



Development Effectiveness Review of
the United Nations Entity for Gender
Equality and the Empowerment of
Women (UN Women)

2011 – 2014

24 April 2015

Preface

This review provides an independent and evidence-based assessment of the effectiveness of the programs supported by United Nations Entity for Gender Equality and the Empowerment of Women (UN Women). The approach and methodology to reviewing multilateral organizations was developed under the guidance of the Development Assistance Committee Network on Development Evaluation (DAC-EVALNET) of the Organization for Economic Co-operation and Development (OECD). It complements the periodic assessments done by the Multilateral Organization Performance Assessment Network (MOPAN) of the multilateral organizations' operational effectiveness. By focusing on development effectiveness and carefully selecting assessment criteria, the reviews seek to avoid duplication or overlap with the MOPAN process.

The review was planned in consultation with the Independent Evaluation Office of UN Women and led by the Policy and Operations Evaluation Department (IOB) of the Netherlands Ministry of Foreign Affairs. It relies on the content of evaluation reports produced by UN Women supplemented with a review of UN Women corporate documents and consultation with staff at UN Women headquarters in New York. The methodology applied is a structured meta-synthesis of the content of a purposive sample of 55 evaluations that is illustrative of UN Women programming. The analysis applied a common set of assessment criteria derived from the DAC evaluation criteria.

The review was conducted by a team from the consulting firm, Goss Gilroy Inc. (Hubert Paulmer, Bruce Goodman and Tasha Truant). The team is grateful to UN Women for its helpfulness and useful, practical advice. Senior Policy Researcher Kirsten Mastwijk coordinated the review on behalf of IOB.

An international advisory group chaired by Senior evaluator Paul de Nooijer from IOB has provided advice on the methodology and the adequacy of evidence in support of conclusions.

The content of this report does not necessarily reflect the official positions, policies or views of the OECD or the Netherlands Ministry of Foreign Affairs. The report is designed for use by UN Women, its Governing Board, the members of the DAC-EVALNET and other interested stakeholders.

We would like to thank everyone who contributed to this review for their valued input and support.

Geert Geut
Interim Director, Policy and Operations Evaluation Department (IOB)
Netherlands Ministry of Foreign Affairs

Table of Contents

Preface	i
Table of Contents	ii
Executive Summary	iv
Acronyms	xiii
1.0 Introduction	1
1.1 Background	1
1.2 Purpose	1
1.3 UN Women: A Global Organization Dedicated to Gender Equality and Empowerment of Women	2
1.4 Evaluation and Results Reporting at UN Women	4
2.0 Methodology	6
2.1 Evaluation Population and Sample	7
2.2 Assessment Criteria	7
2.3 Limitations	8
3.0 Findings on UN Women’s Development Effectiveness	9
3.1 Relevance	12
3.2 Achievement of Objectives	15
3.3 Cross-Cutting Themes	21
3.4 Sustainability	25
3.5 Efficiency	30
3.6 Use of Monitoring and Evaluation	35
4.0 Conclusions	40
5.0 Lessons	43
Annex 1: Development Effectiveness Criteria	45
Annex 2: Evaluation Sample	46

Annex 3: Methodology.....	50
Annex 4: Comparison of Population and the Sample	54
Annex 5: Evaluation Quality Scoring Guide	56
Annex 6: Comparison Between Recent and Initial Years	59
Annex 7: Guide for Review Team to Classify Evaluation Findings...61	
Annex 8: UN Women Corporate Documents.....	70

Figures

Figure 1: UN Women’s Program Expenditures, by Regions and Headquarters, 2011 - 2014.....	3
Figure 2: Coverage and Summary Results for Each Sub-criterion.....	10
Figure 3: Number of Evaluations Addressing Sub-criteria for Relevance	12
Figure 4: Findings for Relevance.....	13
Figure 5: Number of Evaluations Addressing Sub-criteria for Achievement of Objectives	16
Figure 6: Findings for Achievement of Objectives.....	17
Figure 7: Number of Evaluations Addressing Sub-criteria for Cross-Cutting Themes	22
Figure 8: Findings for Effectiveness in Supporting Gender Equality.....	23
Figure 9: Number of Evaluations Addressing Sub-criteria for Sustainability	26
Figure 10: Findings for Sustainability	27
Figure 11: Number of Evaluations Addressing Sub-criteria for Efficiency	30
Figure 12: Findings on Efficiency	32
Figure 13: Number of Evaluations Addressing Sub-criteria for Using Evaluation and Monitoring ...	36
Figure 14: Findings for Using Evaluation and Monitoring.....	37

Executive Summary

Background

This report presents the results of a review of the development effectiveness of the programming supported by the United Nations Entity for Gender Equality and the Empowerment of Women (UN Women). The report utilizes a common approach and methodology developed under the guidance of the Organization for Economic Cooperation and Development's (OECD) Development Assistance Committee Network on Development Evaluation (DAC-EVALNET). The review relies on the content of published evaluation reports produced by UN Women, supplemented with a review of UN Women corporate documents and consultations with the Independent Evaluation Office (IEO) of UN Women.

The purpose of this approach is to work in a coordinated way with related initiatives including the DAC-EVALNET/United Nations Evaluation Group (UNEG) Peer Reviews of United Nations organization evaluation functions and the assessments carried out by the Multilateral Organizations Performance and Assessment Network (MOPAN). It also recognizes that multilateral organizations continue to make improvements and strengthen their reporting on development effectiveness and should eventually be providing regular, evidence-based, field-tested reporting on effectiveness themselves.

Purpose

The purpose of the review is to provide the donor community an independent, evidence-based assessment of the development effectiveness of UN Women's relevance and performance, and support the donor community's relationship with UN Women by identifying lessons for program and policy improvements.

Approach and Methodology

The review, carried out from October 2014 to March 2015, began with a preliminary review of UN Women documents and the identification of the population of UN Women evaluation reports found in UN Women Global Accountability and Tracking of Evaluation Use (GATE) System.¹ The results of the preliminary review indicated that a meta-analysis of the findings of recent evaluations published by UN Women could provide a useful overview of developmental effectiveness for the period 2011-2014 to generate lessons on what works to advance gender equality and women's empowerment.

In consultation with the IEO and the Netherlands Ministry of Foreign Affairs' Policy and Operations Evaluation Department (IOB) staff, the review team subsequently drew a sample of 61 UN Women evaluation reports, published between 2011 and 2014, that was illustrative of UN

¹ <http://gate.unwomen.org/>

Women programming. All corporate and thematic evaluations managed by the IEO, HQ divisions and the Regional Offices (RO) were included in the sampling frame. These evaluations inherently address issues of strategic significance on the agency's triple mandate.² Evaluations were selected to include development programming from the various regions, types of programming and from various years.

The quality of each evaluation was assessed against criteria derived from the accepted OECD/DAC evaluation criteria which themselves reflect the UNEG Norms and Standards for evaluation. Only six reports were excluded due to quality concerns. As a result, 55 evaluation reports were retained for systematic rating in the review.

Each evaluation report was rated, using a four-point scale that ranged from Highly Unsatisfactory to Highly Satisfactory, on six key Development Effectiveness Criteria and nineteen sub-criteria. Based on the findings from the individual evaluation reports, the review team identified factors contributing to both positive and negative findings for each of the six key criteria:

1. Relevance of Interventions
2. Achievement of Development Objectives and Expected Results
3. Cross-cutting Theme (Gender Equality and Environmental Sustainability)
4. Sustainability of Results/Benefits
5. Efficiency
6. Use of Evaluation and Monitoring to Improve Development Effectiveness

The results of the meta-synthesis of evaluation findings were then summarized and presented to UN Women staff prior to the development of this report. At the request of UN Women's IEO, a comparison was made between more recent evaluations (2014/2013) and those produced during the initial years of UN Women (2012/2011). Whenever possible, this distinction was included in the analysis of the assessment criteria.

Coverage of the Assessment Criteria in Reviewed Evaluations

The review established ranges for assessing how well each of the 19 sub-criteria were covered in the 55 evaluations, based on the number of evaluations that addressed each sub-criterion. Coverage could be strong, moderate or weak.³ Only one sub-criterion (the extent to which program supported changes are environmentally sustainable) was weak in coverage and consequently the findings were not reported.

² The unique role of UN Women reflects its strong mandate to work at both normative and operational levels as well as to ensure UN system-wide coordination. This requires an innovative evaluation function that leverages its partnerships and knowledge (UNW/2014/3).

³ Coverage has been reported as strong (if the number of evaluation addressing the sub-criterion is between 41 and 55), moderate (if the number of evaluations addressing the sub-criterion is between 26 and 40), and, weak (if the number of evaluation addressing the sub-criterion is 25 or less)

Limitation: Retrospective Nature of the Review

A review of development effectiveness relies mainly on published evaluation documents and is inherently retrospective, rather than forward-looking. As with any meta-synthesis, there are methodological challenges that argue for careful interpretation and use of the reported findings. In addition to the retrospective nature of a meta-synthesis, the limitations for this review include sampling bias and the challenge of assessing overall programming effectiveness when there are important variations in programming. Despite these limitations, careful selection of the purposive sample of evaluation reports and an intensive process of quality assurance applied during the analysis process give reasonable assurance that the findings reported below are broadly illustrative of the development results achieved by UN supported programming during the period under review.

Findings of the Development Effectiveness Review

Relevance of UN Women Programs

The relevance of UN Women supported programming is covered strongly in the evaluations and findings indicate that the agency's programming is highly relevant. The vast majority of evaluations reported high suitability to target group needs, strong alignment with national development goals and effective partnerships. Evaluations in recent years (2014/2013) as compared to initial years of UN Women (2012/2011) more often rated UN Women supported programming positively in this respect.

The alignment of UN Women's programming with government priorities was the major factor contributing to the relevance of programming. The programming was designed to support the implementation of public policies, supported programming and capacity building in national governments or institutions and was based on partnerships with key stakeholders. The relevance of UN Women's programming was also enhanced through needs assessments and special research of target groups.

Achievement of Objectives

Overall, the issue of objectives achievement had strong coverage and positive findings in the reviewed evaluations. Seventy five percent of the evaluations reported positive findings on the extent to which UN Women programs achieve their stated objectives. Positive results were better with reference to the programs' ability to provide positive benefits for the target population and to support positive changes in national policies and programs, with more than 90% evaluations reporting satisfactory or better. They were somewhat less positive on the number of program beneficiaries reached. When looking at the more recent evaluations, it stands out that they report positive results more often, regarding objectives achievement than those evaluations carried out in 2012/2011.

The ability of UN Women to recognize key actors and to work with them to address gender equality and women’s empowerment issues, most often to develop appropriate legislation, policies and frameworks, was a key factor contributing to the achievement of objectives. UN Women’s advocacy, its active lobbying and its role in building coalitions and networks also facilitated the achievement of program objectives. On the other hand, there were also various factors limiting the achievement of objectives, such as weaknesses in program design, often linked to over ambitious objectives, an unclear theory of change, lack of a results orientation, excessively short time frames and/or limited resources.

Gender Equality

Evaluations indicated strong coverage and reported positive findings regarding UN Women’s contribution to gender equality. Almost 80% evaluations that addressed gender equality reported findings of satisfactory or better.⁴ A key factor in these reported results is the mandate of UN Women to address gender equality and women’s empowerment, which has been well reflected in programs which align with gender equality priorities and apply human rights-based approaches. More recent evaluations reported considerably more positive findings than those produced earlier in relation to effectively addressing gender equality.

