances aiming to bring concepts of women’s human rights and of discrimination against women to a broader and largely illiterate public. While programme reports mention the fact that these events have taken place and that they were successful, they do not provide information on any related experiences or lessons, or on the particular effects that the performances have had on their audience, or how these effects were assessed.

Opportunities for structured exchange, joint discussion of strategic and conceptual questions, and in-depth analysis of common lessons learnt have reportedly also been limited in many cases. This was particularly pointed out in the regional projects and programmes (CEDAWSEAP, HIV/AIDS, SA-Trafficking). Some national co-ordinators have exchanged information one-on-one (CEDAW-SEAP) but others, like in the GRB-LAC project, tended to communicate more with the sub-regional offices than with the other national co-ordinators.

The Local-Global evaluation found that the lack of a structured system of sharing project information within the organisation hampered the ability of the regional offices to build on and integrate the Local to Global project into their programmes.

---

<sup>24</sup> The Nigeria Empowerment evaluation mentions that through its performance management system, UNIFEM “recorded several project achievements and many lesson learnt”. However, these documents are not referenced in the report’s bibliography.



Meanwhile, based on its visit to the Andean sub-region, the MYFF evaluators found that while knowledge-sharing worked rather well within the region, between regions this was difficult because of language barriers, often different contextual challenges and lack of funds to bring people together.

**Box 5.2: Suggestions to Improve Reporting in View of Showcasing Success**

Project staff may want to discuss with the sub-regional office and/or HQ what types of support, tools, or enabling conditions (such as more time allocated for reporting) the team would consider to be helpful for further improving its reporting practice, and which of these are feasible.

The development of a corporate results tracking system – as is currently being prepared by UNIFEM HQ – may be able to help mitigate some of the current reporting challenges e.g. by allowing UNIFEM staff to assign different *codes* to project achievements in order to mark them as ‘a case of something’. Thus, a particular achievement could be coded as an example of progress under outcome 1, but also as an example of an innovative approach. This kind of simplified reporting might also relieve staff members of some of the workload currently created by the multiple obligations of reporting internally and to external donors.

(From the Kosovo GE&WR Evaluation)

## 7.7 Human Resources & Financial Management

A number of evaluations mentioned the importance of having a dedicated programme manager and a **national co-ordinator in the countr(ies)** where UNIFEM’s projects and programmes were implemented. The use of national staff was particularly praised in the CEDAW-SEAP, Caucasus-Peace I and Kosovo GE&WR evaluations.

Meanwhile, as stated in the MYFF and HIV/AIDS evaluations, in countries where UNIFEM does **not have ongoing country presence**, it has been reportedly much harder to be informed about and involved in ongoing interagency work, and to identify opportunities for promoting gender issues within the UNCT. Furthermore, the HIV/AIDS evaluation held that UNIFEM’s lack of country presence limited its contribution to mainstreaming gender equality in the effort to address HIV/AIDS. Lack of country presence can also impact on project directly. The Tanzanian components of the Burundi-Peace project were allegedly negatively affected by UNIFEM having to rely on sister agencies that did not always deliver as well as was expected.

Staff members in most countries were considered very committed and supportive of partners (e.g. Kenya G&G, Kosovo GE&WR, Burundi-Peace). Offices in several countries were praised for their knowledge and experience (Burundi-Peace, Afghanistan-WL, Kosovo GE&WR, Kenya GE&WR).

UNIFEM staff through their professional and personal commitments have ensured that the project achieve maximum results. (Burundi-Peace)

In 2004 the Caucasus-Peace I evaluation concluded, nevertheless, that there was a need to enhance the capacity of the project staff, particularly regarding knowledge related to peace-building and human rights.<sup>25</sup>

A few of the evaluation reports point to staff being over-worked and/or working under great pressure (e.g. CEDAW-SEAP, 16 Days Campaign, Afghanistan-WL, Kosovo GE&WR, Burundi-Peace). In

---

<sup>25</sup> Caucasus-Peace II did not, however, raise this issue, perhaps because the issue had been addressed?

Kosovo, partners were impressed by the ability of the minimalist UNIFEM office to produce results and regarded the staff as adding considerable value to the small resources available. It was also found that the office had difficulties in saying “no” to requests, which contributed to the mismatch between ambition and resources. In the Andean region, the 16 Days Campaign found that UNIFEM staff were overworked partly because partners perceived the campaign as being a UN effort led by UNIFEM and they therefore did not shoulder the planning and execution of the campaign as would have been desired.

The evaluations reported that **staff turnover and vacancies** adversely affected a number of projects and programmes (Local-Global, CEDAW-SEAP, 16 Days Campaign, Burundi-Peace, Kosovo GE&WR, SEE-WR). Furthermore, lack of clarity and shared understanding with regards to roles and responsibility within UNIFEM (country, regional and HQ levels) was raised (CEDAW-SEAP, HIV/AIDS, SEE-WR, Kosovo GE&WR). On the other hand, the Burundi-Peace evaluation found that there was “wonderful and consistent collaboration between the different UNIFEM offices in Nairobi, Bujumbura, Kigali and New York”.

**Financial management** was a concern in some evaluations. Delays in disbursements were experienced in the Kenya G&G, Kenya GE&WR, DRC-Peace and SEE-WR programmes. Financial management and monitoring was uneven or slow in the projects evaluated in Kenya, Burundi, Kosovo and Southern Caucasus. The evaluator of CEDAW-Pacific found that there was a lack of clarity among both programme and financial staff on funds received and those still available for programme activities, making programme decision-making and management at the field level difficult. Lack of access to **Atlas** by the project office was singled out as an impediment in Kosovo, where the office was challenged to track project expenses. Meanwhile, the Caucasus-Peace II evaluation found that the introduction of Atlas had made financial project management less efficient. Because Atlas is designed for large operations and requires high Internet connectivity, the evaluators held that it led to increased workloads for project staff.

## 8 Evaluation Criteria

In the sections that follow, conclusions presented in the evaluations reports that relate to the five evaluation criteria – relevance, effectiveness, sustainability, efficiency and impact – are summarised and discussed.

### 8.1 Relevance

Relevance is the extent to which a development intervention conforms to the priorities and policies of the target group, recipient and donor. A little under two-thirds of the evaluation assessed relevance and deemed UNIFEM’s efforts as being mostly relevant. The evaluators, grouped together by thematic area, summarised relevancy as follows:

#### 8.1.1 Cross-cutting Evaluations

**MYFF:** “Consulted UNIFEM stakeholders widely acknowledged that UNIFEM’s products and services are relevant as they pertain to existing and articulated priorities of UNIFEM partners at local, national, regional, or global levels, and that UNIFEM is highly responsive to upcoming opportunities and needs.”

**CEDAW-SEAP:** “CEDAW-SEAP’s relevance at the *regional* level derives from the fact that, at programme onset, there were no other mechanisms or facilitators able to bring together diverse players over the common theme of CEDAW implementation, nor any regional body to coordinate the collection and sharing of relevant knowledge and experiences gained within the region. CEDAW-SEAP has also been highly relevant in view of UNIFEM’s and CIDA’s respective mandates and corporate strategic priorities.”

#### 8.1.2 EAW Evaluations

**16 Days Campaign:** The evaluation concluded that the campaign was generally relevant to the needs of counterpart organisations; the situation of each country with regard to violence against women; and, the information needs of the general public. It noted that when the Campaign adjusted its strategy to move beyond the women’s movement and include organisations dealing with human-rights, indigenous peoples, Afro-Hispanic groups, youth and men, the relevance of the Campaign improved markedly.

#### 8.1.3 HIV/AIDS Evaluation

**HIV/AIDS:** “The evidence suggests that UNIFEM’s programming is relevant to stakeholders in terms of its timing, content, and approach”. **However**, the evaluation also concluded that “in some cases the Fund’s level of resources or lack of country presence is seen to limit its contribution.”

#### 8.1.4 Governance Evaluations

**Burundi-Peace:** The evaluation indicated that the types of activities funded by the project were very relevant. In terms of the national context and evolving peace situation in Burundi, the project was regarded as playing a unique and important role in ensuring women’s effective participation in the peace process. **However**, the evaluator had reservations “about the activity approach used by UNIFEM in the context of crisis where more sustained uninterrupted support would have been more appropriate.” It further explained that while UNIFEM’s role as a catalyst may be suitable in politically stable situations, it doubted its appropriateness in a crisis-affected country, where a more

sustained approach would have better addressed the more pressing needs of a conflict-ridden country and yielded more results.

**The Kenya GE&WR** evaluation concluded that the programme was “absolutely relevant, timely and firmly grounded within Kenya’s political, social and economic context.” **However**, the evaluators suggested the need to explore how HIV/AIDS could be integrated as a cross-cutting concern to enhance relevance.

**Afghanistan-WL:** The evaluation established that the choice of the three sectors with which UNIFEM was engaged – governance, justice and livelihoods – was appropriate to the needs of women in contemporary Afghanistan.

**Caucasus-Peace II:** “The relevancy of the project has been high, both for the women in the region as for the governments who have committed themselves to international conventions and resolutions on gender equality and women’s role in peace processes. Gender inequalities, such as women’s low participation in decision-making, exist in all the three countries, including Abkhazia, and a greater involvement of women in policy making and peace processes aiming at stability in the region is in the interest of all.”

**DRC-Peace:** The evaluation established that the project responded to real and important needs in terms of capacity development and priorities of Congolese women in the electoral process. **However**, the evaluator implied that it would have been more relevant if the project had been guided by a consultative committee, as was initially foreseen. Instead, the project “opted many times to operate hermetically and took on the role of an executing agent instead of one of facilitator”.

**Kosovo GE&WR:** “UNIFEM’s work has remained highly relevant in the evolving Kosovar and regional context. The project has addressed needs and gaps at national and regional levels, including: i) The continued need for capacity development support for government agencies and CSOs working on gender issues in Kosovo, and for facilitation of cooperation between government and CSOs; ii) The need for strong advocacy for gender equality and women’s human rights at all levels. This need was particularly high as for the past few years political and public debates have been occupied with issues of Kosovo’s final status and the economy; and, iii) the continued need for facilitation of regional exchange through neutral (international) parties.”

**Kenya G&G:** “The idea of a gender and governance programme is very relevant to Kenya today.” **However**, the report implies that some relevance was lost because the programme was poorly conceptualised and designed. Furthermore, because of “the fact that no baseline surveys and or needs assessment were undertaken, there is no guarantee that (the) interventions are consistent with beneficiary priorities or demands.” The evaluators pointed to examples of how the programme design did not adequately take on board the differences in needs and concerns in different communities.

### 8.1.5 Poverty Evaluations

**GRB-LAC:** The project aims were “a response to women’s interests in the region”. Its aim “to enrich the sense of local democracy... is coherent with stakeholders’ agendas”. **However**, the report stated that the project lost some relevancy by not systematically including government actors in the capacity building efforts.

The scope of the **Nigeria Empowerment** evaluation did not include relevance as an evaluation criterion. However, the report did provide findings that suggest that most of the approaches used by the project were relevant – this included involving local structures in the implementation of the project. **However**, one exception was the ICT component that included the provision of ICT equipment and a CD Rom tool. While the computers were mostly put to good use (by the primary school, local council members, children) the women who they were meant for did not use the equipment. Likewise, the CD Rom tool, while of excellent quality, the evaluators found it was far too sophisticated for the use of women farmers and petty traders.

## 8.2 Effectiveness

Effectiveness – the extent to which planned objectives have been achieved – is particularly difficult to attain when the goals are overly ambitious in relation to the resources available. The transformational results that UNIFEM and its partners aim for typically take much more than a couple of years to achieve.

### Box 5.3: Management for Results – Limits Learning from Mistakes:

Accepting capacity development as a long-term process that – besides technical knowledge and skills – also requires local ownership, also means to acknowledge that individuals and organisations have to make mistakes and learn from them, and that they have to progress at a pace that is agreeable and feasible for them. This may mean that processes take longer than they would if led by an external (international) consultant, or that the initial quality of products and processes may not meet ideal external standards.

Working under the format of an externally funded results-oriented programme, however, puts pressure on UNIFEM to demonstrate visible results within a defined, relatively short timeframe. In this context, setbacks or mistakes are usually not regarded as suitable indicators for progress (and are thus not, or only to limited extent, reported upon), although in the long run they may show to have been key factors leading to changes in partners' capacities.

(From CEDAW-SEAP evaluation)

UNIFEM's evaluations from this period report that at an overall level, effectiveness has ranged from fair to high. Outputs goals were generally fulfilled and a majority of interventions achieved significant outcomes as well. These are discussed below, organised by thematic area.

### 8.2.1 Cross-cutting Evaluations

The **Local to Global** project was considered successful in achieving most of the expected outputs. Furthermore, over the years the project contributed to formalising the interaction between the CEDAW Committee and NGOs, creating a more legitimate space for NGOs to play a role in the monitoring of state obligations. “NGOs have been able to prepare better quality shadow reports and summary information based on the needs of the Committee that they learnt from their participation in the project. CEDAW Committee members were able to ask more precise questions to governments and provide sharper Concluding Comments as a result of having greater access to alternative country information and presence of national NGOs at the Committee. The project in some cases has contributed to increased government transparency and government-NGO interaction. Examples include Korea, India, Georgia, Azerbaijan and Nepal.”

The **CEDAW-Pacific** evaluation showed that UNIFEM made a positive contribution by enhancing government and NGO capacity, improving collaboration between government and NGOs and

improving information sharing and networking. UNIFEM was widely recognised as playing a lead role in relation to CEDAW implementation and, in some cases, was the only donor/agency providing support in this area. **However**, in some programme areas progress was slow – constrained in particular by lack of government commitment and limited capacity and resources of women’s departments to fulfil the role of lead agency for CEDAW.

The **MDG** evaluation report concluded that all five pilot projects achieved concrete and complementary results, providing instructive cases of engendering MDG advocacy and reporting. Several lessons and challenges emerged from the pilot projects, providing useful pointers for mainstreaming gender equality in MDG processes on the ground.

The **MYFF** evaluators stated that “There is widespread agreement among consulted partners that UNIFEM is an effective advocate for gender issues and women’s human rights, and that it has been able to achieve remarkable results with very limited resources. UNIFEM is perceived to have worked in accordance with its mandate. During the field visits, nearly all consulted partners stated that they tended to look to UNIFEM if they needed help or expert advice on gender-related topics.” The evaluators did, **however**, encounter criticisms of UNIFEM’s developmental effectiveness related to its level of funding: a) its resources did not currently allow it to be fully present and engaged in all countries; b) that its range of activities and influence was often limited due to a lack of funds; and, c) it was running the danger of “spreading itself too thin” in trying to meet the large existing demands with its limited resources.

The **CEDAW-SEAP** evaluation assessed that the programme had made considerable achievements – even though the full implementation of CEDAW across all sectors and levels of society is a long-term process. The programme had significantly contributed to making CEDAW better known among large and diverse groups of stakeholders in all seven participating countries and among selected regional organisations. CEDAW-SEAP also contributed to strengthening various aspects of national partners’ capacities that were relevant in view of more effective and comprehensive CEDAW implementation at the national level. Furthermore, CEDAW-SEAP’s work had *catalytic* results by enabling processes that were otherwise unlikely to have happened (e.g. initiating and facilitating exchange among governments as well as among NGO partners at national and regional levels) and by enhancing the speed and/or quality of processes (e.g. by assisting governments and NGO partners to use evidence-based and participatory approaches in the process of preparing their respective CEDAW reports).

### **8.2.2 EAW Evaluations**

The evaluator of **India-VAW** maintained that the output considered of greatest importance for VAW-India was the production of 17 resource directories covering 452 districts. The process of compiling the information was regarded as a result in itself, since it was designed to trigger change. The project’s initiatives on men and masculinities – although it was geared towards academic communities – was seen as ground-breaking for India and it spawned research that received international recognition. **However**, regarding the legal literacy component the evaluator concluded that more intensive rather than extensive coverage of issues and locations would have better contributed to outcomes.

In its assessment of effectiveness, the evaluation of the **16 Day Campaign** took into consideration the overall development goal of the campaign – to eliminate harmful practices and discrimination of

women – which of course will take more than a few generations to achieve. Furthermore, since there was no clear strategy to guide the campaign, effectiveness was difficult to measure. Nevertheless, important results were achieved, leading the evaluators to conclude that the campaign was moderately successful. The results were manifested at different levels and include, for example, heightened political commitment in certain instances, awareness among affected women and greater capacities among partners. For instance, as a direct result of the Campaign, hotlines in Ecuador saw an increase of incoming calls by 20% directly after campaigns, and many of the calls originated from more vulnerable women. At the same time, the campaign process contributed to there being capacities in place to address these calls. **However**, to enhance effectiveness, the evaluators concluded that linkages should be sought with other related human rights issues (child rights, nutritional issues, displacement, health) and thereby bring on board the whole UN system, donors and additional public institutions.

### 8.2.3 HIV/AIDS

Overall, the **HIV/AIDS** evaluation concluded that the programme was effective: it contributed to “the integration of gender and human rights into key policies, plans, and activities that address HIV/AIDS at the national level. UNIFEM’s initiatives to empower Positive Women’s networks helped to strengthen their ability to organise and to become more vocal advocates in national and global forums and also had positive effects on the lives of individual women... by piloting Gender Equality Zones, UNIFEM’s programming demonstrated how focused, integrated, community approaches can improve gender equality, empower women economically and socially, and reduce effects of HIV/AIDS on women and their communities.” **However**, two key factors that limited UNIFEM’s efforts to mainstream gender in support of the national response to HIV/AIDS were its lack of resources and limited presence or capacity in country.