The technical expertise of UN Women in the field of gender equality was noted particularly in reference to the positive findings in the areas of addressing the issue of ending violence against women and mainstreaming gender-responsive budgeting. Success in addressing gender equality issues has also been facilitated by UN Women’s participatory approaches and inclusive processes. On the other hand, poor performance measurement frameworks and ill-defined indicators sometimes hindered measurement of how effectively supported programs address gender equality.

Sustainability

In general, the coverage of sustainability ranges from moderate to strong and the reported evaluation findings are mixed. More than 70% of the evaluations reflected positive ratings regarding UN Women’s contribution to strengthening the enabling environment for development and strengthening institutional or community capacity. However, largely negative results (reported by 67% of the evaluations) regarding the likely continuation of benefits after project/program completion pose a challenge for UN Women supported programming. It is worth noting that, in contrast to other evaluation criteria (relevance, achievement of objectives and gender equality), more recent evaluations report positive findings less often on the likelihood of benefits continuing after program completion than those conducted earlier. In contrast, the evaluations from recent years report positive findings for sustainability through institutional

⁴ The remaining 13% of evaluations did report on gender equality results; however, they failed to report if those results met the stated objectives and/or to what extent. Consequently, this led the review team to conclude that for these reports the criterion of gender equality was “not addressed”

and/or community capacity and programs strengthening the enabling environment for development and significantly more often than those conducted in 2012/2011.

The continued reliance by government and civil society partners on external program funding is a key limiting factor for sustainability. Furthermore, deep-rooted negative cultural values and widespread stigma, high staff turnover at UN Women and partner organizations, and the absence of a well-planned exit strategy are also factors that contribute to poor sustainability of program benefits. On the other hand, UN Women's participatory programming methodologies and its inclusive approach to strengthening capacity has aided institutional and community capacity and contributed to a so-called strengthened enabling environment.

Efficiency

The coverage of efficiency in the evaluations ranged from moderate to strong. The findings for efficiency are mixed but less often positive than for other criteria and require careful interpretation. At the same time, it should be noted that approximately 60% evaluations reported positive findings on the cost efficiency of programs despite a moderate coverage. Nonetheless, it is a concern that approximately 75% of the evaluations that addressed the timeliness and efficiency of administrative systems reported negative findings.

In contrast to other criteria (but in line with the findings for the sustainability criteria), the more recent evaluations are likely to show less positive findings than earlier evaluations for cost/resource efficiency, the timeliness of implementation and the efficiency of administrative systems and procedures. However, this difference between earlier and later evaluation findings may be less significant than it appears in these sub-criteria than the overall weakness of efficiency in UN Women program.

Good use of and ability to leverage limited resources are both seen as positive factors in terms of the efficiency of UN Women supported programs. The factors that detract from timeliness are related to UN Women's administrative systems and procedures, notably delays in the release of funds due to cumbersome procedures for funds disbursement and weak program implementation and follow-up procedures. Staff turnover, and internal management challenges, in addition to the transition from UNIFEM to UN Women, are other factors that hinder positive results in efficiency.

Using Evaluation and Monitoring to Improve Development Effectiveness

Coverage for the use of evaluation and monitoring is fairly strong but the results were mixed. The coverage for the effectiveness of monitoring and reporting systems and use of evaluation to improve development effectiveness was strong. On the other hand, coverage was only moderate for the effectiveness of both monitoring and reporting, and RBM systems.

Those evaluations which do report on the effectiveness of evaluation systems and processes and

the use of evaluation results to improve developmental effectiveness⁵ are most often positive. In the first instance (evaluation systems) 66% of evaluations report findings of satisfactory or better while for the second sub-criterion (use of evaluation results) the figure is 90%.

It is notable that the use of evaluations to improve program design is supported by an increasing tendency to prepare detailed management responses to evaluation reports and that these include action plans, assign management responsibility and include measures for reporting on the status of progress in responding to evaluation recommendations.

In contrast, more than 80% of evaluations present negative rating for the effectiveness of systems and processes for results monitoring and reporting and results-based management.⁶ These are the two sub-criteria with the smallest number of evaluations reporting positive results. This is a matter of concern despite recent efforts⁷ by UN Women and remains an area for improvement. Weaknesses in the design of results frameworks, and the use of inadequate or inappropriate indicators and inadequate baseline information, often undermined UN Women's ability to effectively monitor and report on results according to the reviewed evaluations. Reporting was often focused more on the completion of activities rather than on progress in achieving program results.

Recent evaluations more often report positive results than those conducted earlier regarding the effectiveness of evaluation, monitoring and reporting and result-based management systems, and the use of evaluation for improving effectiveness. This trend to more positive results over time is particularly evident for findings on the effectiveness of evaluation systems and processes.

Conclusions

- UN Women **programs are highly relevant** to the needs of target group members and are supportive of the development plans and priorities of program countries. UN Women has also had success in developing effective partnerships with government, NGOs, civil societies and media, especially in promoting gender equality and women's empowerment. The relevance of UN Women programming has also ensured alignment with key national

⁵ The management response to evaluations is separate documents (available online) which in most cases indicated actions taken

⁶ Evaluations findings were primarily based on primary and secondary data collected by the evaluation team when monitoring data was not available or inadequate. The lack of adequate monitoring clearly affects the ability to demonstrate progress or lack thereof against stated objectives. However, it does not necessarily affect the project's ability to achieve results

⁷ Recent efforts by UN Women to improve the evaluation function, monitoring and reporting, and RBM include: appointment of additional personnel (Director of IEO and regional M&E advisors reporting directly to the Director of IEO), the implementation of GERAAS, commissioning of meta-evaluation and meta-analysis report, developing a strategy for strengthening RBM in UN Women including corporate RBM training modules, and initiating the process of building a RBM system. However, it is still too early to assess whether these efforts will lead to improvements.

development plans and policy documents.

- UN Women programs have been **effective in achieving the objectives** of the development programs, in securing positive benefits for target group members and in supporting positive changes in national policies and programs. Whenever UN Women programs have done so, this has often been due to UN Women’s ability to recognize key actors and to work with them to address various gender equality and women’s empowerment issues and develop appropriate legislation, policies and frameworks. The achievement of objectives has also been facilitated by UN Women’s advocacy, its active lobbying, and its ability to convene and play an active role in building coalitions and networks. On the other hand, when UN Women programs failed to meet their objectives, it was reportedly most often due to weaknesses in project design, often linked to over ambitious objectives, unclear theory of change, lack of a results orientation, excessively short time frame and limited resources.
- Regarding the cross cutting themes, as expected UN Women’s performance with respect to **gender equality is highly effective**. This finding can clearly be linked to UN Women’s mandate, which is aligned with gender equality priorities and applying human rights-based approaches. Many evaluations addressed the issue of ending VAW and gender-budgeting and linked their positive findings to the technical expertise of UN Women in the field of gender equality. Success in this area has also been facilitated by UN Women’s participatory approach and inclusive processes.
- The **sustainability of UN Women supported program results is mixed**. The results for the likely continuation of benefits after project/program completion were negative and this poses a challenge for UN Women supported programming. A major factor explaining these findings is the continued reliance by government and civil society partners on external program funding. Deep-rooted negative cultural values and widespread stigma, high staff turnover at UN Women and partner organizations affecting continuity, and the absence of a clear and well-planned exit strategy are other factors that contribute to the relatively poor sustainability of program benefits. On the other hand, UN Women’s participatory programming methodologies and its inclusive approach to strengthening capacity have helped improve institutional and community capacity and contributed to an improved enabling environment.⁸
- The reported results for the **efficiency of UN Women programming are mixed at best**. This is clearly an area for improvement by UN Women. In spite of low coverage, the results for the cost-efficiency of programs were found to be positive and a key factor that contributed to satisfactory findings is the ability to “*achieve more with limited resources*” by leveraging resources and applying internal cost controls. Nonetheless, the absence of appropriate and timely availability of cost data in the program does not allow any reasonable efficiency calculations and/or monitoring. Primarily, administrative processes and/or delays

⁸ Appendix 7 – Guide for DER criteria – sub-criterion 4.3

in the release of funds affected the timeliness of program implementation; and, this was linked to weak systems and procedures for implementation and follow-up. Key factors further limiting efficiency included staff turnover, excessive bureaucratic process requirements and internal management challenges.

- UN Women programs have **positive results for evaluation systems and processes but negative for monitoring and reporting, and results-based management**. The use of evaluations to improve program design (and by extension development effectiveness) is supported by an increasing tendency to prepare detailed management responses to evaluation reports that include action plans, assign management responsibility and include measures for reporting on the status of progress in responding to evaluation recommendations. On the other hand, the findings for performance monitoring and reporting, and results-based management systems are predominantly negative. Weaknesses in the design of results frameworks, especially in the development of appropriate indicators and the measurement and reporting of baseline information often undermined UN Women's ability to effectively monitor and report on results according to the reviewed evaluations. Reporting was often focused more on the completion of activities rather than on progress or lack thereof in achieving program results. Despite UN Women's recent efforts, RBM, monitoring and reporting has remained an area of concern.
- **Recent evaluations (2014/2013) have shown more positive results** than those conducted earlier (2012/2011) in 14 of the 18 sub-criteria reported. More specifically, positive results are significantly higher for recent evaluations for programs and projects making a difference for substantial number of beneficiaries, effectively addressing gender equality, strengthening the enabling environment for sustainability and the effectiveness of systems and processes for evaluation. However, in relation to cost/resource efficiency, the timeliness of implementation and the efficiency of administrative systems and procedures, earlier evaluations report positive results slightly more often than recent evaluations. There are no clear explanations for this shift in these sub-criteria.

Lessons

Lessons for program and policy improvement based on evaluation reports reviewed for this DER, include:

- UN Women is respected as an authority in its strategic focus areas of gender equality and women's empowerment. Sustainable changes in UN Women's strategic areas will take several years with the result that a long term focus is required for both its strategy and programming.
- Many interventions show favourable signs of change; however, with lack of or inadequate monitoring systems, UN Women is not able to capture those results when they do happen. Results-based management is a systemic weakness in UN Women operations

at every level. Many of the programs lack explicit theories of change, measurable results frameworks or adequate monitoring. This is inhibiting both accountability and learning.

- Effective and innovative approaches to enhance inclusiveness in reaching out to women and marginalized groups have helped increase results in gender equality in UN Women supported programs. The approach to include men and religious leaders as active program participants facilitated programs' influence on changes in social and cultural norms, and increased the possibility to progress and realize rights and gender equality for women by creating an enabling environment.
- Strengthening the capacity of organizations at national and community level is one of the key characteristics of UN Women interventions. However, in many cases outcomes of the capacity development aspects of the program are not demonstrable due to weak result-based management and monitoring systems. While programs report on completion of activities or numbers of people trained, evidence is not often available on the impact of capacity development efforts.
- Lack of financial resources and gaps in the technical capacity of national institutions and the political context in a country are threats to the sustainability of the results of UN Women programming. Nonetheless, the commitment of individuals and participatory processes of UN programs have resulted in some examples of locally owned initiatives. Ownership increases when local institutions/stakeholders are involved as early as during the design stages of the program.
- The need for an exit strategy from UN Women program support was highlighted in many of the evaluations. This is an essential element to ensure sustainability and continuity of activities and benefits after program completion.
- In order to improve the reporting of cost-efficiency or value for money of UN Women programs and interventions, it is important to have process and systems in place to ensure tracking of cost data throughout the program/project implementation process.

Acronyms

CO	Country Office
CSO	Civil Society Organization
DAC	Development Assistance Committee
DER	Development Effectiveness Review
DFATD	Department of Foreign Affairs, Trade and Development, Canada
EVALNET	Network on Development Evaluation
GATE	Global Accountability and Tracking of Evaluation Use
GRB	Gender Responsive Budgeting
GERAAS	Global Evaluation Reports Assessment and Analysis System
HQ	Headquarters (UN Women, New York)
IEO	UN Women Independent Evaluation Office
IOB	Netherlands Ministry of Foreign Affairs' Policy and Operations Evaluation Department
M&E	Monitoring and Evaluation
MDG	Millennium Development Goal
MCO	Multi-Country Office
MDG	Millennium Development Goal
MO	Multilateral Organization
MOPAN	Multilateral Organization Performance Assessment Network
NGO	Non-Governmental Organization
OECD	The Organisation for Economic Cooperation and Development
RBM	Results-based Management
RO	Regional Office
UN	The United Nations
UNEG	United Nations Evaluation Group
UNFPA	United Nations Population Fund
UNICEF	United Nations Children's Fund
UNIFEM	United Nations Development Fund for Women
UN Women	UN Entity for Gender Equality and the Empowerment of Women
USD	United States Dollar
VAW	Violence Against Women

1.0 Introduction

1.1 Background

This report presents the results of the review of the development effectiveness of the programs supported by United Nations Entity for Gender Equality and the Empowerment of Women (UN Women). The report utilizes a common approach and methodology developed under the guidance of the Organization for Economic Cooperation and Development's (OECD) Development Assistance Committee Network on Development Evaluation (DAC-EVALNET). The review relies on the content of published evaluation reports produced by UN Women, supplemented with a review of UN Women corporate documents and consultations with the Independent Evaluation Office (IEO) of UN Women.

The study methodology uses a common set of assessment criteria derived from the DAC evaluation criteria (Annex 1). This method was pilot tested in 2010 using materials from Asian Development Bank and World Health Organization and since then has been used to conduct reviews of several multilateral organizations – United Nations Development Programme, World Food Programme, and United Nations Children's Fund (UNICEF).

Since its inception, the process of developing and implementing the review of development effectiveness has aimed to complement the work of the Multilateral Organization Performance Assessment Network (MOPAN). By focussing on development effectiveness and carefully selecting assessment criteria, the reviews seek to avoid duplication or overlap with the MOPAN process⁹.

1.2 Purpose

The objectives of the review are to:

- Provide the donor community an independent, evidence-based assessment of the UN Women's relevance and performance; and,
- Support the donor community's relationship with UN Women, by identifying lessons for program and policy improvements¹⁰.

⁹ From 2015, MOPAN aims to develop and implement an updated methodology that integrates the MOPAN and EvalNet approaches to assessing multilateral organizations' development results.

¹⁰ Terms of Reference – Joint DER of UNPFA, UNCHR and UN Women dated June 3, 2014

1.3 UN Women: A Global Organization Dedicated to Gender Equality and Empowerment of Women

1.3.1 Background and Objectives

In July 2010, the United Nations General Assembly created, by its resolution 64/289, the United Nations Entity for Gender Equality and the Empowerment of Women (UN Women). In doing so, UN Member States took an historic step in accelerating the achievement of the Organization's goals regarding gender equality and the empowerment of women.¹¹

UN Women was created by bringing together four previously distinct parts of the UN system which focused exclusively on gender equality and women's empowerment - the Division for the Advancement of Women (DAW); the International Research and Training Institute for the Advancement of Women (INSTRAW); the Office of the Special Adviser on Gender Issues and Advancement of Women (OSAGI); and, the United Nations Development Fund for Women (UNIFEM).¹²

Grounded in the vision of equality enshrined in the UN Charter, UN Women works for the: elimination of discrimination against women and girls; empowerment of women; and, achievement of equality between women and men as partners and beneficiaries of development, human rights, humanitarian action and peace and security.¹³

1.3.2 UN Women's Strategic Direction

The development results framework for the 2011 – 2013 UN Women Strategic Plan had six main goals that continue to be the focus of Strategic Plan 2014-2017. The first five include:¹⁴

- a) To increase women's leadership and participation in all areas that affect their lives;
- b) To increase women's access to economic empowerment and opportunities, especially for those who are most excluded;
- c) To prevent violence against women and girls and expand access to survivor services;
- d) To increase women's leadership in peace and security and humanitarian response;
- e) To strengthen the responsiveness of plans and budgets to gender equality at all levels.¹⁵

¹¹ UN Women - About us <http://www.unwomen.org/en/about-us/about-un-women>

¹² UN Women – About us <http://www.unwomen.org/en/about-us/about-un-women>

¹³ UN Women Strategic Plan 2011-2013 (UNW/2011/9) p.4.

¹⁴ Several international agreements guide the work of UN Women including – Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) and the Optional Protocol; Beijing Declaration and Platform of Action (PFA); UN Security Council Resolution 1325 (UNSCR 135) on Women, Peace and Security <http://www.unwomen.org/en/about-us/guiding-documents>

¹⁵ UN Women Strategic Plan 2011-2013; and Strategic Plan 2014-2017

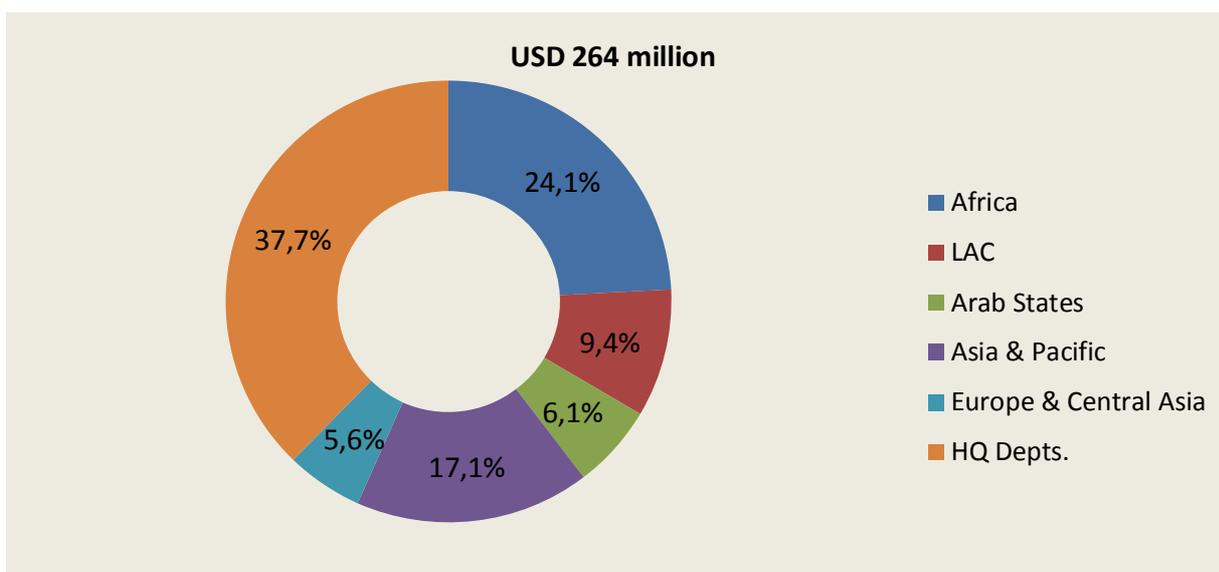
The sixth goal involves support for a comprehensive set of global norms, policies and standards on gender equality and women’s empowerment that is dynamic, responds to new and emerging issues, challenges and opportunities and provides a firm basis for action by Governments and other stakeholders at all levels.¹⁶

1.3.3 UN Women’s Operations and Program Expenditures

As of April 2013, UN Women worked through Regional offices (6), Multi-country offices (6) and Country Offices (48). UN Women is present in 76 countries. At the end of 2013, the agency had 657 staff (an increase of 18.8% over 2012), of which 244 positions were in headquarters (38%) and 413 were in the field (62%). About 75% of staff was female.¹⁷

Revenue and expenses for UN Women reported in 2013 were USD 288 million and USD 264 million respectively¹⁸. Figure 1 presents UN Women’s *program expenditures*¹⁹ by Region and headquarters (HQ) for 2013.²⁰

Figure 1: UN Women’s Program Expenditures, by Region and Headquarters, 2013



Source: UN Women Annual Report 2013-2014; UN Women Progress report 2014 (UNW/2014/2)

¹⁶ UN Women Strategic Plan 2011-2013 (UNW/2011/9) – May 9, 2011

¹⁷ Report on Progress made on the 2011-2013 Strategic Plan – April 2014

¹⁸ UN Women Annual Report 2013-2014

¹⁹ The HQ program expenditures relates to Policy and Programme Bureau, Bureau of Management, Bureau of Strategic Partnerships, Intergovernmental Support

²⁰ Report on Progress made on the 2011-2013 Strategic Plan (UNW/2014/2), p34.

1.4 Evaluation and Results Reporting at UN Women

This section presents a summary of the review team’s examination of UN Women’s evaluation function, policies and practices in place. This should not be viewed as an overall assessment of the evaluation function, policy and practices at UN Women.

1.4.1 Evaluation Policy and Function at UN Women

UN Women evaluations are either corporate or decentralized evaluations. Corporate evaluations are undertaken by the Independent Evaluation Office (IEO) with the support of independent external evaluators to assess issues of corporate strategic significance such as development effectiveness, organizational performance and operational coherence. Decentralized evaluations are conducted by independent external evaluators, but managed by program offices (Country Office, Multi-Country Office, and Regional Office). Corporate and decentralized evaluations may be either be thematic or program/project evaluations.

Evaluation Policy

The current evaluation policy of UN Women (UNW/2012/12), which came into force on January 1, 2013, governs the independent evaluation function and applies to all initiatives supported and funds administered by the organization. The policy is aligned with the norms and standards of the United Nations Evaluation Group (UNEG). At the same time, it supports the unique mandate and role of UN Women²¹ to conduct evaluations responsive to gender equality and women’s rights and to support coordination, coherence and accountability with respect to gender equality and the empowerment of women in the United Nations system.

The policy sets out the purpose of evaluation in UN Women and presents UN Women specific, tailored definition of evaluation and the principles and standards that guide its practice. It covers both the coordination role of UN Women with respect to United Nations system-wide evaluations concerning gender equality and the corporate and decentralized evaluation of the operational, normative support and coordination work undertaken by UN Women itself.

In addition to presenting the evaluation criteria to be used for the evaluations, the policy describes how the evaluation will be used. The policy outlines clear roles and responsibilities with respect to evaluation at all levels of UN Women.²²

²¹ The unique role of UN Women reflects its strong mandate to work at both normative and operational levels as well as to ensure system-wide coordination. This requires an innovative evaluation function that leverages its partnerships and knowledge (UNW/2014/3).

²² UN Women Evaluation Policy (UNW2012/12)

Structure and Organization of the Evaluation Function

The IEO is the custodian of UN Women's evaluation function, which is governed by the Evaluation Policy. The IEO reports directly to the UN Women Executive Director and presents an Annual Report on Evaluation during Executive Board sessions. The Evaluation Advisory Committee, established in December 2013, acts as a forum for the Executive Director and the IEO to further ensure the independence, relevance and quality of its evaluation function.

The IEO has developed a Strategic Evaluation Plan (2014-2017) that was approved by the Executive Director. This provides a results-based framework to strengthen the evaluation function at UN Women in the context of the Evaluation Policy and the UN Women 2014-2017 Strategic Plan.