### 8.2.4 Governance

**Burundi-Peace** achieved most of its key expected output and outcome results and even went beyond. The evaluators firmly held that the project was a good and cost-effective investment yielding strategic results: by supporting Burundian women of all levels and building their capacity, the project strengthened and enhanced women’s effective participation in the peace process. The project developed a unique methodology for engendering peace processes, which the evaluators claimed needed to be fully documented and disseminated, although it had already been replicated in DRC, Somalia, Liberia, and to an extent Timor Leste and Afghanistan.

The **Caucasus-Peace II** project did not fully achieve its outcomes, but significant results were achieved in all areas. The project resulted in improved understanding and capacities of national governments to implement and monitor CEDAW and SCR 1325; it increased participation of women in national policy development; improved capacities of ‘women for peace’ networks to advocate on gender, international human rights and peace-building. It established partnerships between government, national and international organisations. It contributed to a better understanding of the gender and peace issues among groups of students and youth and selected communities. “Members of the Coalition 1325 in Azerbaijan mentioned that the project opened a new a page in women’s movement, as capacity was strengthened and women from diverse parties were brought together in one coalition. Only this project managed to create a culture of dialogue at the national and international level and to develop political tolerance. It introduced the idea of compromise. Thanks to the project the coalition has now strong links with IDP-women, who are the most affected by the war.”

The **DRC-Peace** evaluation maintained that the project had unrealistic and overly ambitious goals. Although it did not benefit from the support or management of an adequate consultative committee and was under funded, it did “carry out important activities and achieved certain sustainable results.”

The **Morocco Political Participation** evaluation did not conclude on the programme’s overall effectiveness. However, the report provided an account of outputs and outcomes that were achieved. These included a relatively successful media campaigns on family law and violence against women; the formation of a lobby group aimed at law reform; and good use of ongoing micro-credit programmes to reach women and inform them about family law. Meanwhile, reaching rural women with information about women’s rights and family law was a challenge, particularly due to cultural issues. Furthermore, the report noted that the percentage of women in parliament dropped by 0,5% during the programming period.

The **Kenya G&G** evaluation highlighted that important outcomes had been achieved in relation to legislation and female representation at different political levels. **However**, poor baselines, inadequate indicators, weak monitoring and lack of training assessments made it nearly impossible for the evaluators to confidently conclude a causal relationship between these results and UNIFEM’s efforts.

The **Afghanistan** programme was unique in being UNIFEM’s first multi-faceted, multi-million dollar programme in a post-conflict country. The evaluation concluded that it was largely achieving its expected outcomes, despite working in a highly pressurised environment and facing a range of difficulties. UNIFEM “maintained a momentum in Afghanistan that fulfilled its programmatic needs”. Furthermore, UNIFEM experimented with piloting a number of projects. Of these, the evaluation established that “success was more evident where implementing modalities were straightforward and a project was clearly conceived and designed. Where matters have been more nebulous, performance was weaker.”

The mid-term evaluation of the **SEE-WR** was positive in terms of effectiveness. For instance, it found that outputs / outcomes related to its constitutional reform pillar have either been achieved or substantial progress has been made.

The **Kosovo GE&WR** project also achieved considerable progress in all of its results areas. A particular strength was its work on strengthening leadership commitment, technical capacity and accountability mechanisms for gender equality in mainstream institutions. UNIFEM played an important leadership role in relation to gender issues within the UN Kosovo Team’s (UNKT) work, and was effective in establishing and strengthening exchange of women leaders at the regional level.

**However**, the Afghanistan-WL, Kosovo GE&WR and SEE-WR programmes/projects **all had comparatively limited results for its objectives in the economic sector**. In Kosovo, UNIFEM did not have a systematic approach to its goal of enhancing NGO capacities in relation to evidence-based advocacy for gender equality in the economic sector. In the SEE-WR project, the slow progress towards the implementation of Women’s Economic Rights was to a large degree seen as UNIFEM’s lack of networks and comparative advantage in this field. The evaluation recommended that UNIFEM develop its economic sector network and forge strong links to institutions and agencies that are driving economic reform and privatisation. Similarly, achievement related to women’s economic security within the Afghanistan-WL programme was also considered a challenge. The evaluation indicated that UNIFEM appeared to have had a reactive response to developing a woman-focused livelihoods approach that had poor cohesion, lacked consistent strategic direction and a prior risk analysis.



### 8.2.5 Poverty Evaluations

**While the CEE-Economic Justice** was a success with regard to laying the foundations of education, production of tools and inspiration; it had only just begun to reap the benefits of the investments. The evaluation concluded that in the next year, “the implementing partners should focus on spreading understanding of economic rights to more women of CEE to the extent it is needed for effective representation and lobbying and on leading and promoting the strategic use of the tools for advocacy”.

The participatory **Kyrgyzstan-Land** project achieved its aim of promoting women’s equal rights to land by successfully employing a comprehensive approach that involved interventions at both policy and grassroots levels, and supporting synergies between them. The project capitalised on the decentralisation process to enhance capacities of rural women to articulate and advocate their needs while at the same time sensitising local governments to the need of considering gender equality in local planning and budgeting. According to the report, the project was most successful when it linked its activities to supporting women in mobilising villagers to identify priority problems at the local level, and then jointly assisting women’s groups and local governments in drafting local development plans and budgets with consideration for gender needs. Furthermore, each focus group held during the evaluation process provided specific examples of women, who upon receiving legal advice, managed to restore their property rights. Moreover, by providing timely and useful training (on land use & management, co-operatives etc.), and using small grants to support upgrading of irrigation canals and roads, agricultural productivity and income were improved in many communities.

The **IFAD** evaluation found that the programme had been effective on the policy level and moderately effective at the learning level. At the operations level, results were mixed. While training at the operations level was seen as effective, the strategy to mainstream gender equality into operations was never adopted.

According to the **Nigeria Empowerment** evaluation, most of the outputs of this project were achieved, except the ICT component (see section above under “relevance”). The report claims that UNIFEM performed particularly well in contributing to reducing poverty and elevating the status of women. By building the capacities of local government, women development centres and community stakeholders, the project created an environment that protected and promoted the rights of the women in project area.

The **GRB-LAC** evaluation concluded that important outcomes had been achieved in relation to gender responsive budgeting. However, poor baselines, inadequate indicators, weak monitoring and lack of training assessments made it nearly impossible for the evaluators to confidently conclude a causal relationship between these results and UNIFEM’s efforts.

The evaluation of the **Moroccan Literacy** project uncovered evidence that the project was not only successful in achieving literacy and numeracy among the women it targeted, it also empowered women by instilling confidence and savoir-faire. However, the project was not able to infuse knowledge of women’s rights or Moroccan family law. While cultural resistance and the level of disempowerment among the women when the project started were partly to blame, the evaluator deems that progress towards these goals would have been better if: i) more time for gender equality and human rights in the training sessions were planned for; ii) trainers were sufficiently knowledgeable in gender equality approaches and women’s rights; and iii) more effort was made to include husbands and male relatives within the context of the programme.

### **Measuring Effectiveness of Support to Networks**

A significant number of UNIFEM's interventions involve supporting networks. One of the primary roles of a network is to build cohesion within the region and to maintain visibility for the region's needs and perspectives on the international level. Neither responsibility, however, generates the kinds of results or impacts that donors seek. Each requires its own sorts of indicators – and recognition that much of the impact is invisible and intangible. In addition, a network should be enabling actors at the national and local levels where the effects ultimately take place. Hence networks may have more indirect than direct effects, and face some limitations over their ability to ensure results. This is quite a challenge for networks to the extent that most donors aim their funding at specific results but not at operations or institution-building itself.

(From CEE-Economic Justice)

## **8.3 Sustainability**

Sustainability is the continuation or longevity of benefits from a development intervention after the cessation of development assistance. Most of the evaluations that address sustainability (e.g. CEDAW-SEAP, HIV/AIDS, Kosovo GE&WR, SEE-WR, Kenya GE&WR, Kenya G&G, Afghanistan-WL, Burundi-Peace, DRC-Peace GRB-LAC, CEE-Economic Justice) imply that sustainability in terms of the mere continuation or duration of results has not been achieved. Financial sustainability was also deemed low:

“Partners and beneficiaries are not yet at a stage where they could continue this programme if donors were to withdraw”. (Kenya G&G)

Building sustainable capacity in governments –where staff turnover often is high – will require consistent and longer-term support. This is particularly true in challenging and volatile contexts (such as Kosovo, Burundi, Afghanistan, Caucasus and Kenya). Several of the evaluations therefore also judged sustainability in terms of ownership, commitment, networking, institutionalisation and ability to attract other donors. In these terms, most of the projects/programmes have gained important ground. However, in all cases, the evaluators deemed that continued financial and technical support would be required to achieve permanent change. For instance, the Burundi-Peace evaluation found that project partners and beneficiaries showed a strong ownership of results, thus contributing to their likely sustainability. However, the precariousness of the institutional and organisational environment in Burundi threatened sustainability of results.

The evaluation of CEDAW-SEAP found that the concept of sustainability underlying UNIFEM's work in the area of implementing CEDAW had largely remained implicit. This has limited learning from experience of what works and what does not, in terms of achieving sustainable results. The evaluations of CEDAW-SEAP, HIV/AIDS and Kosovo GE&WR provide (similar) valuable reflections on sustainability that could provide guidance to how UNIFEM could make sustainability a more explicit concept in its approach. These are included in the box below.

### Box 5.5: Reflections on Sustainability

Sustainability of results is not merely conditional on UNIFEM's partner institutions continuing to work in their exact current form and function. Instead, sustainability is also dependent on the continuation of processes, principles, and approaches to gender equality and women's rights, and their adaptation to evolving contexts. UNIFEM's work to date has been geared towards ensuring sustainability in this dynamic sense. Some key characteristics of UNIFEM's approach that are relevant in this context are:

- *Rights-Based Approach*: UNIFEM has continuously based its work on existing international agreements and instruments for gender equality, and has strengthened the capacity of a broad number of stakeholders (civil society and government) to effectively use these instruments as advocacy tools. Through its rights-based approach UNIFEM has helped to establish a stable point of reference for women advocates that will remain accessible regardless of changes within the Kosovo context.
- *Institutionalisation*: In UNIFEM's support has helped to institutionalise a commitment to gender equality in various public institutions. Having gender equality established as an integral part of the respective organisation's structure and strategies enhances the likelihood that gender issues will continue to be addressed even if individuals leave the organisation.
- *Ownership*: UNIFEM's work has been based on the expressed needs and priorities of local stakeholders and has thus focused on building local ownership for achievements. This increases the likelihood that stakeholders have an interest to defend and expand achievements to date.
- *Accountability*: UNIFEM has worked with both government and with civil society, and has facilitated connections between the two levels. By continuously working with duty bearers and with rights holders, UNIFEM has worked towards establishing public expectations of, and demand for transparent and accountable governance.
- *Networking*: UNIFEM has supported a number of charismatic (male and female) change agents. While some individuals may change their roles and positions over time, most are likely to continue to engage as advocates for gender equality and women's rights. The latter is furthered by UNIFEM's work on creating networks and fostering linkages between different sectors, initiatives and partners.
- *Enabling partners*: UNIFEM has assisted several partner organisations access funding from other donor agencies, thus allowing them to continue and expand their work, as well as enhance their knowledge and skills related to securing resources.

(From the Evaluation of *Advancing Gender Equality and Women's Rights in Kosovo*).

Some of the evaluations brought up the issue of **exit strategies and ethics** in the context of UNIFEM's engagement and subsequent disengagement in communities. The Caucasus Peace II evaluation saw the need for both a strategy for the sustainability of NGO networks and an overall 'exit' strategy. Similarly, the data collected by the HIV/AIDS evaluation suggested that UNIFEM would need to improve the design of transition, replication, or scale up strategies to address expectations raised during the life of the project. India-VAW found a sense of betrayal in a community where a zero violence zone project had been piloted. It suggested the need for clarity on the rights of communities as subjects of initiatives. Furthermore, it recommended a mechanism for an 'ethical review' projects, from pre-proposal, implementation to the reporting stages. This would include preparing a community from the beginning for the eventual exit.

## 8.4 Efficiency

Efficiency measures the outputs – qualitative and quantitative – in relation to the inputs. It is an economic term which signifies that the assistance uses the least costly resources possible in order to achieve the desired results.<sup>26</sup> Relatively few evaluations assessed efficiency. All of those that did concurred with the MYFF evaluation conclusion that UNIFEM has produced remarkable results with

---

<sup>26</sup> As defined by OECD/DAC.

very limited resources (16 Day Campaign, India-VAW, Burundi-Peace, Kosovo GE&WR, Kenya G&G, Caucasus Peace-II, GRB-LAC).

However, efficiency could be enhanced by better financial monitoring. The GRB-LAC and Kenya G&G evaluations held that efficiencies could be gained through more effective results based management (in particular through improved monitoring and evaluation – see section 7.1 above); less staff turnover; and enhanced communication practices (see section 7.7 above). Caucasus Peace-II maintained that the introduction of the Atlas system made financial project management less efficient (designed for large operations, requires high speed Internet connectivity, etc.) and led to increased workload of the project staff. Participants in the Kyrgyzstan-Land evaluation repeatedly noted that political instability and frequent shifts of government officials affected the project's efficiency.

While the DRC-Peace evaluation did not assess efficiency, the evaluators implied that this project's inability to establish a consultative committee with some of the other UN agencies made operating costs higher than it could have been. Transport on the UN planes could have been provided to the project's trainers when they needed to go to the regions if UNIFEM's collaboration with these agencies had worked out. Budget costs for transportation would have been reduced and this money could have been availed for other activities.

## 8.5 Impact

Impact is the totality of effects brought about through a development intervention. It encompasses expected and unexpected, positive and negative, as well as short-term and long-term effects on people, organisations, societies and the physical environment. It is arguably the most difficult and demanding part of evaluation work.

Two evaluations (Morocco-Literacy and Kyrgyzstan-Land) made systematic efforts at establishing evidence of change.<sup>27</sup> Both evaluations showed methodological rigour, worked from baselines and extensively applied participatory techniques to determine change. (The results of these evaluations are discussed below.)

However, none of the other the evaluations between 2004 and 2008 thoroughly assessed the totality of effects brought about through the development intervention in question, or took into account the counterfactual. Moreover, the fact that some terms of references and reports use the term "impact" loosely or incorrectly (e.g. Local-Global, India-VAW), suggests that there has been a lack of clarity within UNIFEM and among some of its evaluators concerning the meaning and implications of this assessment criterion.

A few evaluations (16 Day Campaign, Caucasus-Peace II, SEE-WR, DRC-Peace, CEE Economic Justice, Nigeria-Empowerment) justifiably maintain that assessing impact was not possible to assess due to the insufficient passage of time, lack of baselines and/or problems of inferring causality. Furthermore, assessing impact was made more difficult in that UNIFEM had generally not been systematic in reporting on longer-term effects of capacity-building efforts.

---

<sup>27</sup> Neither of these two reports call themselves "evaluations" and both state that their main aim is to assess change at the community level.

Nevertheless, some evaluations reveal indications of impact, although the methodological rigour applied to reach these conclusions may, in some cases, be questionable. The types of impact described can be loosely divided into micro (community level), meso (institutional level) and macro (policy and legislative level) levels.

### 8.5.1 Macro

At the macro-level, a handful of evaluations concluded that the UNIFEM intervention influenced legislation:

- **SEELINE** noted that at the beginning of the project, none of the countries involved had gender equality laws or specific equality opportunity or anti-discriminatory laws. By the time the report was written, Romania, Croatia and Bosnia Herzegovina passed such laws. The report does not provide much insight into the inferred causality.
- The **SEE-WR** evaluation, mentioned legislative success in Montenegro that it regarded as a result of the programme.
- The **CEDAW-SEAP** report held that there were several examples of UNIFEM-supported exchanges having had direct macro-level impact and provided the example of Vietnamese members of parliament that visited the Philippines and Thailand to learn about the domestic violence laws and subsequently using the emerging information to draft the Vietnamese law on domestic violence.
- The **Kyrgyzstan-Land** evaluation showed that increased legal knowledge among rural women, local public organisations and local government bodies through training offered by UNIFEM contributed to the analysis of cases in which women's land rights had been violated. This in turn served as a solid basis for the advocacy that led to the amendments of the Law on Managing Agricultural Lands. This macro-level change in turn further supported women's land rights at the community level.
- The evaluation of the programme in **Afghanistan** found that by successfully lobbying for the enshrining of women's rights into key instruments, UNIFEM successfully contributed to providing the formal framework for the realisation of those rights.
- The **Burundi-Peace** evaluation deemed that UNIFEM's project significantly contributed to engendering the Arusha Peace Accord. Making their mark on Burundi's history further empowered Burundian women leaders, who according to the evaluation, will "forever strive to fight for their rights and interests".

### 8.5.2 Meso

Indications at the meso-level were less reported on in the evaluations. They tended to relate to raised consciousness of gender equality issues within institutions, resulting in subsequent actions:

- The **SEE-WR** evaluation found that government, political parties and gender equality advocates were empowered by the project. For instance, the report mentioned that gender equality advocates participated in high level policy debates and provided input to drafting of laws and regulations.

- The **CEDAW-SEAP** evaluation reported impact at institutional and individual levels: 85% of survey respondents from governments and NGOs reported that participating in the training sessions provided by the programme impacted their work. Furthermore, several women mentioned that the training has also influenced their personal lives. They reported, for example, that the training “had helped them to become more self-confident, that it had led them to express their opinions and questions more freely, and that it had also positively influenced how they were bringing up their – male and female – children”.
- Within the context of the **Burundi-Peace** project, UNIFEM sensitised and lobbied key national, sub-regional and regional organisations on the importance of allowing Burundian women’s rightful participation in the peace process. The evaluation reported that people from these institutions felt that UNIFEM “had an impact on making the institutional and policy-making environment more women and gender-friendly. The key men involved were sensitised and became more aware of women’s needs and the advantages of including them in negotiations and in the peace process. At the same time, the interviewees said that as compared to before, many of these institutions have become more knowledgeable on the concept of gender equality and show increased sensitivity to gender issues related to the implementation of the Peace Agreement.”
- The **India-VAW** evaluation regarded the “massive exploration” of the subject of men and masculinities as “the most outstanding, strategically creative, innovative and pioneering achievement and impact” of the project. While this assessment is perhaps overenthusiastic, the project did bring men and masculinities to the gender equality agenda in India for the first time. Spin-off effects have included academic research on the subject that even received attention at the international level.

### 8.5.3 Micro

Five evaluations offer indications of some impact at the community level:

- The **Burundi-Peace** evaluation told of how members of URAMA (which means “peace light” in Kirundi) –a refugee women’s group located in two camps in Kasulu, Tanzania – were trained in conflict resolution and mediation through the UNIFEM project. “The 70 trainees were able to go back to their camps and transfer the acquired knowledge and skills to refugee men, women, youth and people responsible for the local court system.... the impact of this training and sensitisation has enhanced community-based conflict resolution and increased women’s awareness as to the importance of their roles in the search for lasting peace.”
- The participatory **Kyrgyzstan-Land** evaluation uncovered evidence that showed that since the project, women became more active in attending meetings and participating in discussions about local development. Women also increased their participation in the courts of the elders, women’s committees, young people’s committees, farmers’ groups, cooperatives, NGOs and health committees. In two provinces, several women were elected as deputies of the local legislatures. Increased women’s participation, in turn, influenced public perceptions of women’s role in the family and society. Villagers noted that women’s participation resulted in visible improvements in their villages. Women’s self-confidence rose as they realised that their actions could change their lives and their villages. Furthermore, using baseline data, the evaluation concluded that by improving land use and management and supporting women’s mini-enterprise development, the

project also helped increase the number of medium and rich households, while the number of poor families decreased.

- The **India-VAW** project included piloting the concept of a “zero violence zone” in a rural and an urban community. While the evaluator did not believe that the level of violence against women had changed as a result of the project, there was “a visible change in the perception on women’s rights, which was scoffed at previously, but is taken seriously even by men and within the families”. Women’s mobility, access to health care and girls’ education were also allegedly improved.
- The **Morocco Literacy** evaluation gathered evidence that showed that women became significantly more autonomous in their mobility in public spaces; enhanced their control over their financial, personal and/or health issues; and were better able assist their children with school. They gained the capacity to participate in society and were more sensitive to issues of the public domain. Furthermore, some of the women took on the role of being change agents by passing on knowledge to other women and family members. However, in others areas there was no impact: most women i) still believed that violence against women was justified in certain circumstances; ii) did not play a greater role in important household decisions; and iii) were not able to gain access to regular income.
- The **Nigeria Empowerment** evaluation reported that women who participated in the project felt “enlightened” and “empowered”. Their confidence was boosted. They claimed that they had more influence in household decisions, increased household earnings and greater economic autonomy. In one project site, the women were excited by the enhanced status of their older girls/younger women who are now able to remain in the villages until marriage and earn an income as opposed to seeking work away from home. A welcomed unintended effect of the project was that it served as a peace-broker between women from warring factions by forcing them to work together and achieve results. However, the evaluator concludes that one single project, particularly one running only two years, is not enough to empower the women economically.

# Annex 1

## TORs for SSA Contract

### **Evaluation Task Manager of the Evaluation of UN Trust Fund 2005-2008 Strategy and Meta-evaluation study of UNIFEM evaluations 2004-2008**

#### **1. Background and purpose**

The purpose of these Terms of Reference is to recruit an independent consultant to undertake 2 evaluation related tasks: the Task management of the external evaluation of the UN Trust Fund in Support of Actions to Eliminate Violence against Women, and a meta-evaluation study of UNIFEM evaluations 2004-2007. The Evaluation Manager will be supervised by the UNIFEM Evaluation Advisor. The Evaluation Unit is an independent entity that reports directly to UNIFEM's Executive Director.

#### **Task 1. Evaluation Task Manager for the Evaluation of the UN Trust Fund**

The UN Trust Fund in Support of Actions to Eliminate Violence against Women ("The UN Trust Fund") is a leading global multi-lateral mechanism supporting national efforts to end one of the most widespread human rights violations in the world. It was established in 1996 by a UN General Assembly Resolution, designating the United Nations Development Fund for Women (UNIFEM) as the UN Trust Fund's Administrator on behalf of the UN System.<sup>28</sup> Primary administrative responsibility rests with UN Trust Fund staff housed at UNIFEM Headquarters in New York, supported by UNIFEM focal points in its sub-regional offices. UN Trust Fund operations, including decision-making on grants, are guided by inter-agency consultative mechanisms at global and sub-regional levels comprised of UN sister agencies, non-governmental organizations and other experts (known as 'PACs' – Programme Appraisal Committees). Each year, UNIFEM reports on the UN Trust Fund's progress to the UN Commission on the Status of Women and the UN Human Rights Council.

#### **Evaluation objectives and approach**

The UN Trust Fund's external evaluation will be focused on assessing the implementation of its 2005-2008 Strategy at both global and country levels. The evaluation will serve as an important source of information for the Trust Fund's main stakeholders and provide key recommendations for strengthening the effectiveness of the Trust Fund's role in ending violence against women and girls, as well as for developing the new Strategy in 2009. It will be a formative evaluation, which will focus on two main aspects; 1) the overall implementation of the UN Trust Fund Strategy; and, 2) selected case studies of UN Trust Fund grantees in selected countries.

The objectives of this evaluation are:

1. To assess the overall implementation and effectiveness of the 2005-2008 Strategy;

---

<sup>28</sup> "The Role of the United Nations Development Fund for Women in Ending Violence Against Women, General Assembly Resolution 50/166 of 22 December 1995.



2. To identify strengths, weaknesses, challenges and current trends in the implementation of the UN Trust Fund activities during the 2005-2008 Strategy period and implications for managerial, administrative, programmatic and funding directions.
3. To provide forward-looking recommendations, including for the development of the new Trust Fund Strategy.

The subject of the evaluation will be to analyze the effectiveness, efficiency and relevance of the UN Trust Fund Strategy. It will provide recommendations for how to strengthen UN Trust Fund activities, particularly in light of the high potential for considerable expansion over the coming years. It will provide pertinent information to key UN Trust Fund stakeholders, including UNIFEM as Administrator, UN partner agencies, donors and other Member States, and interested civil society organizations. It will also serve as a baseline study for future UN Trust Fund evaluations. The evaluation will follow the United Nations Evaluation Group's norms and standards to ensure the highest quality, and will adopt a human rights-based approach and gender perspective.

The evaluation will be conducted respecting the standards and norms of the UNEG, and will be undertaken integrating a human rights and gender equality approach.

## **2. Activities and deliverables of the Evaluation Task Manager**

The Evaluation Task Manager, under the supervision of the UNIFEM Evaluation Advisor, will be responsible for coordinating and managing the external evaluation. These tasks will include:

- Liaising with the Evaluation Team, including to ensure timely delivery of documentation and information requested; organizing meetings and interviews requested by the Evaluation Team with internal and external parties; and facilitating information flow;
- Monitoring the evaluation workplan and ensuring its timely completion, including time-line and deliverables of the Evaluation Team;
- Supporting the establishment and coordination of reference groups (Core Group, of UNIFEM and UN Trust Fund staff; Internal Reference Group, involving other UNIFEM units; External Reference Group, of UN agencies and international non-governmental organizations), facilitating a fully consultative process with members of these groups throughout the evaluation ;
- Organizing debriefings with the UN Trust Fund's main donors;
- Facilitating the preparation and organization of the country site visits, in coordination with the UN Trust Fund staff, including communications with the UNIFEM Sub-regional offices;
- Facilitating the coordination of the Evaluation Team's meetings in New York including the convening of briefings by the Evaluation Team to present preliminary and final findings and recommendations;

- Reviewing and technically assessing the Evaluation Team’s inception report, preliminary and final evaluation products, including following up with the Evaluation Team to ensure Evaluation Team products reflect all requirements for a high quality evaluation;
- Circulating the Evaluation Team’s products for review and compiling comments and feedback from internal and external stakeholders;
- The Evaluation Manager may be requested by the Evaluation Unit to participate in the country missions accompanying the Evaluation Team;
- Other tasks related to the evaluation, upon request by the Evaluation Unit.

The Evaluation Team reports through its Team Leader to the UNIFEM Evaluation Unit, directly to the assigned Evaluation Task Manager. The Evaluation Task Manager will be responsible for managing the evaluation process as a whole, follow up with the Evaluation Team related to their deliveries, the timely application of the work-plan.

***Specific deliverables for Task 1:***

1. Support the technical assessment for the selection process of the Evaluation Team and the recruitment and induction of the evaluation team (beginning November 2008);
2. Supervision of inception report until its completion to fully comply with evaluation quality standards, including sending it for consultation with key stakeholders (end November 2008);
3. Consultation for the establishment of selection criteria for country case studies, completion of selection of 4-6 case studies, and first case study lunched (December 2008);
4. 4-6 case studies completed and reports elaborated, reviewed, consulted and completed (February 2009);
5. Draft evaluation report reviewed, consulted and completed (March 2009).

**Task 2. Metaevaluation of UNIFEM evaluations 2004-2007**

Under the MYFF programming period 2004-2007, UNIFEM conducted 35 evaluations of programmes/projects in different sectors of UNIFEM areas of work. These evaluations were conducted by UNIFEM HQ, SROs or UNIFEM partners on programmes implemented by the organization (Annex 1). These evaluations were used in different ways by UNIFEM offices, as reported in Annual Reports, but were not systematically tracked to analyze how they were used, nor analyzed in terms of their evaluation quality.

As part of UNIFEM Evaluation strategy in trying to strengthen learning and use of evaluations on what works and doesn’t for gender equality and women’s human rights, there is a need to analyze and aggregate the findings of these evaluations in order to systematize the information generated and transform it into accessible knowledge for programming. To do this, consultant support is requested to undertake a meta-evaluation.

The study has a double purpose:

1. Contribute to learning through the identification of key findings, lessons learned and evidence on the key areas of UNIFEM work, namely women's economic security and rights, ending violence against women, HIV/AIDS and gender justice in democratic governance, and any indications of UNIFEM contribution to its overarching corporate goal under the MYFF;
2. Determine and uncover areas for improvement in the quality of the evaluations undertaken, identifying factors that may have weakened the validity, credibility and utility of evaluations conducted, based on UNEG Norms and Standards and UNIFEM guidelines.

The results of the meta evaluation will be used by UNIFEM Evaluation Unit as a basis to have an evidence based analysis of key strategies used by UNIFEM in relevant areas of work and will be shared with UNIFEM Programmatic units in order to better inform their programming choices. It will also be used to identify the most important gaps in terms of evaluation quality in order to fill them in the current evaluation strategy.

The key areas of analysis are:

- Evaluation quality: design of terms of reference, methods and techniques used, implementation processes, presentation of findings, evidence base of findings presented, use of evaluation findings;
- Key findings on programmatic areas: lessons learned on successful strategies used to advance the gender equality agenda, main constraints of programmes implemented, and

### ***Activities and deliverables for Task 2:***

The study will be undertaken based primarily on secondary data as reflected in the evaluation reports, UNIFEM annual reports, and UNEG Norms and Standards. Secondly and if identified as needed, the consultant may gather primary information regarding the evaluation process and use through the conduction of a reduced number of targeted interviews.

- **Stage 0:** Review the matrix of UNIFEM evaluations and a sample of 3 reports to develop a proposed methodology and meta evaluation framework to undertake the study (October);
- **Stage 1:** Undertake a mapping of the evaluations, based on criteria identified in stage 0. The mapping criteria could include the thematic and geographic concentration of evaluations, the purpose, the (type of) aim, the evaluation criteria used, the type of methodology and evaluation techniques used, the input (person days, expenses), etc. A mapping document would be produced. (October)
- **Stage 2:** To the extent possible, examine the validity, utility, credibility, conduct and cost-effectiveness of the evaluations undertaken. Does each evaluation provide valid and justifiable conclusions? Did it ask the right questions? What was left out that might have been included and vice versa? Is it useful to stakeholders? Was the evaluation credible to relevance audiences? Was the team familiar with the context and did it have sufficient evaluation expertise? Was the evaluation appropriately conducted according to legal, ethical, professional standards and

cultural appropriateness (UNEG Standards)? Was the evaluation conducted in an economical, quick and unobtrusive manner? (November)

- **Stage 3:** Analyse the results of the 35 evaluations undertaken. What strengths and challenges do the evaluations expose with regard to UNIFEM's effort to (1) reduce feminized poverty; i) end violence against women; iii) reverse the spread of HIV/AIDS among women and girls; and, iv) achieve gender equality in democratic governance in times of peace as well as war? (December)
- **Stage 4:** Synthesize the data gathered. Draw conclusions, provide recommendations and draft a report would be produced. (January)

### **Task 3. Meta-assessment of UNIFEM evaluations 2008**

In addition to the two previous tasks outlined above, the consultant is required to expand the research and analysis on UNIFEM evaluations to include those evaluations conducted in 2008. Given the current preparation of UNIFEM annual reports from SROs, as well as from HQ sections and the Evaluation Unit, it is necessary to include in the scope of the Meta-evaluation, **all** available evaluations completed by UNIFEM in 2008. This analysis will assess the quality of the 2008 evaluations and identify the key lessons learned.

The scope of 2008 evaluations include 12 evaluation reports, their mapping, review and scoring, additional interviews and in-depth analysis of evaluation findings.

### **3. Ethical code of conduct**

Following United Nations Evaluation Group (UNEG) ethical guidelines, the integrity of evaluation is especially dependent on the ethical conduct of key actors in the evaluation process, including the Evaluation Task Manager. Evaluation Managers have a duty to:

- a) Appoint trustworthy, competent and independent-minded evaluators with the appropriate mix of experience, expertise and competencies, and with an appropriate diversity in gender, ethnicity, religion and language.
- b) Consult with evaluators and other interest groups if significant changes are required to the design or delivery of the evaluation.
- c) Provide the evaluators with access to the documentation and data required for evaluation purposes.
- d) Communicate openly and have respect for people involved in the evaluation and keep the evaluation team informed of changes in circumstances affecting the evaluation.
- e) Respect the evaluators' duty to keep their sources of information anonymous.
- f) Anticipate the different positions of various interest groups and minimise attempts to curtail the evaluation or bias or misapply the results.
- g) Design the evaluation to encourage stakeholders to follow-through and maximize the use of the evaluation results.
- h) Provide all evaluation team members with an opportunity to disassociate themselves from particular judgements and recommendations, with unresolved differences of opinion within the team acknowledged in the evaluation report.

The Evaluation task Manager is required to commit to the UNEG code of conduct, sepecifically to the following obligations:

- **Independence:** Evaluators shall ensure that independence of judgment is maintained and that evaluation findings and recommendations are independently presented.
- **Impartiality:** Evaluators shall operate in an impartial and unbiased manner and give a balanced presentation of strengths and weaknesses of the policy, program, project or organizational unit being evaluated.
- **Conflict of Interest:** Evaluators are required to disclose in writing any past experience, which may give rise to a potential conflict of interest, and to deal honestly in resolving any conflict of interest which may arise.
- **Honesty and Integrity:** Evaluators shall show honesty and integrity in their own behavior, negotiating honestly the evaluation costs, tasks, limitations, scope of results likely to be obtained, while accurately presenting their procedures, data and findings and highlighting any limitations or uncertainties of interpretation within the evaluation.
- **Competence:** Evaluators shall accurately represent their level of skills and knowledge and work only within the limits of their professional training and abilities in evaluation, declining assignments for which they do not have the skills and experience to complete successfully.
- **Accountability:** Evaluators are accountable for the completion of the agreed evaluation deliverables within the timeframe and budget agreed, while operating in a cost effective manner.
- **Obligations to Participants:** Evaluators shall respect and protect the rights and welfare of human subjects and communities, in accordance with the UN Universal Declaration of Human Rights and other human rights conventions. Evaluators shall respect differences in culture, local customs, religious beliefs and practices, personal interaction, gender roles, disability, age and ethnicity, while using evaluation instruments appropriate to the cultural setting. Evaluators shall ensure prospective participants are treated as autonomous agents, free to choose whether to participate in the evaluation, while ensuring that the relatively powerless are represented.
- **Confidentiality:** Evaluators shall respect people's right to provide information in confidence and make participants aware of the scope and limits of confidentiality, while ensuring that sensitive information cannot be traced to its source.
- **Avoidance of Harm:** Evaluators shall act to minimize risks and harms to, and burdens on, those participating in the evaluation, without compromising the integrity of the evaluation findings.