The IEO continues to strengthen its capacity at headquarters and has nine staff members including five mid-level Professional staff and a consultant. A new post of Director of Evaluation was created in 2013, by upgrading the post of Evaluation Chief²³. The evaluation capacity at regional offices is also being strengthened by recruiting a Regional Evaluation Specialist (at P-4 level) in each region. These specialists report directly to the Director of the Independent Evaluation Office. Work is also in progress to strengthen evaluation capacity at country office and multi-country office level.

Quality Assurance

The UN Women publication *A Manager's Guide to Gender Equality and Human Rights Responsive Evaluation* provides the tools necessary for the management of all phases of the evaluation process: preparation, conduct and follow-up/use. The UNEG handbook *Integrating Human Rights and Gender Equality in Evaluation: Towards UNEG Guidance* is also a key reference for all UN Women evaluations.

All decentralized evaluations are quality-assured through the Global Evaluation Reports Assessment and Analysis System (GERAAS, 2013)²⁴ and results are presented at the annual session of the Executive Board through the Annual Report on Evaluation. The recently completed Peer Review (2014)²⁵ acknowledged the efforts of IEO to establish quality control

²³ Report of Evaluation Function of UN Women 2013

²⁴ It is important to note that GERAAS assesses the quality of reports (as a stand-alone document) against a very specific and prescriptive set of UNEG standards for what an evaluation report should look like. However, GERAAS does not measure the quality of evaluation as a process. It is designed to only assess the quality of reports, which is considered a major output of evaluations. The ratings provided are thus indicative of overall evaluation performance, but are not comprehensive and do not represent the actual use of evaluations in enhancing performance, learning and accountability in UN Women (UN Women GERAAS 2013, Meta Evaluation Report Version 1.1, p10).

²⁵ Professional Peer Review of the Evaluation Function of UN Women – Peer Review Panel Final Report – September 2014

systems at decentralized level to improve credibility, quality and use of evaluations. However, the Peer Review Panel noted that there is scope to mitigate the risk of weakened independence and to improve quality assurance at the decentralized level.

1.4.2 Reporting Results and Development Effectiveness

The Peer Review Panel (2014) identified significant strengths in the evaluation systems (centralized and decentralized) in UN Women; and has recommended that this should be fostered to support the goal of developing culture of evaluation (learning and accountability) in the organization.²⁶

Overall, UN Women has shown a consistent effort to improve and report on development effectiveness. Additionally, it is worth noting that results' reporting has been strengthened during 2013 and 2014 (see Annex 6), with the use of DAC criteria to synthesize the lessons learned on development effectiveness from UN Women evaluation reports.²⁷

Improvements in UN Women's reporting and programming processes were acknowledged in the *Executive Summary* of its most recent MOPAN report (2014)²⁸, "UN Women has adopted good practices that are likely to enable contributions to development results". The MOPAN report also reported that UN Women has developed a focus on results and has set up relevant performance management systems, which is likely to facilitate development effectiveness in the future. Findings from MOPAN 2014 summary report indicates that UN Women's work is relevant and is making contributions to various impact areas (strongest in three out of six UN Women impact areas – ending violence against women, women's leadership in peace and security, and advancing global norms and standards on gender equality).²⁹

2.0 Methodology

Based on the results of a preliminary review of evaluation and results reporting in UN Women as summarized above, the review team proceeded to conduct a systematic analysis and synthesis of the findings presented in a sample of UN Women evaluation reports. This chapter briefly presents the methodology used. A more detailed description of the approach and methodology is presented in Annex 3.

²⁶ Professional Peer Review of the Evaluation Function of UN Women – Final Report – September 2014

²⁷ The first report was published in 2014 – Meta-Analysis- WHAT CAN WE LEARN FROM UN WOMEN EVALUATIONS?

²⁸ MOPAN 2014 – Synthesis Report – UN Women - Executive Summary, 2015

²⁹ MOPAN 2014 – Synthesis Report – UN Women – Executive Summary, 2015, p.6

2.1 Evaluation Population and Sample

The sampling process began with an extraction of UN Women evaluations published between 2011 and 2014 as reported on the organization's on-line evaluation database.³⁰ This list was augmented with evaluations identified during discussions with the UN Women Independent Evaluation Office (IEO). As a result, the review team identified 79 evaluations published by UN Women for the period 2011 – 2014: the evaluation universe for this review.

The list of 79 evaluation reports³¹ was further scrutinized to arrive at the final sampling frame of 73 reports by excluding formative evaluations and/or strategic reviews, evaluations that did not have the final report uploaded and reports that explicitly referred only to UNIFEM, rather than UN Women. Reports pertaining to meta-evaluation and meta-analysis were also excluded.³²

Based on inputs and feedback from IEO and Netherlands Ministry of Foreign Affairs' Policy and Operations Evaluation Department (IOB), all corporate and thematic evaluations managed by IEO, HQ divisions and the Regional Offices (RO) were included in the sampling frame. These evaluations by their nature address issues of strategic significance on the triple mandate³³ of UN Women and their inclusion in both the qualitative and quantitative review conducted by the team is expected to enrich and provide a broader perspective of development effectiveness.

Twelve mid-term evaluation reports were then excluded from the sampling frame to arrive at an overall sample of 61 reports for review. This was done so that the evaluations reviewed would all deal with projects that could reasonably be expected to have achieved their stated goals and produced meaningful development results. This sample was further reduced, after the 61 reports were assessed for quality by the review team. Six reports were rejected from the initial sample, as they did not meet the minimum required scores. This left a final sample of 55 reports that were systematically reviewed with findings coded and used for quantitative analysis by the review team (Annex 2).

2.2 Assessment Criteria

The development effectiveness review (DER) methodology does not rely on a particular definition of development effectiveness but rather focuses on essential characteristics of effective multilateral organization programming, derived from the DAC evaluation criteria. These include:

³⁰ <http://gate.unwomen.org/index.html>

³¹ List of Evaluation Reports as of November 25, 2014 in the Evaluation database

³² This was to avoid double counting of results.

³³ The unique role of UN Women reflects its strong mandate to work at both normative and operational levels as well as to ensure system-wide coordination. This requires an innovative evaluation function that leverages its partnerships and knowledge (UNW/2014/3).

-
- Programming activities and outputs are relevant to the needs of the target group and its members;
 - Programming contributes to the achievement of development objectives and expected development results at the national and local level in developing countries (including positive impacts for target group members);
 - The benefits experienced by target group members and the development (and humanitarian) results achieved are sustainable in the future;
 - Programming is delivered in a cost efficient manner;
 - Programming is inclusive in that it supports gender equality and is environmentally sustainable (thereby not compromising the development prospects in the future); and,
 - Programming enables effective development by allowing participating and supporting organizations to learn from experience and the use of performance management and accountability tools, such as evaluation and monitoring to improve effectiveness over time.

The review methodology, therefore, involves a systematic and structured review of the findings of UN Women evaluations, as they relate to six main criteria (Box 1) and 19 sub-criteria that are considered essential elements of effective development and humanitarian programming (Annex 1).

Box 1: Assessment Criteria

1. Relevance of Interventions
2. Achievement of Development Objectives and Expected Results
3. Cross Cutting Themes (Gender Equality and Environmental Sustainability)
4. Sustainability of Results/Benefits
5. Efficiency
6. Using Evaluation and Monitoring to Improve Development Effectiveness

The results reported in Chapter 3 provide UN Women with a reflection of how its own evaluation reports have characterized its development effectiveness over the period from 2011 to 2014. Additionally, as requested by IEO, the review team has also analyzed and compared the results between the recent years (2014, 2013) and formative years of UN Women (2012, 2011), where feasible.

2.3 Limitations

As with any meta-synthesis, there are methodological challenges that argue for careful

interpretation and use of the reported findings. For this review, the limitations include sampling bias, the challenge of assessing overall programming effectiveness when there are important variations in programming (for example, evaluations covering multiple focus areas or only a specific theme or a project) and the retrospective nature of a meta-synthesis (Annex 3).

Limitations specific to the review of UN Women Development Effectiveness include the following:

- The sample of evaluation reports is illustrative rather than statistically representative of UN Women programming;
- The review was not able to report systematically on the effectiveness of UN Women’s programming by focus area or by type of country. There were not sufficient evaluations in each focus area or each country classification included in the meta-synthesis to allow for them to be analyzed separately;
- The 55 evaluation reports, while able to attain at least the minimum quality assessment scores, vary considerable in quality; and,
- The results reported apply to programs implemented as recently as 2013 or as early as 2007 (Annex 4).

Nonetheless, the review of development results reported in Chapter 3 has pointed to at least two key positive strengths:

- They cover a substantial body of UN Women programming;³⁴ and,
- They reflect the consistent application of a common set of evaluation criteria during which the analytical team strove to ensure the findings reported reflect the judgement of the original evaluation report authors, and not that of the review team (training and quality assurance processes are described in detail in Annex 3).

3.0 Findings on UN Women’s Development Effectiveness

This section presents the results of the review as they relate to the six main development effectiveness criteria and their associated sub-criteria. Figure 2 sums up the findings with respect to coverage (the review team’s assessment of whether the coverage is strong, moderate or weak) and then the results for each sub-criterion – including the quantitative proportion of evaluations

³⁴ See Annex 4. Furthermore, the value of programming covered by the evaluations reviewed was USD 329.6 million, as reported in 49 of 55 evaluations (including joint programs). There was no value of programming reported in 6 evaluations.

reporting findings of “satisfactory” or “highly satisfactory” for each of the six major criteria and their associated sub-criteria. It is important to note that the percentages provided are for only those evaluations, which addressed a given sub-criterion. The “n” in column two of Figure 2 represents the denominator used to calculate the percentage of evaluation findings, which were classified as highly unsatisfactory, unsatisfactory, satisfactory or highly satisfactory.

Box 2: Frequency of Observations
 Most = over 75% of the evaluations for which the sub-criterion was covered
 Many = between 51% and 75%
 Some = between 20% and 50%
 Few = less than 20%

In addition, this Chapter includes an overview of the factors that either contributed to or limited the results reported by the evaluations for each criterion. To be consistent the report makes use of terms “most”, “many”, “some” and “few” to describe frequency with which an observation was noted, as a percentage of the number of evaluations addressing the sub-criterion (Box 2). Additionally, the order in which the factors are presented reflects the frequency with which they are mentioned.

In the more in-depth reporting of the rated findings for each criterion, the report presents first the extent to which each review sub-criterion was addressed in the evaluation reports (coverage) and then the results for each sub-criterion – including both the quantitative findings and the results of the qualitative analysis of the factors contributing to or hindering the achievement.