- **Accuracy, Completeness and Reliability:** Evaluators have an obligation to ensure that evaluation reports and presentations are accurate, complete and reliable. Evaluators shall explicitly justify judgments, findings and conclusions and show their underlying rationale, so that stakeholders are in a position to assess them.
- **Transparency:** Evaluators shall clearly communicate to stakeholders the purpose of the evaluation, the criteria applied and the intended use of findings. Evaluators shall ensure that stakeholders have a say in shaping the evaluation and shall ensure that all documentation is readily available to and understood by stakeholders.
- **Omissions and wrongdoing:** Where evaluators find evidence of wrong-doing or unethical conduct, they are obliged to report it to the proper oversight authority.

---

#### 4. Required competencies and skills

##### Evaluation Task Manager

- At least a master's degree related to any of the social sciences, preferably including gender studies, evaluation or social research;
- At least 7 years of working experience in evaluation and/or social research;
- Experience in evaluation of large programmes involving multi-countries and multiple stakeholders;
- Proven experience as evaluation manager with ability to lead and work with other evaluation experts;
- Demonstrating strong analytical ability and communication skills;
- Ability to work with the organization commissioning the evaluation, the Evaluation Team and with other evaluation stakeholders to ensure that a high quality product is delivered on a timely basis;
- Proven strong ability as evaluation manager and team work facilitator;
- Experience and background on human rights based approach to programming and gender equality;
- Experience working with the UN and with multi-stakeholders essential: governments, NGOs, and the UN/ multilateral/bilateral institutions and donor entities.
- Familiarity with the UNEG standards and norms for evaluations;
- Fluent in English and working knowledge of at least another UN language.

## **5. Time frame and payments**

On average an estimated 75% of full time person days per month (depending on the tasks and the progress of the evaluation itself) for a six month period, starting on October 13, 2008 – March 31, 2009. The payments will be issued on a monthly basis based on a record of actual days worked. The total budget estimated for the consultancy work is in a range between 48,000 and 52,000 USD, to be agreed with selected consultant.

It is estimated that an additional period ending 8 May 2009 will be required to undertake research and analysis on the completed 2008 evaluations. The total budget estimated for this additional task ranges between 4,800 and 7,000 USD. All other conditions and competencies remain as for the previous tasks.

## Annex 2 Quality Assessment & Scoring Framework

### *Quality Assessment & Scoring Framework for UNIFEM META-Evaluation*

The quality of the evaluations was assessed using 16 different parameters in the table below. The parameters have been largely derived from the UNEG standards. For 11 of the parameters, a score between 0 and 3 was awarded. For the parameters “findings”, “analysis” and “conclusions”; a scale of 0 to 5 was used since these areas are relatively more important than the other parameters. For the remaining 3 parameters – i) the inclusion of safeguards, ii) debriefing of stakeholders and/or ensuring a validation process at the end of the evaluation; and, iii) the existence of a management response to the evaluation – a score of 0 or 1 was given. An aggregation of the scores provides the “quality score” which is out of a total of 48 possible points.

	<b>Good (3)</b>	<b>Average (2)</b>	<b>Weak (1)</b>	<b>Not provided (0)</b>
<b>1. Basic key information</b>	<p>The title pages include the name of:</p> <ol style="list-style-type: none"> <li>1. The subject (i.e. activity, programme, policy etc.) being evaluated;</li> <li>2. Name and organisation(s) of the evaluators;</li> <li>3. Name of the organisation(s) that commissioned the evaluation.</li> <li>4. The date.</li> </ol>	3 of the 4 components in column 2 are provided.	Only 1 or 2 of the components in column 2 are provided.	No basic information is provided.
<b>2. Executive summary</b>	<p>The executive summary is a short stand-alone synopsis of the substantive elements of the evaluation report. It provides the uninitiated reader with a clear understanding of what was found and recommended and what has been learnt from the evaluation. It includes:</p> <ol style="list-style-type: none"> <li>1. Brief description of the subject being evaluated;</li> <li>2. Context, present situation, and description of the subject;</li> <li>3. Purpose of evaluation;</li> <li>4. Objectives of evaluation;</li> <li>5. Intended audience;</li> <li>6. Short description of methodology, including rationale for choice of methodology, data sources used, data collection &amp; analysis methods used, and major limitations;</li> </ol>	The executive summary includes half or more of the components in column 2.	The executive summary includes less than half of the components in column 2 and cannot be understood by the uninitiated reader.	No executive summary included.



	<p>7. Most important findings &amp; conclusions;</p> <p>8. Main recommendations.</p>			
<p><b>3. Purpose of the evaluation, objectives and scope</b></p>	<p>The purpose of the evaluation is described including:</p> <ul style="list-style-type: none"> <li>• why the evaluation is being done</li> <li>• how it will be used</li> <li>• what decisions will be taken after the evaluation is complete.</li> <li>• The context of the evaluation is described to provide an understanding of the setting in which the evaluation took place.</li> </ul> <p>The evaluation report provides a clear explanation of the objectives and scope of the evaluation. It includes the following:</p> <ol style="list-style-type: none"> <li>1. The limits of the evaluation are acknowledged.</li> <li>2. The original evaluation questions are explained, as well as those that were added during the evaluation.</li> <li>3. An explanation of the evaluation criteria used is provided</li> <li>4. Performance standards or benchmarks used in the evaluation are described.</li> </ol>	<p>The purpose of the evaluation is described including half or more of the components in column 2.</p> <p>Evaluation objectives and scope are explained briefly. A limited discussion on the evaluation criteria and performance standards and benchmarks is provided.</p>	<p>Minimal information is provided on the purpose of the evaluation.</p> <p>The evaluation reports states the objectives and scope of the evaluation but provides no explanation.</p>	<p>No purpose is provided.</p> <p>The scope and objectives of the evaluation is not provided.</p>
<p><b>4. Context of subject</b></p>	<p>An explanation of context is provided that contributes to the utility and accuracy of the evaluation. This includes an analytical account of the contextual background (institutional, political, socio-economic, cultural, geographical or any other relevant context information). Relevant statistics to support the contextual analysis are provided.</p>	<p>The context is provided but it is limited in scope or includes information that is not directly important for the evaluation.</p>	<p>The context is described in minimal terms and can barely be understood by the uninitiated reader.</p>	<p>No contextual information is provided.</p>
<p><b>5. Description of the subject</b></p>	<p>The subject being evaluated is clearly described.</p> <p>Information is also provided on:</p> <ol style="list-style-type: none"> <li>1. Overall purpose &amp; goals of the intervention</li> <li>2. Logic model and/or the expected results chain and intended impact</li> <li>3. Implementation strategy and key assumptions</li> <li>4. Importance, scope and scale of the subject being evaluated</li> <li>5. Recipients / intended beneficiaries</li> </ol>	<p>More than half of the components in column 2 are provided.</p>	<p>The subject is described in minimal terms (less than half of the information in column 2), making it very difficult to understand the evaluation. The subject can barely be understood by the uninitiated reader.</p>	<p>No description of the subject is provided.</p>

	<p>6. Budget figures</p> <p>7. Organisation and management of the project &amp; programme</p> <p>8. Stakeholders - their roles &amp; contributions to the subject being evaluated (financial resources, in-kind contributions, technical assistance, participation, staff time, training, leadership, advocacy, lobbying, and any contributions from primary stakeholders, such as communities). An attempt is made to clarify what partners contributed to which outcome.</p>			
<b>6. Recommendations</b>	<p>Recommendations are:</p> <ol style="list-style-type: none"> <li>1. firmly based on evidence and analysis</li> <li>2. relevant (to subject, ToR &amp; objectives of the evaluation)</li> <li>3. realistic, with priorities for action made clear</li> <li>4. formulated in a clear and concise manner</li> <li>5. prioritised to the extent possible and state responsibilities and the time frame for their implementation.</li> </ol>	3 of the components in column 2 are fulfilled.	Recommendations are provided but only 1 or 2 of the components in column 2 are provided.	No recommendations are provided.
<b>7. Lessons learnt</b>	<p>(Not all evaluations generate lessons).</p> <ol style="list-style-type: none"> <li>1. Lessons drawn represent contributions to general knowledge.</li> <li>2. Lessons are well supported by the findings and conclusions of the evaluation and are not a repetition of common knowledge.</li> <li>3. The analysis presents how lessons can be applied to different contexts and/or different sectors, and takes into account evidential limitations such as generalizing from single point observations.</li> </ol>	Some of the lessons represent contributions to general knowledge.	The lessons provided do not represent contributions to general knowledge.	No lessons learnt are provided-
<b>8. Gender equality and human rights</b>	<ol style="list-style-type: none"> <li>1. The evaluation uses i) rights-based language, ii) analyses results in relation to the human rights framework and integrates a gender and human rights perspective in the methodology.</li> <li>2. The evaluation report indicates the extent to which gender issues and relevant human rights considerations were incorporated where applicable. The evaluation examines :</li> </ol>	The report addresses half of the components in column 2.	The addresses less than half of the components in column 2.	Gender and human rights are not addressed.

	<p>a) If the project/programme was based on a sound gender analysis, to what extent gender equality issues were taken into consideration in programming, and the subject being evaluated promoted gender equality and gender-sensitivity.</p> <p>b) If the project/programme is based on a sound rights-based analysis and whether the subject being evaluated was informed by human rights treaties and instruments;</p> <p>c) How gaps were identified in the capacity of rights-holders to claim their rights, and of duty-bearers to fulfil their obligations, including an analysis of gender and marginalised and vulnerable groups, and how the design and implementation addressed these gaps;</p> <p>d) How the subject being evaluated monitored and viewed results within the rights framework.</p>			
<b>9. Stakeholder participation</b>	<ol style="list-style-type: none"> <li>1. Stakeholders participate actively meaningfully throughout the entire evaluation process.</li> <li>2. Participation includes both primary and secondary stakeholders (key stakeholders) and a rationale for why and a description how the different stakeholders were selected for participation.</li> <li>3. The methodology involves extensive use of participatory techniques that are clearly described.</li> </ol>	<p>There are some participatory elements in evaluation process, mainly for data gathering purposes – such as the use of focus groups, questionnaires, workshops, etc.</p> <p>The evaluation process is consultative.</p>	<p>There is evidence of some consultative processes. Participation is not active.</p>	<p>Stakeholder participation did not form part of the evaluation process.</p>
<b>10. Clear communication</b>	<p>Clear, precise and professional language used. Correct terminology and grammar are applied. Text is highly reader - friendly. Useful graphs and tables are included.</p>	<p>Language used is sometimes difficult to understand, imprecise and/or contains grammatical errors.</p>	<p>Poor language skills. Verbose. Unclear, imprecise and difficult to read. Emotive language.</p>	
<b>11. Ethical safeguards</b>	<p>The evaluation report includes a discussion of the extent to which the evaluation design included ethical safeguards where appropriate and expressly respected human rights and gender equality principles. This includes protection of the confidentiality, dignity, rights and welfare of human subjects, including children, and respect for the values of the</p>			<p>Ethical safeguards are not discussed.</p>

	beneficiary communities. (1 point)			
<b>12. Validation process</b>	An evaluation should include a verification and validation process with key stakeholders to provide an opportunity to correct factual errors and confirm the validity of conclusions. It can take the form of workshop(s), seminar(s), e-conference(s) etc. before the evaluation team drafts the evaluation report. (1 point)			
<b>13. Management response</b>	UNIFEM requires that an explicit response to an evaluation is prepared by the governing authorities and/or management entities that are addressed by its recommendations. As per the UNEG Norms, this may take the form of a management response, action plan and/or agreement clearly stating responsibilities and accountabilities. (1 point)			

	<b>Excellent (5)</b>	<b>Good (4)</b>	<b>Average (3)</b>	<b>Weak (2)</b>	<b>Very Weak (1)</b>	<b>Not provided (0)</b>
<b>14. Findings</b>	<ol style="list-style-type: none"> <li>Findings cover all of the evaluation objectives.</li> <li>There is evidence of findings being based on richness of data – including use of different data collection approaches (primary and secondary); data sources; types of data (quantitative and qualitative).</li> <li>The report makes a logical distinction in the findings, showing the progression from implementation to results with an appropriate measurement and analysis of the results chain, or a rationale as to why an analysis of results was not provided.</li> <li>Findings regarding performance, inputs for the completion of activities, or process achievements are distinguished clearly from outputs, outcomes and impact.</li> <li>Inputs, outputs, and outcomes / impacts are measured to the extent possible (or an appropriate rationale given as to why not).</li> </ol>	<p>The findings meet 5-6 of the components listed in column 2.</p>	<p>The findings meet 4 of the components listed in column 2.</p>	<p>The findings meet 3 of the components listed in column 2.</p>	<p>The findings meet 2 or less of the components listed in column 2.</p>	<p>No findings are provided.</p>

	<ul style="list-style-type: none"> <li>6. Outcomes and impacts include any unintended effects, whether beneficial or harmful.</li> <li>7. Any multiplier or downstream effects of the subject being evaluated are included. To the extent possible, these are measured either quantitatively or qualitatively.</li> </ul>					
<b>15. Analysis</b>	<ul style="list-style-type: none"> <li>1. The analysis is structured with a logical flow. Data and information are presented, analysed and interpreted systematically.</li> <li>2. Results attributed to the subject being evaluated are related back to the contributions of different stakeholders. There is a sense of proportionality between the relative contributions of each, and the results observed. (If such an analysis is not included in the report, the reason why it was not done has been clearly indicated.)</li> <li>3. Reasons for accomplishments and enabling factors are identified and analysed to the extent possible.</li> <li>4. Reasons for difficulties and constraining factors are identified and analysed to the extent possible.</li> <li>5. External factors contributing to the accomplishments and difficulties are identified and analysed to the extent possible, including the social, institutional, political, financial or environmental situation.</li> <li>6. Analysis of both threats and opportunities to build on is provided.</li> <li>7. The analysis determines how factors that have contributed to successes/failures will affect the future, or whether it could be replicated elsewhere.</li> </ul>	The analysis meets 5-6 of the components listed in column 2.	The analysis meets 4 of the components listed in column 2.	The analysis meets 3 of the components listed in column 2.	The analysis meets 2 or less of the components listed in column 2.	No analysis is provided.
<b>16. Conclusions</b>	<p>Conclusions are:</p> <ul style="list-style-type: none"> <li>1. substantiated by findings and analysis consistent with data collected and methodology,</li> <li>2. represent insights into identification and/or solutions of important problems or issues.</li> <li>3. add value to the findings.</li> </ul>	The conclusions meet 6 of the components listed in	The conclusions meet 5 of the components listed in	The conclusions meet 4 of the components listed in column 2.	The conclusions meet 3 or less of the components listed in column 2.	No conclusions are provided.

4. focus on issues of significance to the subject being evaluated, determined by the evaluation objectives and the key evaluation questions. column 2. column 2.
5. are provided for each evaluation criterion.

Furthermore:

6. Simple conclusions that are already well known and obvious are not useful are avoided
7. Tentative conclusions regarding attribution of results, include a discussion on what is known and what can plausibly be assumed in order to make the logic from findings to conclusions more transparent and credible.

## Annex 3 Criteria for determining an Evaluation or Review

The United Nations evaluation group (UNEG) defines an evaluation as follows:

An evaluation is an assessment, as *systematic* and *impartial* as possible, of an activity, project, programme, strategy, policy, topic, theme, sector, operational area, institutional performance, etc. It focuses on expected and achieved accomplishments, examining the results chain, processes, contextual factors and causality, in order to understand achievements or the lack thereof. It aims at determining the *relevance, impact, effectiveness, efficiency and sustainability* of the interventions and contributions of the organisations of the UN system. An evaluation should provide *evidence-based* information that is *credible, reliable* and *useful*, enabling the timely incorporation of findings, recommendations and lessons into the decision-making processes of the organisations of the UN system and its members.

This definition and the UNEG evaluation standards have constituted the basis for determining whether each of the 2004 to 2007 reports received by UNIFEM's Evaluation Unit qualifies as an evaluation.

According to UNEG, a review is the periodic or *ad hoc* often rapid assessments of the performance of an undertaking that do not apply the due process of evaluation". Reviews tend to emphasise operational issues. While an evaluation is generally considered more comprehensive and/or more in-depth than a review, distinguishing between an evaluation and a review is not always clear-cut.<sup>29</sup>

A review has many of the same characteristics as an evaluation. It can cover the same criteria, be evidence-based, systematic and impartial. Like an evaluation, it is a management tool. However, a review is related to the monitoring function of an intervention since it primarily serves as a tool for those responsible for the process of a given intervention. (In fact, reports that call themselves "external monitoring reports" can usually be classified reviews). Meanwhile, an evaluation is a management tool for organisations as a whole, with the aim of learning lessons from better practices which can be reflected in future work and strategies. It is also an important agent of change and plays a critical function in enhancing accountability.

To assess whether the 2008 reports are evaluations or reviews, the meta-evaluation has expanded on the UNEG definition of a review and developed the following criteria:

A review –

- focuses on performance and *operational* aspects

---

<sup>29</sup> In some cases, a "review" is deliberately called an "evaluation" and vice versa, for tactical reasons. For instance, if managers would like to raise the status of a review, it is called an evaluation. Likewise, particularly when a highly sensitive or controversial intervention is evaluated, the term "review" or "stock-taking" may be used to down-play the importance of the exercise. While some of the "evaluations" analysed in this exercise are actually reviews, it is not possible to determine whether UNIFEM managers have called them so for tactical reasons or because they have a different understanding of what constitutes an evaluation.

- has a more limited scope –it often only looks at the extent to which objectives are being or will be achieved (effectiveness)
- is conducted during a shorter timeframe
- is usually conducted while the project is on-going (often during its mid-term)<sup>30</sup>
- may be conducted internally, but can also be conducted by external consultants or both
- is less likely to focus on lessons that can be generalised on an institution-wide scale
- may not apply the due process of evaluation
- may be less rigorous in its approach to insure impartiality
- may be less rigorous in triangulating data.

---

<sup>30</sup> While evaluations in theory can be conducted at mid-term, when the timeframe of the project/programme is less than 5 years, there are often not enough results to assess relevance, effectiveness, impact, efficiency and sustainability – particularly when the programme/project is aimed at contributing to socio-economic, legal, socio-political and socio-cultural changes, which all take several years to have effects. Thus, the assessments of UNIFEM's efforts conducted during a project's mid-term are usually reviews in nature.



## Annex 4 Profile of Assessments undertaken from 2004 to 2007

### Evaluations from 2004 to 2007

#### 1. Gender and Economic Justice in European Accession and Integration (2004)

<i>Implementation period evaluated:</i>	2002-2003
<i>Thematic area:</i>	Poverty
<i>Region:</i>	CEE-CIS
<i>Budget:</i>	\$ ?
<i>Goal:/Activities:</i>	The project had three goals: (1) to promote gender equality and economic issues for women in CEE (by means of activities related to the EU Accession process), (2) to foster better understanding and cooperation between their region (“other countries of Europe”) and the EU and, in the process (3) to strengthen KARAT as a regional network.

This evaluation looks at results and performance. While the report sets out to look at impact, what is evaluated is actually the extent to which outcomes have been achieved. While the report is sometimes chatty and emotive, the evaluation represents a solid effort at gathering and analysing data. A questionnaire has been used but what data was gathered this way is not clear in the report. What questions were asked and the results of the questions is not presented.

*Quality score:* 33 **Good**

#### 2. Engendering Peace Building and Mediation Processes in Burundi (2004)

<i>Implementation period evaluated:</i>	2000-2003
<i>Thematic area:</i>	Governance
<i>Region:</i>	Africa: Burundi and Tanzania
<i>Budget:</i>	\$ ?
<i>Goal:/Activities:</i>	Strengthened participation of women in conflict resolution at the policy level; increased and operational gender mainstreaming in regional institutions; increased visibility of women’s peace initiatives and effective advocacy roles.

Given that this evaluation was conducted in less than three weeks, the resulting report is quite good. The aim of the evaluation can be summed as assessing effectiveness, relevance and sustainability. The report identified critical gaps, challenges, unintended results and lessons learnt. In a straightforward way, it seems that the evaluators systematically addressed the

areas they were asked to assess. There is some confusion and misuse of the terms impact and outcome, but this appears to stem from confusion within the project itself. More information on the background, including the monetary size of the project would have been desired. It seems that if these evaluators had been given more time, this report could have been even stronger.

*Quality score: 30 Average*

### **3. Evaluation of the Southeastern European Legal Initiative (SEELINE) (2004)**

<i>Implementation period evaluated:</i>	2000-2002
<i>Thematic area:</i>	Governance
<i>Region:</i>	CEE-CIS South-eastern Europe
<i>Budget:</i>	\$ 662,000
<i>Goal:/Activities:</i>	Engender legal frameworks of 12 countries in south-eastern Europe by strengthening national and regional efforts to lobby government to ensure the protection of women's rights.

The structure of report is unconventional. Much of the report consists of detailed descriptive text, documentary style – including the information on the methodology used. The project manager is referred to by first-name. The report does not include an executive summary, list of abbreviations or any of the required annexes. It appears that the evaluator has more of a legal background than knowledge of evaluation. The evaluation purports to address impact, but there appears to be a misuse of the term.

*Quality score: 17 Weak*

### **4. Women for Conflict Prevention and Peace Building in the Southern Caucasus (2004)**

<i>Implementation period evaluated:</i>	2001-2004
<i>Thematic area:</i>	Governance
<i>Region:</i>	CEE-CIS; Azerbaijan, Armenia, Georgia
<i>Budget:</i>	?
<i>Goal:/Activities:</i>	Support work's groups & networks to promote peace?

A strength of the report is that the evaluator uses a consistent and insightful rights-perspective in the analysis. Furthermore, the evaluator provides useful in-depth reflections on the context and ideas and for future opportunities. Performance and management issues are addressed. However, this evaluation was conducted in less than three weeks, which is reflected in the quality. The report is missing annexes. The executive summary does not allow an uninitiated reader to understand the subject or evaluation sufficiently. The project aims are not clearly conveyed. There is minimal information on the methodology used. The report provides what appear to be sensible recommendations, but they are not clearly presented as such. In some respects, the report likens a review. The reasoning and

knowledge conveyed in the report suggests that has there been more time available for the evaluation process, the report would have been of better quality.

*Quality score: 25 Average*

## **5. From Global to Local: A Convention Monitoring and Implementation Project (2004)**

<i>Implementation period evaluated:</i>	1997-2004
<i>Thematic area:</i>	Cross-cutting
<i>Region:</i>	Global
<i>Budget:</i>	\$ 756,000(for 7 years)
<i>Goal:/Activities:</i>	The project brings women civil society activists from countries that are presenting their periodic reports to the CEDAW Committee in New York to observe the committee meetings and participate in a two-week long practical training and lobbying process.

This evaluation aims to assess three different criteria – effectiveness, relevance and impact – but the evaluation questions related to these criteria suggests the terms are not used in an entirely conventional way. The evaluation was conducted as a desk study. A survey of UNIFEM regional offices and participants was undertaken, but the questionnaires are not included in the report. Of the 219 participants who participated in this project, only 11 returned questionnaires. These participants are identified in the report which presents ethical problems. Seven participants were contacted for interviews, but only 2 responded. Thus, a limitation of this evaluation is the available of sufficient data, but the report does not recognise this. Nevertheless, the evaluation has presented its findings in a systematic and logical way; undertaken analysis of solid quality and presented well-formulated recommendations.

*Quality score: 33 Good*

## **6. IGAD Gender Strategy in the Context of Globalisation and the Feminisation of Poverty (2005)**

<i>Implementation period evaluated:</i>	2001-2004
<i>Thematic area:</i>	Poverty
<i>Region:</i>	Africa: Djibouti, Eritrea, Kenya, Ethiopia, Sudan, Somalia, Uganda
<i>Budget:</i>	\$ ?
<i>Goal:/Activities:</i>	Empower African women to realise their economic rights and security in the context of globalisation and the feminisation of poverty.

This evaluation is very weak. Considering that the one-person evaluation team was asked to answer 28 evaluation questions pertaining to relevance, effectiveness, efficiency, sustainability and impact in under 14 days, this result is not surprising. The report structure and presentation of findings, conclusions and recommendations are not systematic. Much of

the data provided does not directly pertain to the project but refers to the general context. It is not clear how findings are backed by evidence. That a single evaluator managed to conduct 56 interviews, refer to 76 documents; conduct surveys and write a 34 page report in 14 days is impressive. The report appears to be written by someone with academic knowledge of the subject area, but not necessarily of evaluation techniques.

*Quality score: 15 Very weak*

### **7. Prevention of Trafficking in Women and Children in South Asia (2005)**

*Implementation period evaluated:* 2002-2004?  
*Thematic area:* Poverty  
*Region:* Asia: Nepal, India, Sri-Lanka, Bangladesh  
*Budget:* \$ ?  
*Goal/Activities:* ?

While this evaluation has generated some information, perspectives and evidence that seems valid and relevant, it is a very weak effort. There is no information on the project, its design, objectives, strategies or outputs. For the uninitiated reader, the report begs more questions than it answers. There are conclusions that lack backing from sufficient evidence. Parts of this document are written without sentences being complete. The three case studies are in bullet form, 2-3 pages long. The evaluation skills of the team appear to be minimal. This evaluation is interestingly an output mentioned in the evaluation of the Programme for Mainstreaming Gender in Poverty Reduction in Asia (see below).

*Quality score: 14 Very weak*

### **8. Reinventing India: Action for Empowerment to Address Violence against Women (2005)**

*Implementation period evaluated:* 2002-2004  
*Thematic area:* Violence  
*Region:* Asia - India  
*Budget:* \$ ?  
*Goal:/Activities:* Strengthen capacity of women's organisations & networks through research & information dissemination; increase women's understanding of and access to their legal rights; deepen and expand the forms of communication used to report an address violence against women and promote positive make role models.

The evaluation aims to address relevance and effectiveness, although these specific terms are not used. UNIFEM's and its partners' performance is examined. In its conclusions it also addresses efficiency and sustainability. The report structure is messy and the language used is emotive in several places. ("beautiful evolution", "simply brilliant" "extremely backward rural setting", field visit was "a deeply moving experience"). The report notes a number of activities were completed, but it does not attempt to explain why. Some statements are not backed by evidence or examples. For instance, the report makes the dubious assertion that

“The strategy of bringing out posters is to create one of the most effective ad time-tested mediums of sensitisation and information dissemination”. The evaluation also claims that the workshops (in which in only total 916 Indians participated) and posters “filled a very critical gap in the movement against VAW in a very substantial manner and made a high impact in touching very closely the lives of an awesome mass of humanity across the country”. However, the report offers no analysis of posters, where and how they were put up. It does not even a number of how many were produced. More than half of the required report content is not included. While this is a weak evaluation, the report does include several important findings.

*Quality score:* 16 **Weak**

### **9. UNIFEM Programming in HIV/AIDS (2006)**

<i>Implementation period evaluated:</i>	2002-2005
<i>Thematic area:</i>	HIV/AIDS
<i>Region:</i>	Global. Kenya, Nigeria, Rwanda, Senegal Zimbabwe, Cambodia, Thailand, India, Brazil, Honduras.
<i>Budget:</i>	\$ 3,000,000
<i>Goal:/Activities:</i>	Strengthened legal and policy frameworks to promote gender equality, women’s empowerment and stigma reduction in community infected, affected or at risk of HIV/AIDS; Heightened awareness of gender dimensions of HIV/AIDS among decision-makers & general public; Integrated community approaches contributing to gender equality.

While this evaluation is of top standard, more elaboration on the evaluation methodology and/or inclusion of a more clear and concise presentation of recommendations would have earned it an even higher mark. The evaluation assesses performance, effectiveness, relevance and sustainability of the programme. It analyses lessons learnt and recommends strategic direction. It makes a solid attempt to analyse gender equality and human rights perspectives throughout. The report is well-written, well-presented and uses data adeptly.

*Quality score:* 43 **Excellent**

### **10. Promoting Women’s Human Rights and Enhancing Gender Equality in Democratic Governance in Kenya (2006)**

<i>Implementation period evaluated:</i>	2004-2006
<i>Thematic area:</i>	Governance
<i>Region:</i>	Kenya
<i>Budget:</i>	SEK 54,000,000 (circa 7 million USD)
<i>Goal:/Activities:</i>	Engender the legal and policy framework for securing women’s human right; increase options, choices and capacities for Kenyan women in order to enhance women’s

organising, influencing and participation; support actions for gender responsive budgeting, resource planning, allocation and expenditure for gender equality.

This report is a mid-term evaluation. It is a relatively solid piece of work. Effectiveness, relevance, sustainability and performance are assessed. The report structure is somewhat confusing with the methodology being presented after 10 pages that describe the programme's context, rational and goal structure.

Quality score: 36 **Good**

### **11. Progressing the Implementation of CEDAW in the Pacific (2006)**

*Implementation period evaluated:* 2003-2006  
*Thematic area:* Cross-cutting  
*Region:* Asia: Fiji, Papua New Guinea, Samoa, Solomon Islands, Vanuatu and Tuvalu  
*Budget:* \$ 515,274  
*Goal:/Activities:* to build the capacity of the Governments and NGOs for effective implementation, monitoring and reporting on CEDAW; to facilitate sustainable partnerships at the national level between Governments and civil society for this purpose; and, to capture, expand and make accessible regional analytical skills and strategic knowledge for the implementation of CEDAW at the national and regional level.

Overall, this is a solid report. The team has gathered lots of data. It made a point of contacting participants of UNIFEM training and in some cases gathered data on how this training may have affected the work of the trainees. The recommendations are well supported by findings and conclusions. In line with the ToR, the evaluation assessed programmatic and organisational effectiveness. There are findings in the report that would support important conclusions on relevance, but this criterion is not mentioned in the ToR. Likewise, it would have been useful to include sustainability as a criterion. The structure of the report is not optimal. After a section that fairly systematically approaches results and findings, there is section called "analysis and recommendations", which effectively contains additional findings, conclusions and recommendations. This is followed by another section with six concluding sentences and 18 bullets good practice and lessons learnt.

Quality score: 33 **Good**

### **12. Women's Rights to Land in Kyrgyzstan (2006)**

*Implementation period evaluated:* 2004  
*Thematic area:* Poverty  
*Region:* CEE-CIS Kyrgyzstan  
*Budget:* \$ ?

*Goal:/Activities:*

Contribute to the efforts of government and civil society to achieve MDGs on poverty by targeting gender equality. Bring in the voices of rural women to advocate for their rights, influence important policy decisions and contribute to the process of development and implementation of the national poverty reduction strategy.

This evaluation calls itself a “participatory assessment”. Indeed, it is the only evaluation that scores full points on participation. The methodology used is innovative and is laid out, in relative detail, in the terms of reference. The terms of reference were discussed in a one-day workshop with a broad range of stakeholders to ensure that proposed methodology was viable. The evaluation addresses effectiveness, performance, and impact. The evaluation results are clearly presented in various forms of tables and diagrams. The concluding chapter would have been a little stronger if evaluation criteria were assessed in a more evident manner. The recommendations could also have been more concise and clearly formulated. Background on the project and context is sparse in the report but included in the terms of reference. Nevertheless, the report is inspiring and a commendable effort. The evaluators have made a credible effort to establish effectiveness and impact. The ToR and the evaluation represent good practice.

*Quality score: 33 Good*

### **13. Mainstreaming Gender in Poverty Reduction: IFAD-UNIFEM Gender Mainstreaming Programme in Asia (2007)**

*Implementation period evaluated:*

2004-2006

*Thematic area:*

Poverty

*Region:*

Asia: Bangladesh, Nepal, India, Pakistan, Sri Lanka; Cambodia, China, Mongolia, Laos Indonesia, Vietnam

*Budget:*

\$ 715,000 million

*Goal:/Activities:*

To enable women to overcome poverty and to increase their role in participatory governance and economic activities.

A very clear, well-argued and systematic evaluation. It is however, commissioned by and geared towards IFAD, so all recommendations address IFAD and not UNIFEM, who was the implementing partner. A mediocre self evaluation conducted by UNIFEM is included as an annex. It is output-focused and draws conclusions without backing of evidence e.g. “increased social acceptance of women’s economic contributions and rights to productive resources”.

*Quality score: 37 Good*

### **14. Mainstreaming Gender into the Electoral Process in DRC (2007)**

*Implementation period evaluated:*

2005-2006

*Thematic area:*

Governance

*Region:*

Africa: Democratic Republic of Congo

*Budget:*

\$ 1,46 million

*Goal:/Activities:* Ensure that the electoral process be inclusive and that women and other marginalised groups could actively participate in all stages of the elections.

Considering that this evaluation was undertaken in under 3 weeks, it is of high quality. The evaluation assesses effectiveness, relevance and sustainability. It examines performance and identifies lessons learnt. When assessing effectiveness, the report concentrates on outputs because of the lack of a baseline to assess changes against. Furthermore, attribution was considered a significant problem, given the number of actors involved in this area. It would have been nice to see an aggregation of the results from the questionnaire and reference to perspectives voiced during the focus group meetings. The report is relatively short but provides clear conclusions and recommendations.

*Quality score:* 35 **Good**

### **15. Supporting Women’s Leadership in Rebuilding Afghanistan (2007)**

*Implementation period evaluated:* 2002-2006  
*Thematic area:* Governance  
*Region:* Asia: Afghanistan  
*Budget:* \$ 9.8 million  
*Goal:/Activities:* To increase options and opportunities for Afghan women to transform the overall development of Afghanistan into a more equitable and sustainable process.

This report is highly informative and contains useful analysis, insight and conclusions. The structure of the report, however, lacks a clear flowing logic. The evaluation does not benefit from the report being divided into two parts. A couple of section titles are misleading or odd. Presentations of data, analysis and conclusions are not always systematised. The recommendations could benefit from being formulated in full sentences. Nevertheless, overall, this is an evaluation of good quality. The report uses abbreviations to an unnecessary extent that affects readability.

*Quality score:* 35 **Good**

### **16. Women for Conflict Prevention and Peace Building in the Southern Caucasus, Phase 2 (2007)**

*Implementation period evaluated:* 2004-2006  
*Thematic area:* Governance  
*Region:* CEE-CIS; Azerbaijan, Armenia, Georgia  
*Budget:* \$ 2,1 million  
*Goal:/Activities:* Develop a transformative peace agenda for Southern Caucasus which ensures the protection of women’s rights.



This is a solid report. It assesses relevance, effectiveness, sustainability, efficiency and performance. Lessons learnt are identified. Findings are presented by country, outcome and output. The conclusions section is of high quality. The background section could have been strengthened by including relevant statistics and a stronger rights perspective in the analysis. Overall a commendable report.