Figure 2: Coverage and Summary Results for Each Sub-criterion

Criteria and Sub-Criteria	n*	Coverage Level‡	Satisfactory Ratings (%) †	Unsatisfactory Ratings (%) †
1.Relevance				
1.1 Programs and projects are suited to the needs and/or priorities of the target group.	51	Strong	88%	12%
1.2 Projects and programs align with national humanitarian and development goals.	45	Strong	98%	2%
1.3 Effective partnerships with governments, bilateral and multilateral development and humanitarian organizations and non-governmental organizations for planning, coordination and implementation of support to development and/or emergency preparedness, humanitarian relief and rehabilitation efforts.	51	Strong	73%	27%
2. Achievement of Objectives				
2.1 Programs and projects achieve their stated humanitarian and development objectives and attain expected results.	47	Strong	74%	26%
2.2 Programs and projects have resulted in positive benefits for target group members.	53	Strong	96%	4%
2.3 Programs and projects made differences for a substantial number of beneficiaries and where appropriate contributed to national humanitarian and	26	Moderate	58%	42%

Criteria and Sub-Criteria	n*	Coverage Level‡	Satisfactory Ratings (%) †	Unsatisfactory Ratings (%) †
development goals.				
2.4 Programs contributed to significant changes in national humanitarian and development policies and programs (including for disaster preparedness, emergency response and rehabilitation) (policy impacts) and/or to needed system reforms.	44	Strong	91%	9%
3.Cross-Cutting Themes – Inclusive Development which is Sustainable				
3.1 Extent to which multilateral organization supported activities effectively address the cross-cutting issue of gender equality.	48	Strong	79%	21%
3.2 Extent to which changes are environmentally sustainable.	2	Weak	–	–
4.Sustainability of Results/Benefits				
4.1 Benefits continuing or likely to continue after project or program completion or there are effective measures to link the humanitarian relief operations, to rehabilitation, reconstructions and, eventually, to longer term humanitarian and development results.	48	Strong	33%	67%
4.2 Projects and programs are reported as sustainable in terms of institutional and/or community capacity.	49	Strong	71%	29%
4.3 Programming contributes to strengthening the enabling environment for humanitarian and development.	36	Moderate	78%	22%
5.Efficiency				
5.1 Program activities are evaluated as cost/resource efficient.	33	Moderate	61%	39%
5.2 Implementation and objectives achieved on time (given the context, in the case of humanitarian programming).	40	Moderate	33%	67%
5.3 Systems and procedures for project/program implementation and follow up are efficient (including systems for engaging staff, procuring project inputs, disbursing payment, logistical arrangements etc.).	51	Strong	35%	65%
6.Using Evaluation and Monitoring to Improve Development Effectiveness				
6.1 Systems and process for evaluation are effective.	40	Moderate	65%	35%
6.2 Systems and processes for monitoring and reporting on program results are effective.	51	Strong	14%	86%
6.3 Results based management systems are effective.	40	Moderate	13%	87%
6.4 Evaluation is used to improve humanitarian and development effectiveness.	55	Strong	91%	9%
* n = number of evaluations addressing the given sub-criterion ‡ Strong: n=41– 55; Moderate: n= 26– 40; Weak: n = 25 or less † Satisfactory ratings includes “satisfactory” and “highly satisfactory”; unsatisfactory ratings includes “unsatisfactory” and “highly unsatisfactory”				

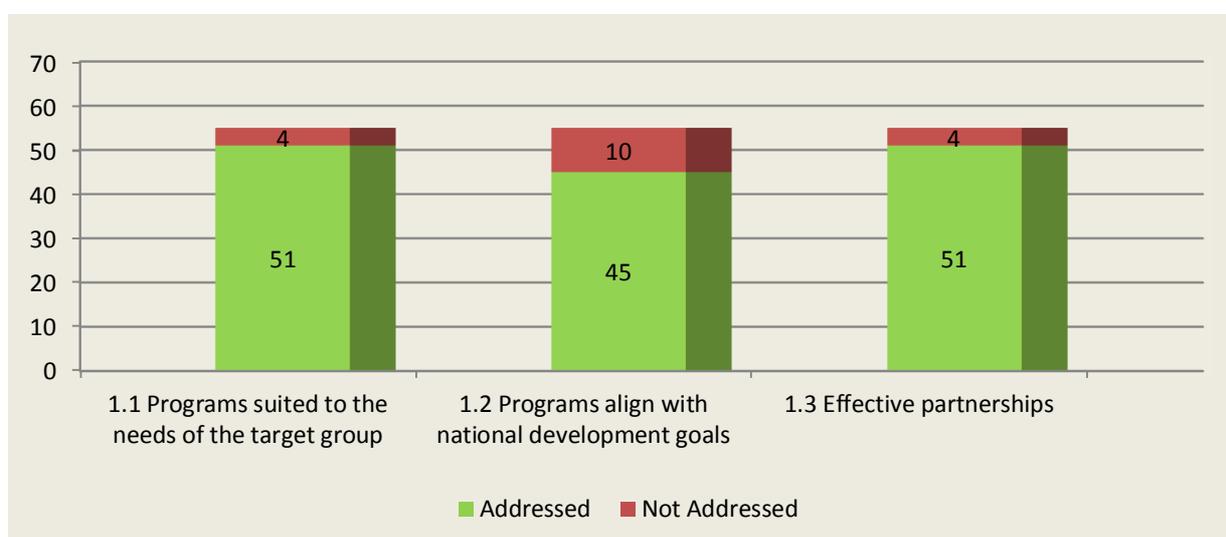
Source: DE Review Team

3.1 Relevance

3.1.1 Coverage

There is strong coverage of all three sub-criteria with respect to relevance (Figure 3). About 93% (51/55) of the evaluations covered programs suited to the needs of the target group (sub-criterion 1.1) and effective partnerships (sub-criterion 1.3). Only marginally fewer of evaluations (88%) address programs aligned with national development goals (sub-criterion 1.2). Coverage is higher in recent evaluations than earlier evaluations for all three sub-criteria (Annex 6).

Figure 3: Number of Evaluations Addressing Sub-criteria for Relevance



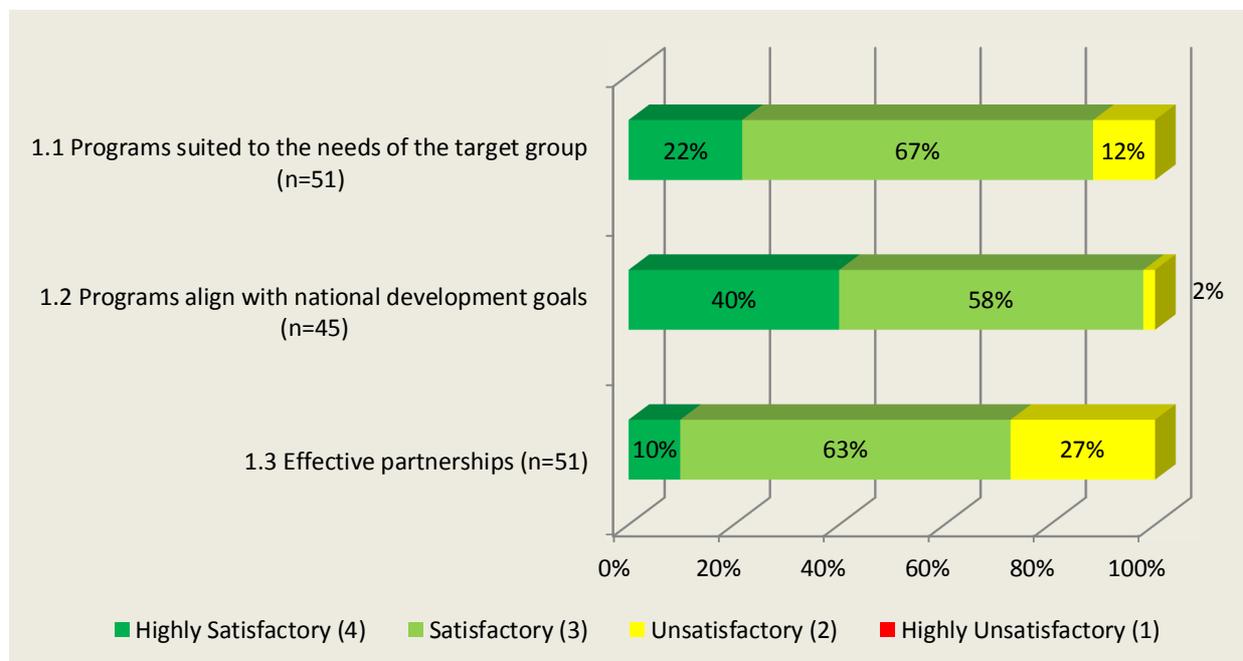
3.1.2 Key Findings

The findings indicate that UN Women programming is highly relevant. Most of the evaluations (89% - 45/51) reported satisfactory or better for the suitability of UN Women programs to the needs of the target group (sub-criterion 1.1). Almost all (98%) of the 45 evaluations, that addressed the sub-criterion, reported alignment with national development goals (sub-criterion 1.2). With regard to UN Women programs developing effective partnerships (sub-criterion 1.3), 73% (37/51) of the evaluations reported satisfactory or better findings. None of the evaluations reported highly unsatisfactory in the three sub-criteria (Figure 4).

Evaluations in recent years (2014, 2013) reported higher satisfactory or better ratings than the evaluations from the initial years of UN Women (2012, 2011) in all three sub-criteria (1.1, 1.2 and 1.3), and more specifically in suitability to the needs of the target group (93% and 83% respectively) and developing effective partnerships (79% and 65% respectively) – see Annex 6. It should be noted, that new structures and ways of working have been established in 2012 and

2013, including major decentralization³⁵ and increased capacity in the field,³⁶ in addition to the Evaluation Policy of UN Women coming into effect in January 2013.³⁷

Figure 4: Findings for Relevance



Evaluations identified a wide range of partners that were involved in UN Women programming. These include national governments, civil societies (CSOs) and non-governmental organizations (NGOs), other UN agencies, international organizations, bilateral donors and media.

Specific examples of relevance from the evaluation reports include:

- Implementation of the project (Mehwar Centre Project), within the framework of Ministry of Women Affairs' nine-year National Strategic Plan, to combat violence against women. It was directly relevant to the second and third goals of the National Plan (Palestine);
- Alignment of the project to national priorities and the existing international commitments of the Governments of the countries under Convention to Eliminate all forms of Discrimination Against Women (CEDAW). It was also relevant for the two countries in their effort to fulfill the requirements to become part for European Union – obligations with reference to gender equality (Serbia and Montenegro);

³⁵ For example, Multi-country offices (MCOs) were established in 2012.

³⁶ Progress made on UN Women Strategic Plan 2011-2013 – Executive Director's Report April 2014 (UNW/2014/2)

³⁷ UN Women Evaluation Policy, October 2012 (UNW/2012/12) p.1

-
- Identifying and integrating the needs of the target group into the program through a participatory planning workshop in which representatives from organizations for young women participated (Brazil);
 - Implementation of the project was in partnership with Ministry of Municipality and Public Works and the State Ministry of Women Affairs. Partnerships from earlier projects helped in project design (Iraq); and,
 - Successful coordination of various agencies including government agencies, private sector (foreign recruitment agencies) and development partners in designing and implementation of program (Nepal).

3.1.3 Contributing Factors

The evaluations identified a number of factors either contributing to or detracting from the relevance of UN Women interventions. In addition to alignment with national development goals (sub-criterion 1.2) such as national plans, national programs, policies and action plans and development frameworks, a few evaluations also pointed out to alignment of UN Women supported program with regional priorities as a strong positive factor in achieving relevant development effectiveness.

Positive factors contributing to relevance noted by many evaluations include:

- The competence of UN Women in establishing effective (strategic) partnership with responsible partners (with a focus on gender equality and women's empowerment), for implementation – including government ministries and agencies, NGOs and CSOs other UN agencies/donors and media (32 evaluations); and,
- Effective targeting of a wide range of beneficiaries by UN Women programs (23 evaluations). Examples of groups and organizations targeted by UN Women supported programs include - human rights organizations, human rights advocates (including women holding political positions), national women's institutions / agencies, relevant ministries, internally displaced persons, war victims, indigenous women, organizations/networks working with women living with HIV, women's centres, and, community groups, in addition to women and girls in general.

Some evaluations noted:

- The effective use of needs assessment and / or other diagnostic exercises including, situation analysis, livelihood assessments, risk assessments, socio-economic and conflict assessments (17 evaluations). Effective use of recommendations from earlier evaluations and inputs from stakeholder was also noted (also see sub-criterion 6.4); and,

-
- The use of participatory approaches and inclusive processes for program planning and implementation, and, selection of beneficiaries increased the relevance of UN Women program (12 evaluations).

In general, the negative factors relate to weaknesses in the same areas as the positive factors, they often represent the absence of the essential practices noted above. Factors detracting from relevance noted by a few evaluations include:

- Inadequate involvement of partners including weak coordination/policy dialogue engagement, unequal sharing of information and non-involvement of partners at the design stage or during program revisions limited the relevance of the program. A failure to create synergy among partners carrying out various activities due to lack of sharing of information was also pointed as a factor limiting effective partnerships (8 evaluations);
- A weak or simply absent partnership strategy also limited relevance – this included lack of partner assessment/mapping, working with too many partners, a short-term approach to partnership, and the absence of defined roles and responsibilities for partners (6 evaluations); and,
- Weaknesses in detailed needs assessment processes also limited overall relevance as reported in some evaluations. This also included the lack needs assessment at community level and insufficient consultations with stakeholders (5 evaluations).

Two evaluations also noted that a challenging or negative political environment and institutional dynamics could limit program relevance.

3.2 Achievement of Objectives

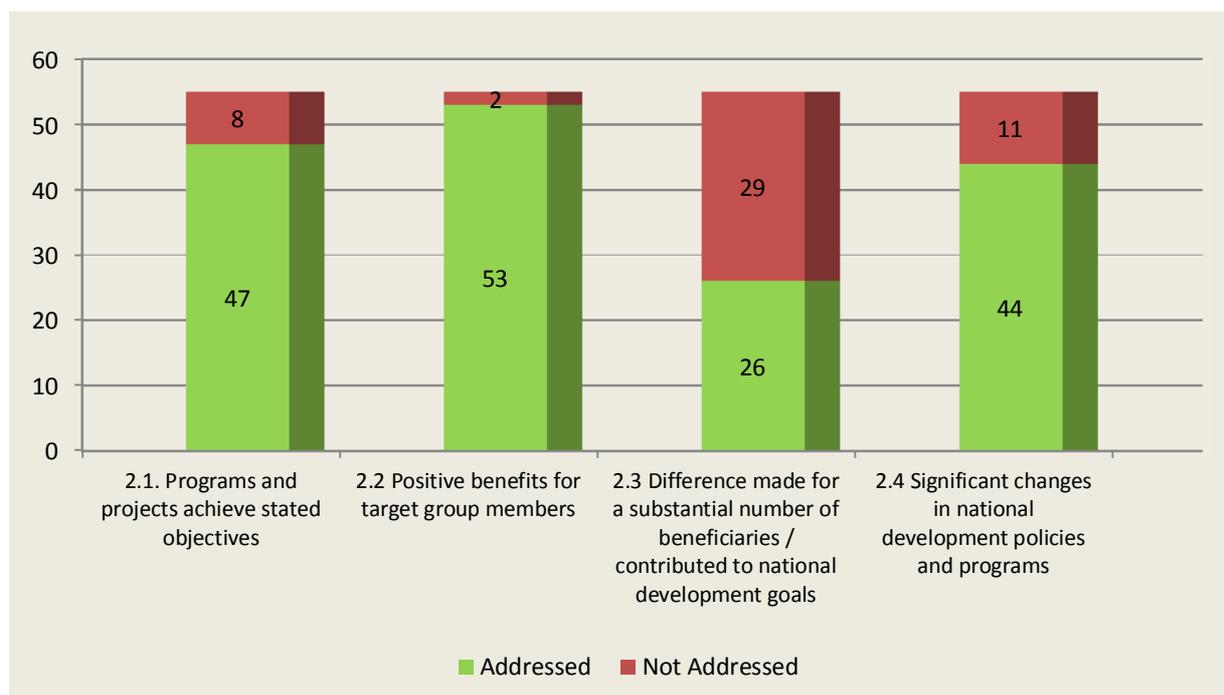
3.2.1 Coverage

Overall, there is a strong coverage of sub-criteria relating to achievement of objectives in the sampled evaluations (Figure 5). This holds true for achievement of objectives by programs and project (sub-criterion 2.1), positive benefits for target group members (sub-criterion 2.2), and for significant changes in national policies and programs (2.4). However, for difference made for substantial number of beneficiaries / contribution to national development goals (sub-criterion 2.3) the coverage declines to moderate level, with only 26 evaluations including relevant findings.³⁸

³⁸ It should be noted that sub-criterion 2.3 coverage is based on whether the evaluation had reported on either quantitative or qualitative results on beneficiaries reached. (see also Annex 7 sub-criterion 2.3 scale). The evaluation is not able to report on reach of beneficiaries due to lack of monitoring data (discussed in sub-criteria 6.2 and 6.3)

The relatively lower coverage of sub-criterion 2.3 is explained by the fact that many evaluations did not provide either quantitative or qualitative estimate of the number of beneficiaries reached by UN Women supported programs. Evaluations also noted that programs were often implemented with limited resources. Coverage is higher in recent evaluation than earlier evaluations for all four sub-criteria of objectives achievement (Annex 6).

Figure 5: Number of Evaluations Addressing Sub-criteria for Achievement of Objectives



3.2.2 Key Findings

In general, the review findings on UN Women supported programs achieving stated objectives were positive (Figure 6). Many of the evaluations (74% - 35/47) reported findings of satisfactory or better in terms of program and projects achieving their stated objectives (sub-criterion 2.1). It means that UN Women programs were reported to achieve more than 50% of their objectives, including the most important ones, in 74% of the evaluations.

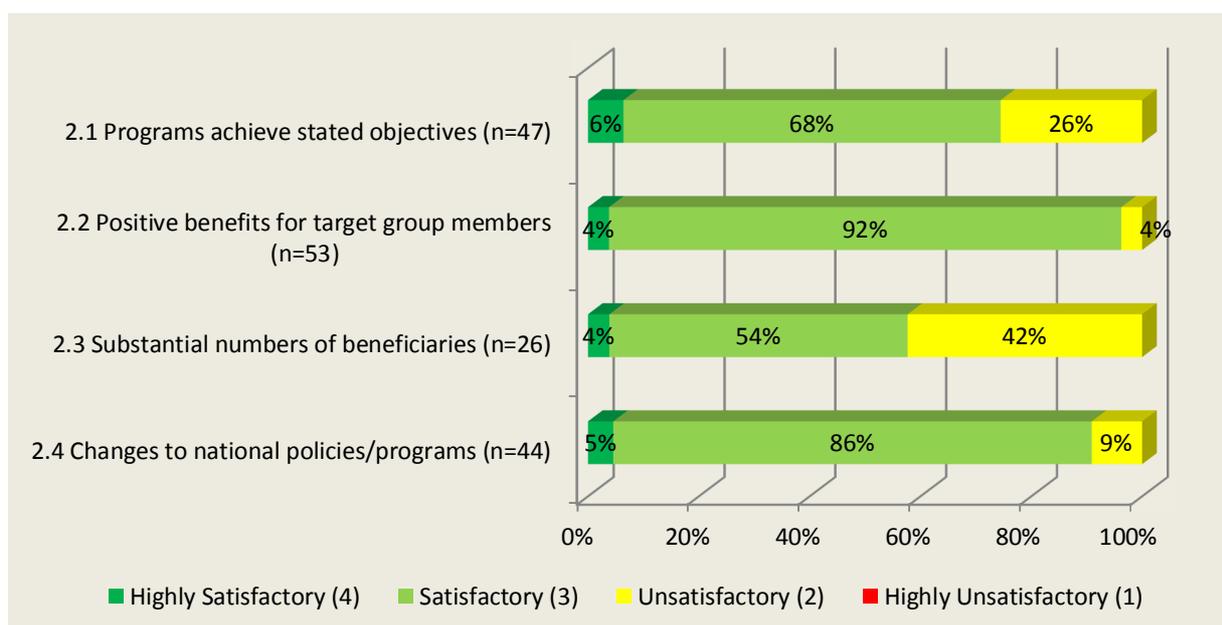
Very strong positive results were seen for programs' ability to provide positive benefits for the target beneficiaries (sub-criterion 2.2)³⁹ and changes made to national policies and programs

³⁹ Evaluations were able to report on positive changes experienced by target group members (at individual, household or community level) based on data collected as part of the evaluation, even if monitoring project data was not available. Data collection methods noted most often in evaluation reports include key informant interviews, focus groups and site visits.

(sub-criterion 2.4) with 96% of the evaluations (51/53) and 91% (40/44) evaluations respectively reporting satisfactory or better findings. However, the findings were only moderately positive for differences made for substantial numbers of beneficiaries by UN Women programs (sub-criterion 2.3)⁴⁰, with only 58% of the evaluations (15/26) reporting satisfactory or better results.

None of the evaluations reported highly unsatisfactory findings for the four sub-criteria (2.1, 2.2, 2.3 and 2.4) under achievement of objectives (Figure 6).

Figure 6: Findings for Achievement of Objectives



The review did not assess the differences in the achievement of objectives among various thematic areas or by regions due to the smaller number of evaluations in each area. However, the review did look at the differences in results across different periods. With reference to achievement of objectives, evaluations in recent years (2014, 2013) reported higher satisfactory or better ratings than the evaluations from the formative years of UN Women (2012, 2011) in all four sub-criteria (2.1, 2.2, 2.3 and 2.4). Most notably in differences made for substantial number of beneficiaries (73% and 36% respectively) followed by contribution to significant changes in national development policies and programs (96% and 83 % respectively) and programs and projects achieving objectives (80% and 68% respectively) – see Annex 6. Also note that highly satisfactory results were noted only in recent (2014, 2013) evaluations (Annex 6). This

⁴⁰ In sub-criterion 2,3, evaluations have not been able to report on substantial number of beneficiaries, partially to due to lack of monitoring data

improvement in the two year period 2013-2014 is consistent with the pattern of findings reported under criteria one: Relevance.

Types of Objectives Achieved

Providing an exhaustive list of objectives achieved by UN Women programs and assessing how they contribute to the six strategic goals (impact areas) is beyond the scope of this review. However, it is possible to mention some specific and key achievements pointed out in the evaluation reports. These objectives were achieved in partnership with government, civil societies, local women's organizations and local communities. It should be noted that the evaluations do not attribute the achievement of objectives exclusively to UN Women or to its support.

The objectives achieved by UN Women supported programs focus on/contribute to one or more of the impact areas⁴¹ of UN Women (Box 3).

Box 3: UN Women's Impact Areas

UN Women program results as cited by evaluations focused on/contributed to:

1. Increased women's leadership and participation (24 evaluations);
2. Increased women's access to economic empowerment and opportunities (14 evaluations);
3. Prevention of violence against women and girls and expand access to services (24 evaluations);
4. Increased women's leadership in peace, security and humanitarian response (13 evaluations);
5. Strengthened the responsiveness of plans and budgets to gender equality at all levels (14 evaluations); and,
6. A comprehensive set of global norms, policies and standards on gender equality and women's empowerment is in place that is dynamic, responds to new and emerging issues, challenges and opportunities and provides a firm basis for action by Governments and other stakeholders at all levels (9 evaluations).