*Quality score: 34 Good*

### **17. Gender and the Millennium Development Goals (2007)**

<i>Implementation period evaluated:</i>	2004-2005
<i>Thematic area:</i>	Cross-cutting
<i>Region:</i>	Cambodia, Kenya, Kyrgyzstan, Morocco, Peru
<i>Budget:</i>	\$ 820,000
<i>Goal:/Activities:</i>	Produce innovative examples of how to bring gender equality perspective to MDG advocacy, reporting and monitoring.

The evaluation assesses effectiveness and gathers lessons learnt. It is a report of high quality. Its conclusions section, however, could have contained more substance. It was commissioned and managed jointly between UNDP and UNIFEM.

*Quality score: 37 Good*

### **18. Evaluation of UNIFEM's MYFF System (2007)**

<i>Implementation period evaluated:</i>	2004-2007
<i>Thematic area:</i>	Cross-cutting
<i>Region:</i>	Global
<i>Budget:</i>	\$
<i>Goal:/Activities:</i>	Reduce feminised poverty and exclusion; end violence against women; halt & reverse the spread of HIV/AIDS among women & girls; achieve gender equality in democratic governance.

This is well-written, well structured and well-presented evaluation. This evaluation would have been graded as excellent if it had received marks for lessons learnt.

*Quality score: 40 Good*

## Reviews from 2004 to 2007

The six reviews scored as weak or very weak (below 21 points out of 42):

### 1. Arab Regional Trust Fund for the Empowerment of Women, Phase I (2006)

*Implementation period evaluated:* 2006  
*Thematic area:* VAW  
*Region:* Arab region

This assessment is entitled a “mid-term review”. The data-gathering was systematic but was exclusively desk-based. What are actually outputs are referred to as “impact” in the report. The methodology involved minimal triangulation and verification of data. The report is overly descriptive and contains minimal analysis. The recommendations are often not formulated in a clear and concise manner and not firmly based on evidence and analysis.

*Quality score: 14 (very weak)*

### 2. Arab Women Parliamentarians Project (2007)

*Implementation period evaluated:* 2007  
*Thematic area:* Governance  
*Region:* Arab region

This assessment is entitled a “mid-term review”. Undertaken by the same consultant who reviewed the *Arab Regional Trust Fund for the Empowerment of Women*, it is very similar in approach and style. The data-gathering was relatively systematic but exclusively desk-based and was not analysed to any greater degree. The questions asked of stakeholders in a survey were often formulated in a way that promotes predictable responses. The concept of impact is used incorrectly. The methodology involved minimal triangulation and verification of data. Several of the recommendations are vague – e.g. “to ensure successful results and sustainability it is recommended that UNIFEM shall strengthen the role and engagement of the national counterparts in the countries where they are not strongly engaged”. The report is overly descriptive, contains little analysis, (for instance, no explanations are given for why certain outputs have not yet been achieved). It is a stock-taking exercise or, at best, a review without analysis.

*Quality score: 13 (very weak)*

### 3. Strengthening the Network of Home Based Workers SE Asia (2004):

*Implementation period evaluated:* 2004  
*Thematic area:* Poverty  
*Region:* Asia

This assessment calls itself an internal review. It was drafted as part of (only) a two-day participatory workshop held with the project teams. There is minimal evidence of triangulation and the document reads like a progress report.

*Quality score: 13 (very weak)*

### 4. Strengthening the Network of Home Based Workers South Asia (2005):

*Implementation period evaluated:* 2005  
*Thematic area:* Poverty  
*Region:* South Asia

This assessment was conducted at mid-term. The aim of the report resembles that of a mid-term review since it consists of taking stock of the situation and providing pointers, at least some of which seem useful, for future direction of the project. This includes identifying key areas of progress, gaps, trends, good practice and lessons. No evaluation criteria are used. It is not clear who undertook the evaluation. The structure of the report is unlike that of an evaluation. There is virtually no explanation about the context and very little about the project. The report consists of 4 parts, 3 covering a country each and one focusing the regional operation – but there is no section that aggregates findings and conclusions from these 4 parts. Recommendations are not delineated and need to be searched for in the text.

*Quality score: 14 (very weak)*

### **5. Stock-Taking of the Railways-UNIFEM Project on Equalizing Gender Relationship in the Context of the HIV/AIDS Epidemic (2005):**

*Implementation period evaluated:* 2005  
*Thematic area:* HIV/AIDS  
*Region:* India

This assessment is clearly a review, or, as stated in the title, a “stock-taking” exercise. It was conducted in under six weeks. No evaluation criteria are used but achievements are studied. The information on how the project was implemented is very detailed, operational in focus and not very evaluative. However, it is likely to be a useful account for capturing knowledge.

*Quality score: 24 (average)*

### **6. Gender and Justice in Afghanistan (2005)**

*Implementation period evaluated:* 2005  
*Thematic area:* Governance  
*Region:* Afghanistan

:

The objective of this study was to “review the current state of work on Gender and Justice Programme”. It was conducted over a three-week period. The study has a limited scope and a strong operational focus.

*Quality score: 20 (weak)*

The reviews have not been included in the analysis of the aggregate mapping and scoring exercise that is discussed in sections 3 and 4.

## Neither evaluations or reviews

As discussed above, some of the assessments do not purport to be evaluations, while other reports mistakenly call themselves evaluations:

1. The report *Elimination of Violence Against Women in Central Asia* (2004) calls itself a “final progress report” to the German government, which is exactly what it is.
2. The report *Empowering Women in Global Production Networks: Innovation in NTFP-based Livelihoods in India* (2006) calls itself a “paper” that “as it relates to community-based strategic change in gender relations and reduction in the feminisation of poverty”. While the report contains useful recommendations and a valuable analysis (an interesting read indeed!), it does not aim to evaluate the project, discuss its set-up in any detail, systematically provide findings or present a methodology.
3. The report *UNIFEM Peace & Security Programme in Sierra Leone* (2005) is a trip report prepared by a consultant who participated in a joint UNIFEM-donor mission of six days. The aim of the mission was to assess to what degree UNIFEM’s input made a difference to its key partners to subsequently inform the future programming process. The approach was not systemic or sufficiently evidence-based.
4. *Capacity Building Workshop for Mainstreaming Gender Analysis in HIV/AIDS Programming in the Caribbean* (2006) –is an independent assessment of a single training workshop. While training workshop assessments are commonly called “evaluations”, they are (or should be) considered part of the monitoring framework of a capacity-building effort.<sup>31</sup>
5. *Savings Mobilisation and Productive Investment Pilot in Iligan and Jagna - Asia, Pacific and Arab States Regional Programme on Empowering Women Migrant Workers in Asia* (2005): This is an 11-page double-spaced document that sets out to “evaluate” the project from “a gender-sensitive enterprise development perspective” and “highlighting strengths, problems and making concrete recommendations”. No evaluation criteria are used. More than 70% of the report is its recommendations chapter. There is no methodology section and it is unclear how data was gathered. There is minimal information on the project and no systematic presentation of findings. The recommendations are not firmly based on evidence and analysis, but rather seem like a number of ideas provided by the consultant. It is not clear whether the assessment was undertaken at mid-term or at the end of the project. This does not meet the criteria for a review but contains some information that could be of use to the project.
6. *Promotion of Women’s Human Rights through the Elimination of Violence against Women SE Asia* (2004). This report was written by the programme’s manager at the end of her tenure. The report is not impartial and there is no evidence of the data being triangulated. The report includes personal observations and opinions.

---

<sup>31</sup> On the whole, however, the evaluation reports reveal that UNIFEM’s monitoring effort of training activities is often weak. The reports directly or indirectly point to the general lack of follow-up after workshops. To adequately determine the effects of capacity-building, regular ex post surveying of participants is a useful tool. There is no evidence from the reports that this has taken place.

These six reports have been eliminated from the list of evaluations studied by the meta-evaluation.

# Annex 5 2008 Meta-Assessment

## 1 Introduction

As part of the effort to improve evaluations and strengthen the learning from them, UNIFEM's Evaluation Unit has identified the need to analyse its evaluations conducted since 2004. The Part 1 of the UNIFEM Meta-Evaluation includes the mapping and quality assessment of evaluations from the MYFF period – 2004 to 2007. This report, meanwhile, looks at evaluations conducted in 2008 and completed by February 10, 2009.

The meta-assessment consisted of the following activities:

- g) A database was established in which the characteristics of 13 reports were mapped. This included gathering information on the subject, timing, length, management, date and place of each evaluation. The content the reports and their annexes were also examined to assess the extent they complied with UNEG standards. In all, 8 aspects of the evaluations were entered in the database.
- h) Each report was assessed to determine whether it met the criteria for an evaluation or review, or whether it constituted another type of assessment altogether. The criteria used for this assessment are presented in Chapter 2.
- i) The quality of each evaluation/review was analysed. Based on UNEG standards, a framework including 16 parameters for assessing the quality of the evaluations was devised to score each report. The framework and scoring system is included in Annex 2.
- j) The data collected through the mapping and scoring exercise were aggregated and analysed.
- k) The results of the 9 evaluation reports were reviewed. Findings common to several reports and conclusions relating to relevance, effectiveness, efficiency, sustainability and impact were analysed. This analysis has been integrated in Part 2 of the meta-evaluation. Combining the results of the 2008 evaluations with those of the MYFF period to ensure relevance and illustrate the pervasiveness of certain issues since 2004.

The 2008 meta-assessment faced similar limitations as the overall meta-evaluation. These are discussed in the introduction of the main report.

### 1.1. Structure of the Report

This report consists of five chapters. The assessments from 2008 are introduced in Chapter 2. Chapter 3 maps the characteristics of the evaluations. The quality of each evaluation, in relation to UNEG standards, is scored in Chapter 4.

## 2 UNIFEM Assessments Undertaken in 2008

Upon applying the criteria to distinguish evaluations from reviews and other types of reports (included in Annex 3), the 13 reports<sup>32</sup> from 2008 period have been classified as follows:

Table 4: UNIFEM: Reports from 2008 Analysed in the Meta Evaluation

	<b>Name of Report</b>	<b>Thematic area</b>
<b>OTHER TYPES OF REPORTS</b>		
1	Assessment of Effectiveness of <b>IAWG Capacity Building</b> Project in Mainstreaming Gender in HIV/AIDS	HIV/AIDS
<b>REVIEWS</b>		
2	Capacity Building for Mainstreaming Gender in <b>HIV/AIDS Programming in the Caribbean</b>	HIV/AIDS
3	Evaluability of <b>UN Reform Process in Uruguay</b>	Cross-cutting
4	The UN contribution to the implementation of the <b>Paris Declaration</b>	Cross-cutting
<b>EVALUATIONS</b>		
5	Accountability for Protection of <b>Women's Human Rights (SEE)</b>	Governance
6	Advancing Gender Equality and Women's <b>Rights in Kosovo</b>	Governance
7	Evaluating the <b>Gender &amp; Governance Programme Kenya</b>	Governance
8	Social Communication, Political Participation and Democratic Governance for Women's Empowerment in Morocco	Governance?
9	<b>Functional Literacy</b> and Women's Empowerment in Morocco	Poverty?
10	<b>Engendering Budgets: Making Visible Women's Contributions to National Development in Latin America</b>	Poverty
11	Economic Empowerment of Women in Nigeria	Poverty
12	Days of Campaign against Violence against Women in Latin America and the Caribbean	Violence
13	Progressing the Implementation of <b>CEDAW</b> in Southeast Asia	Cross-cutting

In the section that follows, each assessment is summarised. Basic information on the evaluation subject is provided.<sup>33</sup> How the quality score provided for each evaluation below was determined is discussed in Chapter 3.

<sup>32</sup> Another 9-10 assessments have been reported to the Evaluation Unit as having been initiated in 2008 but not completed by February 2009.

<sup>33</sup> In the following sections, the information on the project/programme is derived from the assessment reports. It is therefore possible, depending on the quality of the reports, that the information is incomplete or erroneous.

## 2.1. 2008 Evaluations

### 1. Progressing the Implementation of CEDAW in Southeast Asia

<i>Implementation period evaluated:</i>	2004-2008
<i>Thematic area:</i>	Cross-cutting
<i>Region:</i>	SE Asia: Cambodia, Laos, Thailand, Vietnam, Timor-Leste, Indonesia, Philippines.
<i>Budget:</i>	\$6,806,999 (Canadian funding)
<i>Goal:/Activities:</i>	“CEDAW SEAP is a regional programme designed to facilitate the realisation of women’s human rights in SE Asia through more effective implementation of CEDAW.”

This evaluation is enlightening. From an evaluation perspective, it should be showcased within UNIFEM as a high quality evaluation that comes close to meeting every UNEG standard to the full. In terms of its results, it is a “must-read” for staff concerned with CEDAW implementation. Not only does it uncover important findings, the report offers reflections and deliberations that go beyond meeting the demands of the terms of reference. The report makes excellent use of meaningful text boxes. Its discussions on institutional learning, capacity development, results-based management and sustainability are applicable to much of UNIFEM’s work. The report is well-written, well-structured and nicely presented. The evaluation criteria used in the assessment include relevance, effectiveness and sustainability. Performance with regard to these criteria is also analysed. The annexes should have been re-submitted with the final report.

*Quality score: 43 (Excellent)*

### 2. Advancing Gender Equality and Women’s Rights in Kosovo

<i>Implementation period evaluated:</i>	1999-2008
<i>Thematic area:</i>	Governance
<i>Region:</i>	CEE-CIS, Kosovo
<i>Budget:</i>	\$1,468,137.55
<i>Goal:/Activities:</i>	“To mainstream gender in the emerging political and legal frameworks, structures and institutions in Kosovo; develop capacity to analyse budgets from a gender perspective, especially at a local level; and, to link partners and advocacy in Kosovo with networks and advocacy at the sub-regional level.”

This is a solid evaluation. It could have provided more information on the Kosovar context, in particular the situation of gender equality and women’s rights. Furthermore, more information on the project would have been desired. For instance, how much did the project cost? How were the funds allocated? Nevertheless, the findings, analysis, conclusions and recommendations are of high quality. The report provides constructive reflections on the concept of sustainability, results-based management, institutional learning and partnership with government.

*Quality score: 38 (Good)*



### 3. Evaluating the Gender & Governance Programme Kenya

<i>Implementation period evaluated:</i>	2006-2008
<i>Thematic area:</i>	Governance
<i>Region:</i>	Africa: Kenya
<i>Budget:</i>	\$ 4,987,947
<i>Goal:/Activities:</i>	Aims to ensure focused, coordinated, and long-term support towards women's enjoyment of human rights and participation in democratic governance. It is managed by UNIFEM; implemented by 30 organisations; and, governed by a donor steering committee of 8 bilateral agencies.

This is a very ambitious study. The evaluators have used several methodological tools and seem to have gathered a wealth of data. Evidently, the evaluators faced a difficult assignment given: i) the number of donors and other stakeholders involved; ii) erratic results of an over-ambitious programme; and, iii) inconsistent programme logic. The resulting report is heavy and long. The style of the report is chatty at times, but bold typeface helps guide the reader. The overall objective and structure of the programme are not clearly conveyed. The executive summary mainly consists of key findings and recommendations in table form. It is shame that such a data-rich report does not present its recommendations in a more clear and concise manner. Regarding the data gathering tools, the evaluation used 3 separate questionnaires. It is however not clear what the response rate to the questionnaire was and there is no discussion on the accuracy of the surveys undertaken. One questionnaire has 72 questions, excluding sub-questions, many of which require qualitative answers. This is more than what many stakeholders can be expected to handle. The evaluation criteria used in the assessment include relevance, effectiveness, efficiency and sustainability. Management performance is also assessed.

*Quality score: 29 (Average)*

### 4. Accountability for Protection of Women's Human Rights (SEE)

<i>Implementation period evaluated:</i>	2006-2009
<i>Thematic area:</i>	Governance
<i>Region:</i>	South-eastern Europe: Bosnia-Herzegovina, Serbia, Montenegro, Kosovo
<i>Budget:</i>	\$ ? (Canadian funding)
<i>Goal:/Activities:</i>	Aims to advance gender equality and women's human rights in line with international agreements, covenants, and commitments.

This is a mid-term evaluation of a regional project. It could qualify as a review, given the timing of the assessment (mid-term – after only a couple of years of implementation). However, the external evaluator went beyond focusing on operational and performance issues, triangulated data and addressed relevance, effectiveness, efficiency, sustainability, coherence and impact. A questionnaire was used with a respectable 60% response rate. Thirty 30 stakeholders were interviewed.

*Quality score: 33 (Good)*

## **5. Functional Literacy and Women's Empowerment in Morocco**

<i>Implementation period evaluated:</i>	2006-2008?
<i>Thematic area:</i>	EVAW? Poverty?
<i>Region:</i>	Africa: Morocco
<i>Budget:</i>	\$ ? (USAID funded)
<i>Goal:/Activities:</i>	Aims to empower women through provision of functional literacy courses to women which also inform about the Moroccan family law.

The report states that its goal was to examine and assess results; observe whether the results produced lead to sustainable changes; and, propose ways to improve future efforts. Sustainability, efficiency, relevance and performance have not been addressed. The evaluation gathered extensive qualitative data from 88 beneficiaries using five different tools. Relying on a baseline produced at the start of the project, the evaluation is able to provide concrete evidence of change in a systematic way. The report provides scant information about how the project was managed and its background. There are three brief sentences in the executive summary which constitute recommendations. Meanwhile, the conclusions and recommendations chapter does not include recommendations, but instead a list of questions asking how the training was conducted. These are supposedly directed at the project managers but could have been expected to have been answered by the evaluation itself! The terms of reference for the evaluation are not included. It is unclear if the report was conducted ex-post. The project has been classified as related to UNIFEM's thematic area of eliminating violence against women. While violence against women is addressed in the literacy training, it would seem that the project is much more relevance to the area of reducing the feminisation of poverty – where poverty is understood from a holistic perspective that includes a state of lacking power.

*Quality score: 27 (Average)*

## **6. Engendering Budgets: Making Visible Women's Contributions to National Development in Latin America**

<i>Implementation period evaluated:</i>	2005-2007
<i>Thematic area:</i>	Poverty
<i>Region:</i>	LAC: Argentina, Brazil, Bolivia, Honduras, Peru, Peru.
<i>Budget:</i>	\$ US 678,500 (\$313,000 provided by UNV)
<i>Goal:/Activities:</i>	The project's aims include to institutionalise women's participation in political and economic decision making and accountability process; and, enhance capacities for applying gender analysis in the planning, monitoring and evaluation of revenue and budgetary allocations at national, sub-national and local levels.

This evaluation is unique in that it seems to be the only ex-post evaluation undertaken in the period since 2004. While its English translation is not without fault, more bothersome is the use of evaluation pages to discuss the team's relational difficulties with UNIFEM. Such matters should be kept out of the evaluation. Confusing for the reader is that the first few

pages of the evaluation discuss a project that was eventually not included in the assessment. The overall objective of the programme and the budget are not clearly conveyed. A more comprehensive analysis of the regional approach of the project would have been welcome. The evaluation criteria used in the assessment include relevance, effectiveness, efficiency and sustainability. Management performance is also assessed. The report strongly criticises the lack of monitoring and evaluation systems. The point is hammered home many times but unlike the Kosovo and CEDAW-SEAP evaluations, the evaluators do not provide constructive reflections and ideas of how this could be undertaken.

*Quality score: 27 (Average)*

## **7. Social Communication, Political Participation and Democratic Governance for Women’s Empowerment in Morocco**

<i>Implementation period evaluated:</i>	2006-2008
<i>Thematic area:</i>	EVAW? Governance?
<i>Region:</i>	Africa: Morocco
<i>Budget:</i>	\$ ?
<i>Goal:/Activities:</i>	Improve coverage of sensitisation and communication; improve the political representation of women in parliament and enhance gender analysis in government administration?

This evaluation was the most incomplete in terms of content – there is no executive summary, no terms of reference, no tools, no lists of persons interviewed or documents reviewed. No contextual analysis is provided, the methodology is insufficiently described. The structure of the report and a mix of fonts make the report somewhat challenging to read. Its strengths lies in the evidence gathered in relation to the advocacy activities addressing rural populations, women involved in micro-credit and the public at large – which was assisted by the existence of the programme’s monitoring systems that provided a wealth of data. The other part of the evaluation, which covered the programme component of mainstreaming gender in the public sector, contains less analysis. The programme is classified contributing to UNIFEM’s thematic area of eliminating violence against women. Violence against women is an important theme, but it would seem that classifying this programme as contributing to the goal of achieving gender equality in governance is more accurate.

*Quality score: 24 (Weak)*

## **8. Economic Empowerment of Women in Nigeria**

<i>Implementation period evaluated:</i>	2006-2008
<i>Thematic area:</i>	Poverty
<i>Region:</i>	Africa: Nigeria
<i>Budget:</i>	\$ 275,000
<i>Goal:/Activities:</i>	Aimed to improve access to information for economic decision-making for women in 3 communities through increased use of ICT facilities, training, farming equipment and facilitation in the community.

This evaluation looked at results, effectiveness of strategies used and impact on beneficiaries. The evaluation used appreciative inquiry as its methodological approach. While this was useful in assessing the effectiveness of different strategies used, the approach focuses on successes and therefore does not look at whether every planned output and outcome have been achieved. Although there are some very useful recommendations, particularly for replicating the project in other Niger Delta communities, the recommendations section is somewhat weakened by including data and conclusions mixed up with the recommendations. There is an ethical problem with the report since the names of beneficiaries are included next to each quote.

*Quality score: 21 (Weak)*

### **9. 16 Days of Campaign against Violence against Women in the Andean Region**

<i>Implementation period evaluated:</i>	2004-2007
<i>Thematic area:</i>	EVAW
<i>Region:</i>	LAC: Ecuador, Venezuela, Bolivia, Colombia, Peru
<i>Budget:</i>	\$ 144,000? \$ 104,307?
<i>Goal:/Activities:</i>	Raising awareness about gender-based violence as a human rights issue at the local, national, regional levels to eliminate violent and discriminatory practices against women.

This evaluation looks at efficiency, effectiveness, relevance and sustainability. One weakness of the evaluation is that there is no section that describes the context, background and content of the campaign. The evaluation is unique in that it quantifies the responses from stakeholders during interviews and provides clear graphics to illustrate them. However, the validity of addressing evaluation criteria by asking stakeholders to provide their assessment is questionable in terms of ensuring triangulated findings. The concluding section is somewhat weak.

*Quality score: 25 (Average)*

## **2.2. 2008 Reviews**

Three assessments are classified as reviews:

### **1. Capacity Building or Mainstreaming Gender in HIV/AIDS Programming in the Caribbean:**

<i>Implementation period evaluated:</i>	2006-2007
<i>Thematic area:</i>	HIV/AIDS
<i>Region:</i>	LAC: Caribbean
<i>Budget:</i>	\$ ?
<i>Goal:/Activities:</i>	“To support the development and implementation of training programmes in gender analysis and HIV and AIDS for policy makers and programmers engaged in HUIV and AIDS work within regional development

institutions, including the donor community and at national level.”

This report is misleadingly called an “Impact Evaluation Assessment”: it does not look at impact (as defined by the international development community) and is not an evaluation. The study is relatively short and has a limited scope. It attempts to look at effectiveness (although the term is not used) but other evaluation criteria are not applied. The focus of the evaluation is on operational training issues. A questionnaire was used to survey trainees, but the information was not quantified. The evidence gathered does not seem to be triangulated but relies on the perspectives of the trainees who responded to the questionnaire. The terminology is often sloppily used (“gender” being used as a substitute for “women”; while “gender equality” is not used at all) At best the report is a review.

*Quality score: 13 (Very weak)*

## 2. The Evaluation of the UN contribution to the Implementation of the **Paris Declaration**.

<i>Implementation period evaluated:</i>	2007-2010
<i>Thematic area:</i>	Cross-cutting
<i>Region:</i>	Global
<i>Budget:</i>	\$ ?
<i>Goal:/Activities:</i>	The UN system’s implementation of the Paris Declaration which involves “efforts to harmonise, alignment and managing aid for results with a set of monitorable actions and indicators.”

This report is of high standard. While called an evaluation, it is a review in nature. It has been undertaken during the on-going effort to implement the Paris Declaration. The Report is geared towards providing ideas and guidance on how to improve on-going implementation. The reflections made focus on the immediate term. While effectiveness is addressed to some extent, no other evaluation criteria are used. No lessons are drawn that contribute to general knowledge and that can be applied to different contexts.

*Quality score: 39 (Good)*

## 3. The Evaluability of **UN Reform Process in Uruguay**

<i>Implementation period evaluated:</i>	2007-2010
<i>Thematic area:</i>	Cross-cutting
<i>Region:</i>	LAC: Uruguay
<i>Budget:</i>	\$ 95,5 million (entire One UN Programme budget)
<i>Goal:/Activities:</i>	In line with the General Assembly resolution 59/250), rationalise UN country presence, pilot the establishment of “One UN in Uruguay, with One Leader, One Programme, One Budget, and if appropriate, One Office.”

This assessment was commissioned by UNEG. It is the first of a three-phased approach to evaluating the pilot initiatives for Delivering as One, namely: (i) an assessment of the evaluability of Delivering as One (DaO); (ii) an independent process evaluation of the pilot experience to be completed by September 2009; and (iii) an independent evaluation of the results and impacts of the pilot experience by September 2011. The report assesses evaluability as opposed to evaluation criteria. Because it analyses operational aspects and performance, it can be regarded as a type of review. No lessons are drawn. While the report includes some recommendations, they are embedded in the text.

*Quality score: 34 (Good).*

### **2.3. Neither an evaluation nor a review**

The *Assessment of Effectiveness of IAWG Capacity Building Project in Mainstreaming Gender in HIV/AIDS* is neither an evaluation nor a review:

<i>Implementation period evaluated:</i>	2004-2008?
<i>Thematic area:</i>	HIV/AIDS
<i>Region:</i>	LAC: Caribbean sub-region
<i>Budget:</i>	\$ ?
<i>Goal:/Activities:</i>	“To strengthen inter-agency collaboration regionally and nationally in HIV/AIDS programme design, implementation and monitoring from the gender perspective.”

Most of its 10 pages consist of text in bullet form. A questionnaire was undertaken, but the data gathered was not quantified. The context is not provided and the subject is barely described. Effectiveness is the only criteria looked at. Stakeholders are referred to by first name and the report is virtually impossible to understand for the uninitiated.

### 3 Mapping of Evaluation Characteristics

The evaluator has mapped 8 aspects of the 9 evaluations. These include:

10. The **subject** of the evaluation (project, programme, thematic cluster or policy/strategy)
11. The **thematic area** that the evaluation falls into
12. The **timing** of the evaluation
13. The **region** the evaluation was undertaken in
14. The **type** of evaluation team (internal, external or mixed)
15. Who **managed** the evaluation
16. The **length** of the evaluation process
17. The **report content** of the evaluation reports

The following sections provide a quantitative analyse of the evaluations' characteristics. It is important to note that the number of evaluations from 2008 is relatively small, particularly considering the range of thematic areas, regions and number of countries in which UNIFEM has projects and programmes. Furthermore, the 13 reports originally submitted as evaluations only represent about half of the assessments initiated in 2008 – the others have yet to be finalised.

#### 3.1. Subject

In 2008, there were 5 project evaluations, 3 evaluations covering programmes and one evaluation covering a campaign. No policy/strategy, joint or thematic cluster evaluations undertaken.

#### 3.2. Thematic areas

Three evaluations undertaken were related to the thematic area of *increasing gender justice in democratic governance in stable and fragile states* (hereinafter referred to as “governance”). Two evaluations relate to *enhancing women's economic security and rights* (hereinafter referred to as “poverty”). One evaluation relates to CEDAW and is classified as “cross-cutting”. There were no evaluations related to halting the spread of HIV/AIDS among women and girls (hereinafter referred to as “HIV/AIDS”)

Three evaluations are reported as related to *reducing the prevalence of violence against women* (hereinafter referred to as ERAW). Among the latter evaluations, the classification of being related to ERAW is questionable for two of them. While eliminating violence against women cross-cuts both projects, given the focus and activities undertaken, the *Functional Literacy and Women's Empowerment in Morocco* would be more appropriately classified as relating to enhancing women's economic security and rights;<sup>34</sup> while *Social Communication, Political Participation and Democratic Governance for Women's Empowerment in Morocco* is strongly related to governance.

---

<sup>34</sup> Where poverty is understood in a holistic rights-based since and includes the concept of disempowerment.

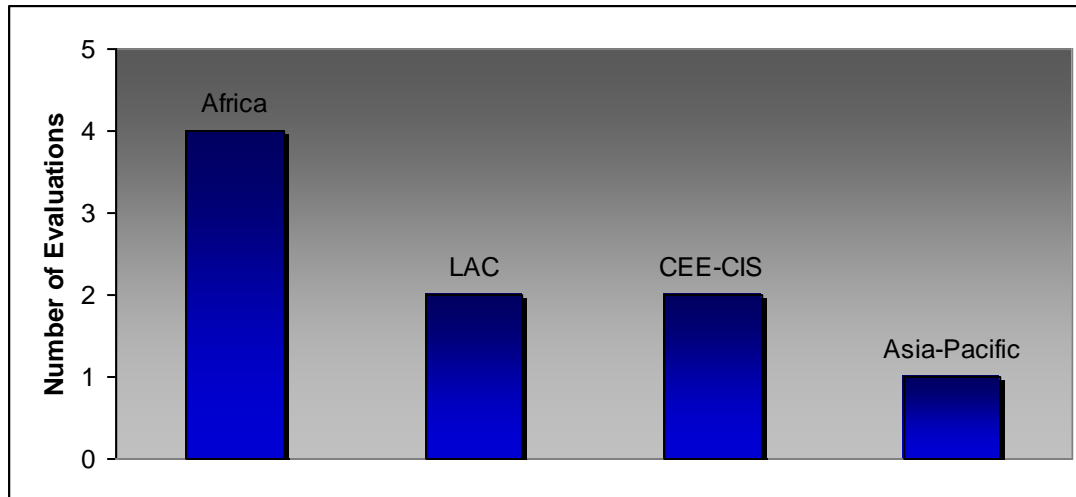
### 3.3. Timing

Seven of the evaluations were conducted at the end or towards the end of the programme/project period. One evaluation was conducted at mid-term. The evaluation, *Engendering Budgets: Making Visible Women's Contributions to National Development in Latin America* was an ex-post evaluation, the only one of its kind in the period 2004 to 2008.

### 3.4. Region

The 9 evaluations cover all of UNIFEM's four regions. Of the African evaluations, all four were focused on a single country – Kenya, Nigeria and two evaluations covering Moroccan projects. The evaluations from the Asia-Pacific and LAC regions were regional in scope. One evaluation from CEE-CIS covered one country (Kosovo), while the other covered several countries in the Balkans (Bosnia-Herzegovina, Serbia, Montenegro, Kosovo). There were no global evaluations.

Figure 10: Number of Evaluations by Region



### 3.5. Team

All of the evaluations were conducted by external evaluators. Four of the evaluations were conducted by teams of consultants.

### 3.6. Management

Six of the evaluations were commissioned and managed by UNIFEM's sub-regional or country offices. One evaluation was managed by the Geo Section in headquarters, the project office in Kosovo and regional office (Bratislava), with added support from the Evaluation Unit. The Evaluation Unit managed one evaluation. The management of the Kenyan evaluation was undertaken by UNIFEM's regional office jointly with the Embassy of Norway.



### 3.7. Length of evaluation process

It appears that two evaluations took between 6 to 12 weeks and two evaluations were completed in 3 to six months. It was unclear how long the process lasted for the remaining 5 evaluations.

### 3.8. Report content

According to the UNEG standards, an evaluation report should include a table of contents, an executive summary and a list of abbreviations. The annexes should include the terms of reference, the list of persons consulted/interviewed, list of documents reviewed and the methodological framework and/or data collection tools used.

Three of the reports managed to include all of these in the report. Two reports substantially lacked this type of content.

Table 5: Contents of the Evaluation Reports

Evaluation title	ToC	Exesum	Acronyms	ToR	Interviewees	Documents	Tools
Functional Literacy and Women's <b>Empowerment in Morocco</b>	Yes	Yes	Yes	No	No	Yes	Yes
<b>Engendering Budgets: Making Visible Women's Contributions to National Development in Latin America</b>	Yes	Yes	Yes	Yes	No	Yes	Yes
Advancing Gender Equality and Women's <b>Rights in Kosovo</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Accountability for Protection of <b>Women's Human Rights (Balkan)</b>	Yes	Yes	Yes	No	Yes	Yes	Yes
Social Communication, Pol.Participation and Dem. Governance for Women's <b>Empowerment in Morocco</b>	Yes	No	No	No	No	No	No
<b>16 Days of Campaign</b> against Violence against Women in LAC	Yes	No	No	No	No	No	No
Evaluating the <b>Gender &amp; Governance Programme Kenya</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Progressing the Implementation of <b>CEDAW</b> in Southeast Asia	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Economic Empowerment of Women in <b>Nigeria</b>	Yes	Yes	Yes	No	No	Yes	No

## 4 Assessment of Evaluation Quality

The quality of the evaluations was assessed using 16 different parameters that have been largely derived from the UNEG standards. For 11 of the parameters, a score between 0 and 3 was awarded. For the parameters “findings”, “analysis” and “conclusions”; a scale of 0 to 5 was used since these areas are relatively more important than the other parameters. For the remaining 3 parameters – i) the inclusion of safeguards, ii) debriefing of stakeholders and/or ensuring a validation process at the end of the evaluation; and, iii) the existence of a management response to the evaluation – a score of 0 or 1 was given.

The table in Annex 2 of the UNIFEM Meta-Evaluation provided the framework for the assessment. It specifies how scores for each parameter were awarded for each evaluation (the criteria are also included in boxes in most of the sections). The scores were aggregated to obtain a “quality score”. The highest attainable score is 48 points. The 2008 evaluations received a total average score of 30 points. However, it is important to note that the number of evaluations conducted – nine in all – **is too small to make definitive conclusions** about evaluation trends or to make any cross tabulations.

It is also important to view the score for each evaluation as not only pertaining to the evaluators, but also the commissioner(s) and manager(s) of the evaluation. If the terms of reference of the evaluation are weak, unclear and not aligned with the UNEG standards, the resulting evaluation is not likely to score well. Likewise, during the evaluation process, the manager(s) of the evaluation are responsible for guiding the evaluation to live up the UNEG standards.