Selected examples of achievement of expected results include:

- Greater access to support services (women's protection centres, free legal support and referral services) for victims of violence against women (Afghanistan);
- Enhanced women's income through livelihoods programming in internally displaced persons camp, reduced women's movement in search of firewood as a source of income (Sudan);
- Increased number of government institutions and officials are more aware of and convinced about the value of need for a gender perspective in strategic planning and national budgeting processes. This has led to significant number of initiatives undertaken by the civil servants including legislative and regulatory changes enabling implementation of gender-responsive

⁴¹ UN Women Strategic Plan 2014-2017

-
- budget and amendments to budget call circulars (Western Balkans);
 - Formation of a coalition of women’s organizations for active involvement in national issues related to women and gender equality. The coalition has continued to be in the forefront in spearheading women’s issues in constitutional reforms and implementation (Kenya);
 - Increased participation of women living with HIV in community engagements and national HIV-related meetings highlighted the public tolerance and shift in attitude towards women living with HIV (Cambodia, Jamaica, Kenya, and Papua New Guinea);
 - Resolution of legal disputes over land, property and inheritance rights for rural women at the community level through project-sponsored legal clinics. Furthermore, increased awareness and knowledge on gender equality and women’s right to land, property and inheritance by government officials and state agency representatives (Kyrgyzstan);
 - Changes in legal/policy frameworks at the national and regional level with gender equality and gender-based violence (Rwanda);
 - Mainstreamed gender in the Law of Victims and Land Restitution and Law on Rural Development (Colombia); and
 - Increased number of male advocates for gender equality due training (Pacific Island countries); and,
 - Establishment of Gender Info base to serve as a repository for gender disaggregated data to make informed policy decisions and develop appropriate action plans for women (South Asia).

One of the evaluation reports reviewed, the evaluation on joint gender programs,⁴² reported on reduced tolerance and increase in number of gender-based violence cases reported, improved women’s civil and political participation, increased national financial allocations to gender-related priorities, greater awareness of gender-related rights, greater networking and advocacy capacity among women’s groups, and, enhanced national evidence base on gender issues.

3.2.3 Contributing Factors

Factors that made a positive contribution to the achievement of program objectives are presented in this sub-section.⁴³

Many evaluations pointed to the following positive factors:

- The ability of UN Women to identify and bring several actors together and work with them

⁴² Thematic Evaluation on Joint Gender Programmes (JGP) in the UN System (2013)

⁴³ As noted in Box 2, the contributing factors are presented in relative order of the frequency with which they were found in the evaluation reports and classified as “most” (found in more than 75% of the relevant evaluations), “many” (between 51% and 75%), “some” (between 20% and 50%) and a “few” (less than 20%).

to address various gender equality and women's empowerment issues including violence against women (VAW), gender-responsive budgeting (GRB), women's human rights and others, to develop appropriate policies and frameworks. This includes gender mainstreaming, sensitizing budgetary actors with gender perspective and facilitating health providers as change makers; in addition to encouraging participation from different levels – national, sub-national and community, (all 55 evaluations address one or more of these issues).⁴⁴ This also included the organization's ability to engage in effective sensitization and capacity building. A specific example is the effective mobilization of regional high-level decision makers on gender issues (particularly sexual and gender based violence) and strengthening commitment through regional declarations – as done for the Kampala Declaration;

- Advocacy and active lobbying role of UN Women in developing national actions plans/strategy on VAW, media and advocacy strategy on gender equality and women's empowerment, and/or drafting of legislation on human trafficking, and regulations for women centres (21 evaluations); and,
- The positive role of UN Women in building/strengthening structures and coalitions among government ministries, CSOs, UN agencies and other partners. This includes program management committees, steering committees, national referral mechanisms, and committees at decentralized levels (19 evaluations).⁴⁵

Some evaluations noted:

- Using appropriate approaches in combination or as stand-alone for better achievement of objectives and continuity of benefits (10 evaluations). Selected examples include: multi-sectoral coordination; holistic approaches to program activities (psychological counselling, food, medical and vocational training); increasing the number of male advocates; use of gender focal points in Ministries; using a broad range of (innovative) initiatives, use of media and hotlines; use of multi-stakeholder approach to program planning and, GRB champions as drivers of change; and,
- Inclusive and participatory process in implementing projects and programs (8 evaluations).

Other positive factors mentioned by evaluations related to UN Women's own capacities including effective program support, staff technical expertise and the responsiveness of staff (5 evaluations).

Negative factors highlighted by some evaluations include:

- The excessively short time frame for projects that hindered the achievement of results at the

⁴⁴ It should be noted that while all the evaluations addressed the gender equality issues, they did not report if they achieved the stated objectives and hence the "not addressed" in the coverage section.

⁴⁵ It should be noted this is about building structures and coalitions while the first bullet is about identifying, and mobilizing people.

outcome level (18 evaluations); and,

- Inadequate monitoring and reporting on the progress of results. Evaluations cited weak results framework and performance measurement tools, lack of baseline data, dependency on inconsistent Implementing Partner reports, and/or overall an inadequate system for measuring and reporting progress (17 evaluations). While this may not affect achievement of objectives directly, it affects monitoring of progress made and hence hinders the ability to make changes in delivery and improve results, as required. In a few cases, programs achievements were only known after an assessment by an (independent) evaluation team, as monitoring of progress had not taken place or was not adequate during the implementation phase (see further discussions in sub-criteria 6.2 and 6.3).

Other detracting factors noted include:

- A project design with too limited resources negatively affected the reach of the project (9 evaluations);
- Weak program design sometimes hindered achievement of objectives – this included a missing or weak theory of change, over ambitious goals, and an insufficient orientation on results (9 evaluations); and,
- Socio-political constraints and socio-cultural difficulties hinder achievement of objectives (3 evaluations).

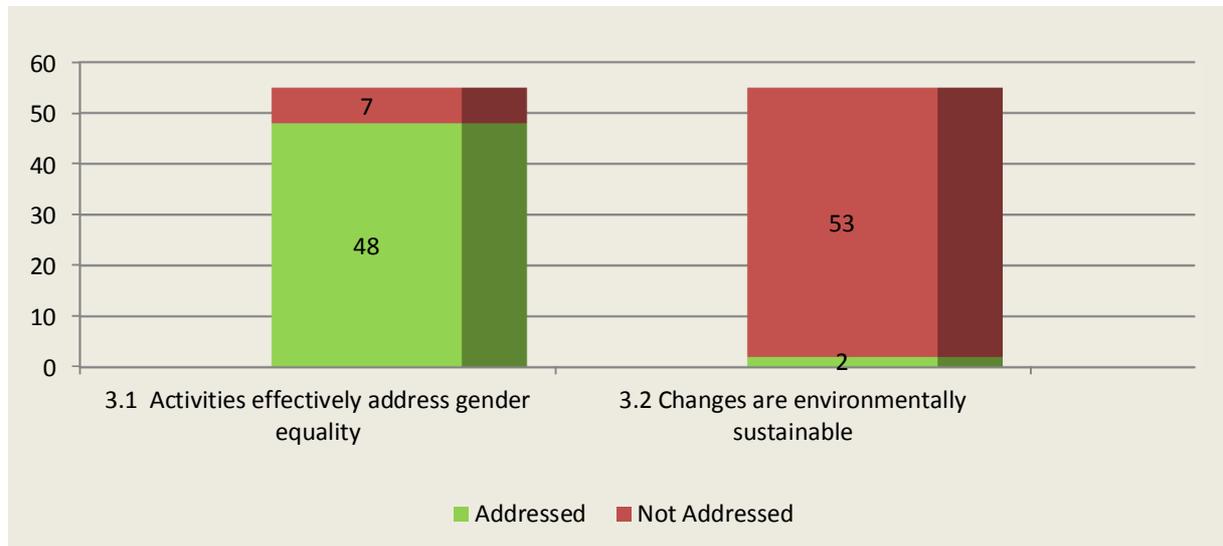
3.3 Cross-Cutting Themes

3.3.1 Coverage

The levels of coverage of the two cross-cutting themes present a contrasting picture (Figure 7). The review indicated a strong coverage for effectively addressing gender equality (sub-criterion 3.1) – with relevant findings addressed in 48 of 55 evaluations. On the other hand, the coverage of whether changes are environmentally sustainable (sub-criterion 3.2) was weak and addressed in only two evaluation reports.⁴⁶ As in earlier two criteria (relevance and achievement of objectives), the recent evaluations had better coverage than earlier ones for addressing gender equality.

⁴⁶ The two evaluations also addressed gender equality and the findings were “satisfactory” for both cross-cutting criteria.

Figure 7: Number of Evaluations Addressing Sub-criteria for Cross-cutting Themes



It is of interest to note that only 87% of the evaluations in the sample addressed the question of gender equality for UN Women programs. In order for the review to consider this sub-criterion adequately covered, evaluations had to not only identify the extent to which the program or project in question had incorporated gender equality objectives but also whether they had been achieved.⁴⁷ Gender equality could also be addressed in an evaluation report if it assessed the program’s success in mainstreaming gender equality. On the other hand, if an evaluation simply reported the portion of girls or women receiving benefits compared to boys or men, the rating for this sub-criterion was labeled as ‘not addressed’ because there was no reference to the program’s success in addressing equality.

Given the very weak level of coverage of environmental sustainability, with just two evaluations addressing the sub-criterion 3.2, the review is not able to report on the effectiveness of UN Women supported programs in this area.

3.3.2 Key Findings in Gender Equality

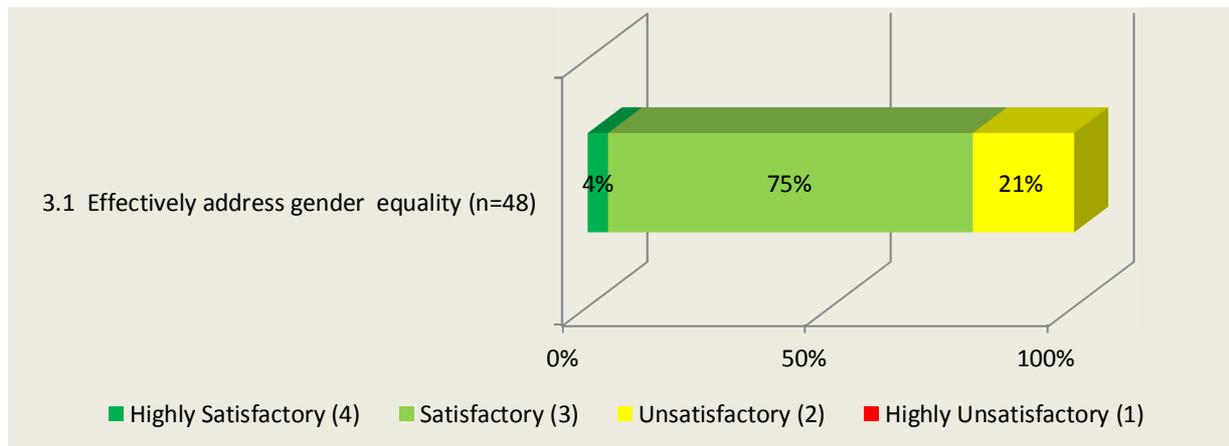
The evaluations report substantially positive findings for UN Women programs’ effectiveness in addressing gender equality (sub-criterion 3.1). Most evaluations (79% - 38/48) that addressed gender equality reported satisfactory or better results. None of the evaluations reported highly

⁴⁷ Some evaluations reported gender equality results; however, they failed to report if those results met the stated objectives and/or to what extent they met the stated objectives. In this case also the review team noted it as “not addressed”.

unsatisfactory (Figure 8).