Using multiples of 8, a grade range was established as follows:

**Very weak:** 16 points or under (0 evaluation)

**Weak:** 17 to 24 points (0 evaluations)

**Average:** 25 to 32 points (5 evaluations)

**Good:** 33 to 40 points (1 evaluation)

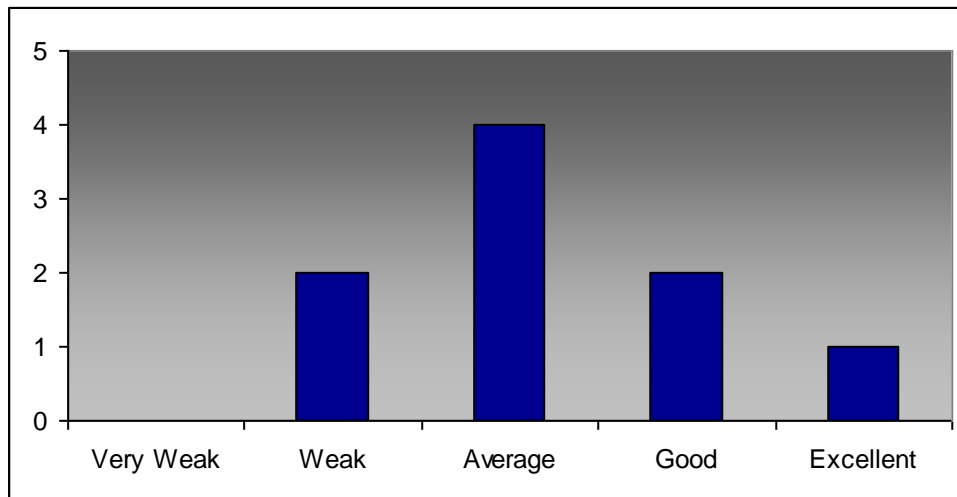
**Excellent:** 41 or more (1 evaluation)

*Table 6: List of 2008 Evaluations Scored*

	<b>Evaluation Title</b>	<b>Score</b>
1	Progressing the Implementation of <b>CEDAW</b> in Southeast Asia	43 (excellent)
2	Advancing Gender Equality and Women’s <b>Rights in Kosovo</b>	38 (good)
3	Accountability for Protection of <b>Women’s Human Rights (SEE)</b>	33 (good)
4	Evaluating the <b>Gender &amp; Governance Programme Kenya</b>	29 (average)
5	<b>Engendering Budgets:</b> Making Visible Women’s Contributions to National Development in Latin America	27 (average)
6	Functional Literacy and Women’s <b>Empowerment in Morocco</b>	27(average)
7	<b>16 Days of Campaign</b> against Violence against Women in the Andean Region	25 (average)
8	Social Communication, Pol. Participation and Dem. Governance for Women’s <b>Empowerment in Morocco</b>	24 (weak)
9	Economic Empowerment of Women in <b>Nigeria</b>	21(weak)

The graph below illustrates how the scored evaluations fared when given the corresponding grade:

Figure 11: Average Evaluation Score for Evaluation Conducted in 2008



#### 4.1. Basic key information

The basic key information that the title page (or the first few pages) of an evaluation report should include are the subject (i.e. activity, programme, policy etc.) being evaluated; name of the evaluators; name of the organisation(s) that commissioned the evaluation; and the date. In all cases, the title pages of evaluation reports contain this information. However, in some cases it became clear who commissioned the report only after opening the document. The average quality score was **3.0**.

#### 4.2. Executive summary

Two-thirds of the evaluations had an executive summary. Of those that did, under half scored full points. The average score out of 3.0 for the quality of the executive summary was **1.4**.

#### 4.3. Evaluation purpose, objectives, scope and methodology

Three evaluations scored full points for explaining the purpose, objectives and scope of the evaluation and providing a clear account of the methodology applied. Typically, evaluations failed to explain the limitations of the evaluation methodology and process. The average score for the presentation of the purpose, objectives, scope and methodology was **2.2**.

#### 4.4. Context of subject

Surprisingly, only one evaluation scored full points for providing an analytical account of contextual background (institutional, political, socio-economic, cultural, geographical and/or any other relevant context information) that contributed to the utility and accuracy of the evaluation. This includes providing relevant statistics to support the contextual analysis. Two evaluations provided no contextual information at all, and six evaluations did not

provide an adequate account of the context. The average score for the contextual analysis was **1.7**.

#### **4.5. Description of the subject**

One evaluation clearly described the subject being evaluated. The other evaluations did not provide sufficient information. The majority did not provide budget figures for the project/programme being assessed and some did not provide information how the project was organised or managed. Another common weakness of many of the evaluations – including some of the strong evaluations – was the failure to clearly state the overall goal of the project/programme. Furthermore, they tended not to provide detailed information on the project/programme’s stakeholders and their respective roles and contributions. They also often failed to explain the expected results chain and intended impact of the programme/project being evaluated. The average score for describing the evaluation subject was **1.9**.

#### **4.6. Findings**

The quality of the findings was scored on a scale from 0 to 5. Three evaluations were successful in presenting findings that are based on rich data. In these evaluations there is evidence of different types of data collected from a range of sources. Outputs, outcomes and possible impact were systematically covered. Multiplier or downstream effects were also addressed. Separate findings were also presented that related to performance. Four evaluations scored 3 or under because they were not able to present findings so that outputs, outcomes and possible impacts were systematically covered. Among these was the ex-post evaluation, which according to the report, was constrained by the inadequate goal formulations of the programme. Three of these evaluations did not present any findings relating to performance, making it was impossible to understand the structure and management of the project/programme. The average score was **3.8**.

#### **4.7. Analysis**

The quality of the analysis was scored on a scale from 0 to 5. The three evaluations with quality findings also presented structured, logical and measured analysis. These reports provide insight into the enabling and constraining factors – both internal and external to the subject. Possible threats and opportunities to build on in the future are expanded upon. Meanwhile, most of the evaluations with weaker findings were not able to present an analysis of high quality. However, one evaluation received a slightly better analysis score than findings score because the report presented a structured analysis that relatively systematically looked into internal and external enabling and constringing factors. Another evaluation had based its findings on a variety of data from several sources but did not approach the analysis in a sufficiently structured way. The average score for analysis is **3.8**.

#### **4.8. Conclusions**

The quality of the conclusions was scored on a scale from 0 to 5. While the conclusions of an evaluation are often the part that a reader remembers best, only one evaluation scored full points for conclusions. Six evaluations scored 3 or less. In most cases, these evaluations did not adequately tie together the findings and analysis, sum up the main points and lay the groundwork for the recommendations that are made later in the report. Furthermore, they did

not provide substantiated conclusions on the evaluation criteria assessed. The average score for the quality of the conclusions was **3.2**.

#### **4.9. Recommendations**

Three evaluations scored full points for the quality of the recommendations. One evaluation did not include any recommendations at all. A common problem of the recommendations was that they were not clearly and concisely formulated and presented. This would include explaining who the recommendation is directed at, what needs to be done, how and when. The evaluation should also prioritise recommendation to the extent possible. Sometimes recommendations were followed by extensive reasoning why the recommendation is needed, instead of ensuring that the findings and conclusions preceding the recommendation contained this reasoning. Recommendations were sometimes not numbered, making it difficult for end-users and management responses to refer to them. **2.0**.

#### **4.10. Lessons learnt**

Not all evaluations generate lessons. However, if it is the case that no lessons were generated, it should be stated in the report. Only two evaluations provided adequate lessons learnt. One evaluation provided poor lessons learnt and the remaining six assessments provided no lessons at all. As a result, the average score for the quality of the lessons provided was low, at **0.8**.

#### **4.11. Gender equality and human rights**

Since virtually all evaluations concern gender equality and therefore discuss gender issues in one way or another, all evaluations scored at least one point in this area. Six evaluations to some extent had a rights perspective, used rights-based language and/or to a limited extent analysed results in relation to the human rights framework. These include the CEDAW-SEAP evaluation, Kosovo GE&WR, Kenya G&G, Kosovo GE& WR, SEE-WR and the two evaluations in Morocco. All six of these evaluations deal directly with women's rights, CEDAW and/or governance. However, none of the evaluations fully lived up to UNEG's current standards in this area. The average score for considering gender equality and human rights issues was medium: **1.7**.

#### **4.12. Stakeholder participation**

All the evaluations tended to be more or less consultative. The 4 evaluations that scored "2s", used focus groups, questionnaires and involved a large number of stakeholders. No evaluation scored a "3". The average score was **1.6**.

#### **4.13. Clear communication**

Just under half of the evaluations used clear, precise and professional language and earned full points in this area. The others scored "2s". A couple of the evaluations made good use of graphs, boxes, illustrations and/or tables. The average score was **2.4**.

#### **4.14. Ethical safeguards**

None of the evaluations included a discussion of the extent to which the evaluation design included ethical safeguards where appropriate. This would have included protection of the confidentiality, dignity, rights and welfare of human subjects, including children, and respect

for the values of the beneficiary communities. One evaluation report has an ethical problem since the names of beneficiaries are included next to each quote.

#### **4.15. Validation process**

An evaluation, particularly one that purports to be consultative, should include a verification and validation process with key stakeholders. This provides an opportunity to correct factual errors and confirm the validity of conclusions. Such a process is likely to augment the quality of the workshop and its future usefulness. It can take the form of workshop(s), seminar(s), e-conference(s) etc. before the evaluation team drafts the evaluation report. From the information provided in the reports, 3 evaluations undertook some form of validation process with stakeholders before the report was finalised.

#### **4.16. Management response**

Evaluation requires an explicit response by the management and governing bodies addressed by its recommendations. There is a management response to only 1 evaluation. This can partly be explained by the fact that the evaluations were undertaken relatively recently.

#### **4.17. Comparisons with previous years**

The quality score of the evaluation from 2004 to 2007, the 2008 aggregate score is the same – 30 points or “average”. Nevertheless, as mentioned before, the number of evaluations completed by UNIFEM each year and gathered by the Evaluation Unit is relatively small. As a result, an evaluation of high or low quality greatly impacts on the overall score for the evaluations of that year. Trends are therefore not easily detectable. A cautious supposition may be that quality is making small steps in the right direction since none of the evaluations from 2008 scored “very weak”. Judging from the aggregate score and given the greater emphasis on a rights approach in UNIFEM, it might also be possible to say that the consciousness of integrating human rights perspectives into evaluations and ensuring more consultative evaluation processes have improved slightly.

## Annex 6 List of documents consulted

### UNIFEM Reports

1. **16 Days Campaign** against Violence against Women in the Andean Region (2008)
2. Accountability for Protection of **Women's Human Rights** (2008)
3. Advancing Gender Equality and Women's **Rights in Kosovo** (2008)
4. Advancing Gender Equality and Women's **Rights in Kosovo** (2008)
5. Arab Women Parliamentarian Project (2007 Review)
6. Assessment of Effectiveness of **IAWG Capacity Building** Project in Mainstreaming Gender in HIV/AIDS (2008)
7. Capacity Building for Mainstreaming Gender in **HIV/AIDS Programming in the Caribbean** (2008)
8. Capacity Building **Workshop** for Mainstreaming Gender Analysis in HIV/AIDS Programming in the Caribbean (2006)
9. Capacity Building **Workshop** for Mainstreaming Gender Analysis in HIV/AIDS Programming in the Caribbean (2006)
10. CIDA-UNIFEM Accountability for Protection of **Women's Human Rights** (2008)
11. Days of Campaign against Violence against Women in Latin America and the Caribbean (2008)
12. Economic Empowerment of Women in Nigeria (2008)
13. Elimination of Violence Against Women in **Central Asia** (2004)
14. Elimination of Violence Against Women in **Central Asia (2004)**
15. Empowering women in global production networks - NTFP India **Gum Karaya** Project (2007)
16. Empowering women in global production networks - NTFP India **Gum Karaya** Project (2007)
17. **Engendering Budgets: Making Visible Women's Contributions to National Development in Latin America** (2008)
18. **Engendering Budgets: Making Visible Women's Contributions to National Development in Latin America** (2008)
19. Engendering Peace Building and Mediation Processes in **Burundi** (2004)
20. Evaluability of **UN Reform Process in Uruguay** (2008)
21. Evaluating the **Gender & Governance Programme Kenya** (2008)
22. Evaluating the **Gender & Governance Programme Kenya** (2008)

23. Evaluation of the Southeastern European Legal Initiative (**SEELINE**) (2004)
24. Evaluation of UNIFEM's **MYFF** System (2007)
25. From **Global to Local**: A Convention Monitoring and Implementation Project (2004)
26. Functional Literacy and Women's **Empowerment in Morocco** (2008)
27. **Functional Literacy** and Women's Empowerment in Morocco (2008)
28. Gender and Economic Justice in **European Accession** and Integration (2004)
29. Gender and the **Millennium Development** Goals (2007)
30. **IGAD Gender Strategy** in the Context of Globalisation and the Feminisation of Poverty (2005)
31. Mainstreaming Gender in Poverty Reduction: **IFAD-UNIFEM** Gender Mainstreaming Programme in Asia (2007)
32. Mainstreaming Gender into the **Electoral Process in DRC** (2007)
33. Prevention of **Trafficking** in Women and Children in South Asia (2005)
34. Progressing the Implementation of **CEDAW** in Southeast Asia (2008)
35. Progressing the Implementation of **CEDAW** in Southeast Asia (2008)
36. Progressing the Implementation of **CEDAW** in the Pacific (2006)
37. Promoting Women's Human Rights and Enhancing Gender Equality in Democratic **Governance in Kenya** (2006)
38. Promotion of Women's Human Rights through the Elimination of **Violence against Women** SE Asia (2004)
39. Promotion of Women's Human Rights through the Elimination of **Violence against Women** SE Asia (2004)
40. Reinventing India: Action for Empowerment to Address **Violence against Women**
41. **Savings Mobilisation** & Productive Investment Pilot in Iligan & Jagna: Regional Programme on Empowering Women Migrant Workers in Asia (2005)
42. **Savings Mobilisation** & Productive Investment Pilot in Iligan & Jagna: Regional Programme on Empowering Women Migrant Workers in Asia (2005)
43. Social Communication, Political Participation and Democratic Governance for Women's **Empowerment in Morocco** (2008)
44. Social Communication, Political Participation and Democratic Governance for Women's Empowerment in Morocco (2008)
45. Stock-Taking of the **Railways-UNIFEM Project** on Equalizing Gender Relationship in the Context of the HIV/AIDS Epidemic (2005 Review)



46. Supporting Women's Leadership in **Rebuilding Afghanistan** (2007)
47. The UN contribution to the implementation of the **Paris Declaration** (2008)
48. UNIFEM Peace & Security Programme in **Sierra Leone** (2005)
49. UNIFEM Peace & Security Programme in **Sierra Leone** (2005)
50. UNIFEM Programming in **HIV/AIDS** (2006)
51. UNIFEM Evaluation Strategy
52. Women Economic Empowerment Project in **Nigeria** (2008)
53. Women for Conflict Prevention & Peace Building in the **Southern Caucasus, Phase 1** (2004)
54. Women for Conflict Prevention & Peace Building in the **Southern Caucasus, Phase 2** (2007)
55. Women's Rights to Land in **Kyrgyzstan** (2006)

## Other Documents

1. Dale, Reidar. *Evaluating Development Programmes and Projects*. Second Edition. Sage Publications, 2004
2. Danida. *Evaluation Guidelines*. Ministry of Foreign Affairs, 1999
3. Davidson, E. Jane. *Evaluation Methodology Basics: the Nuts and Bolts of Sound Evaluation*. Sage Publications, 2005
4. Earl, Sarah, Fred Carden and Terry Smutylo. *Outcome Mapping. Building Learning and Reflection into Development Programmes*. IDRC, 2001
5. Ljungman, Cecilia. "A Rights-Based Approach to Development", in Mikkelsen *Methods for Development Work and Research A New Guide for Practitioners*, Second Edition, Sage Publications 2005.
6. Mikkelsen; Britha. *Methods for Development Work and Research - a New Guide for Practitioners*, Chapters 3,7 and 8. Sage Publications 2005
7. Scrivens, Michael. *Key Evaluation Checklist*: [www.wmich.edu/evalctr/checklists](http://www.wmich.edu/evalctr/checklists)
8. Sida. *Looking Back, Moving Forward*. Sida evaluation Manual 2<sup>nd</sup> Revised Edition. Sida 2007
9. Stufflebeam, Daniel L. "Evaluation checklists: practical tools for guiding and judging evaluations". *American Journal of Evaluation*, 22(1), 71-80
10. UNEG "Ethical Guidelines for Evaluation" 2007
11. UNEG "Norms for Evaluation in the UN System", 2005

12. UNEG “Standards for Evaluation in the UN System” 2005
13. Universalia, “IUCN Meta-Evaluation An Analysis of IUCN Evaluations 1994 – 2000” IUCN, 2000
14. Wiles, Peter “ALNAP Meta-Evaluation”. ALNAP 2004

## Annex 7 List of Persons interviewed

Coello, Raquel	Programme Technical Coordinator, Quito, Ecuador
Ishikawa, Shoko	Programme Manager, CEDAW South East Asia Programme, Bangkok, Thailand
Jusaeva, Sagipa	Programme Specialist, Almaty, Kazakhstan
Marcelino, Elena	Programme Specialist, UNIFEM HQ
Miriti, Linette	Programme Specialist, Planning and Monitoring & Evaluation, Nairobi, Kenya
Rojas, Katrina	Team Leader HIV/AIDS evaluation
Sandler, Joanne	Deputy Director, UNIFEM HQ
Sanz, Belen	Head, Evaluation Unit, UNIFEM HQ
Wenderoth, Anette	Team leader, Kosovo GE&WR; CEDAW-SEAP, team member MYFF evaluation