Consistent with findings for the first two criteria, evaluations in 2014 and 2013 reported higher satisfactory or better ratings than evaluations published in 2011 and 2012 – 89% and 68% respectively in relation to how well UN Women programs address gender equality (Annex 6).

Figure 8: Findings for Effectiveness in Supporting Gender Equality



Selected examples of success achieved by UN Women programs include:

- Increased access and participation in education by girls - in program areas (Ethiopia);
- Improved access for women to decentralized legal and paralegal services (Mali and Sierra Leone)
- Reduction in the number of cases reported on the incidence of violence against women in project communities (Palestine);
- Increased income of women through livelihood activities and small enterprises (Sudan, Gaza);
- Women Centres (specifically within the targeted camps in Sudan) facilitated the opportunities for social mobilization and sense of belongingness for women; and,
- Increased efforts by civil servants and government official to mainstream gender in their activities, processes and budgets (see Box 4).

Box 4: Gender-Responsive Budgeting (GRB) in Western Balkans

Results achieved with reference to gender-responsive budgeting include:

- “An increased number of governmental institutions, in particular, budgetary actors are now more aware and convinced about the value of and need for a gender perspective in strategic planning and budgeting processes.”
- “Initiatives undertaken by civil servants and government institutions to mainstream gender in sector policies and budget processes include: legislative and regulatory changes (which enable and support GRB implementation), amendments to budget call circulars, and the development of GRB specific strategies and action plans.”

Project 76742 "Promoting Gender Responsive Policies in the Western Balkans" Final evaluation, 2013, p.4

3.3.3 Contributing Factors

UN Women, in accordance with its core mandate addresses gender equality and women’s empowerment. Evaluations reviewed indicate that in general UN Women programs align with gender equality priorities and apply human rights-based approaches. As a result, it should be noted that many of the factors mentioned in achievement of objectives (Criterion Two) are also relevant and applicable to this sub-criterion.

Many evaluations report on results relating to efforts aimed at ending VAW, and frequently cited the technical expertise of UN Women as the positive factor in achieving results in the area is (24 evaluations).

Other positive factors cited by some evaluations include:

- Adopting of participatory approach and inclusive processes by UN Women has enabled better capacity building at all levels and also facilitated national and community level ownership (20 evaluations);
- UN Women programs have provided support to the development of national action plans for gender, ending violence against women, policies and strategies, collection of gender disaggregated data and strengthening the gender info-base through capacity building and sensitization at various levels (20 evaluations); and,
- An ability to identify strategic partners and creating coalitions, committees, national/regional networks and working groups for advocacy and to implement programs (19 evaluations).

Negative factors hindering achievements in gender equality, as reported by some evaluations include:

- Program designs which include short-timeframes and a project-to-project approach,

-
- ambitious goals, lack of theory of change and weak results-orientation (18 evaluations);
 - Poor performance measurement frameworks and which hindered measurement of progress on achievements due to ill-defined indicators and a lack of baseline data (17 evaluations) – this has also been discussed in detail in sub-criteria 6.2 and 6.3; and,
 - A challenging political context coupled with deep-seated negative cultural values and widespread stigma (10 evaluations).

Other negative factors cited by evaluations include – male dominated organizational structures, absence of involving men in program implementation, lack of progress in policy development, and, non-implementation of policies due to shortfalls in budgets (see also Box 5).

Box 5: Factors Hindering Gender Mainstreaming

Various factors hinder gender mainstreaming. These are summarized well by National AIDS Coordinating Agencies (NACAs) in the five programme countries. These include:

- a) “Reliance on a few individuals for the mainstreaming of gender equality considerations within NACAs’ structures and programming;
- b) Absence of inadequacy of M&E systems capturing the gender dimensions of HIV programming impacts;
- c) Absence of accountability mechanisms for gender mainstreaming at different levels of NACAs and in other government ministries; and,
- d) Staff turnover”

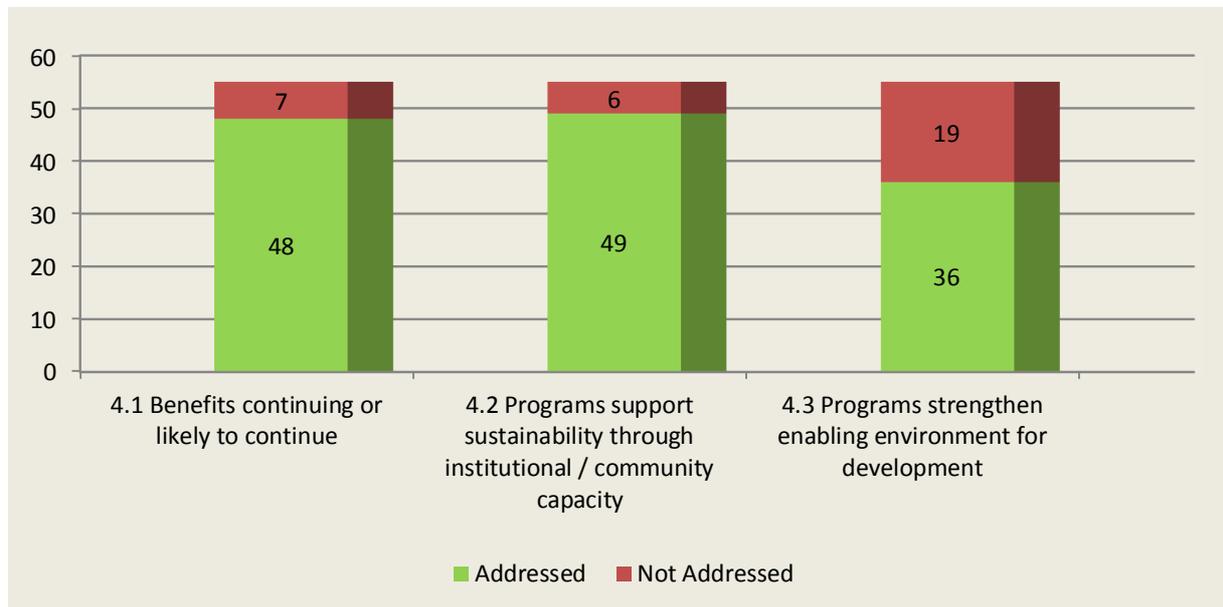
End of Programme Evaluation of EC-UN Women programme Supporting Gender Equality in the Context of HIV/AIDS, 2013 p.33

3.4 Sustainability

3.4.1 Coverage

The coverage for two of the three sub-criteria with respect to sustainability was strong (Figure 9). The probability of benefits continuing after program completion (sub-criterion 4.1) and sustainability through institutional/community capacity (sub-criterion 4.2) were covered in 48 and 49 evaluations respectively. However, the coverage was moderate for contributing to a strengthened enabling environment for development (sub-criterion 4.3), as this sub-criterion was addressed in only 36 evaluations. Notably, the coverage in recent evaluations was higher than earlier evaluations for all three sub-criteria (Annex 6).

Figure 9: Number of Evaluations Addressing Sub-criteria for Sustainability



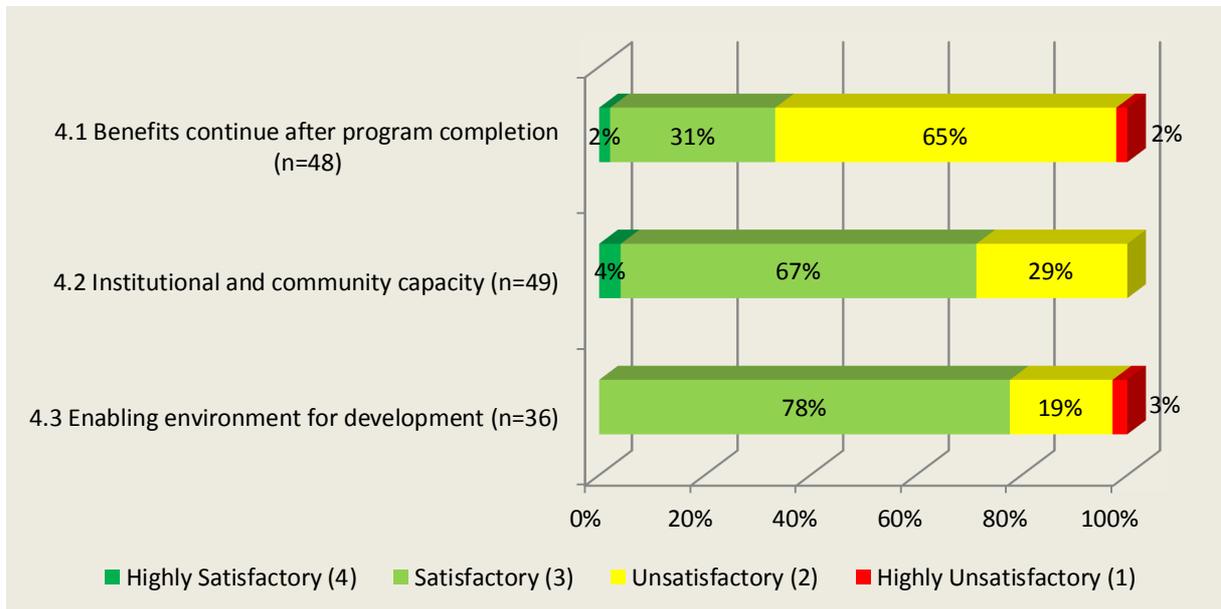
3.4.2 Key Findings

Overall, the results with respect to sustainability are mixed (Figure 10). Evaluations reported positively on two sub-criteria – programs support sustainability through institutional and/or community capacity (sub-criterion 4.2) and programs strengthen enabling environment for development (sub-criterion 4.3) – with 71% (35/49) and 78% (28/36) respectively reporting satisfactory or better. As, discussed earlier there is improved reporting (better coverage) on enabling environment in recent evaluations.

However, only 33% (16/48) of evaluations reported the likelihood of benefits continuing after program completion (sub-criteria 4.1) as satisfactory or better.

Interestingly (and in contrast to the first three main criteria addressed above), the recent (2014, 2013) evaluations more often get lower satisfactory or better rating than earlier (2012, 2011) evaluations regarding the likelihood of benefits continuing after program completion (sub-criterion 4.1). There was no explicit explanation found for this shift. On the other hand, evaluations from recent years more often get satisfactory or better ratings for sustainability through institutional and/or community capacity (sub-criterion 4.2) than those in earlier years (Annex 6). The evaluation reports of recent years were significantly more positive for programs strengthening enabling environment for development (sub-criterion 4.3), with 90% reflecting findings of satisfactory or better, compared to earlier evaluations (63%).

Figure 10: Findings for Sustainability



The results for institutional/community capacity building and enabling environment (sub-criteria 4.2 and 4.3) are more positive than for benefits being sustained in the future (sub-criterion 4.1) in the absence of UN Women support. This may reflect aspects relating to UN Women programming at the country level, as pointed out by the evaluation reports reviewed. For example, a project-to-project approach may not adequately address ownership and exit strategy and hence will hinder sustainability. On the other hand, UN Women’s approach to developing the capacities of its partners and target groups; and, UN Women’s focus on policy engagement and advocacy that to create an enabling environment with regard to gender equality and women’s empowerment will enhance sustainability.

On the other hand, the relatively poor results reported for sub-criterion 4.1 often seem to be cited in relation to the inability of the supported implementing partner and/or national and local government to assume the costs of UN Women supported programs in the absence of continuing support.

Specific examples on sustainability from evaluations include:

- Integration of gender considerations into planning and budgeting processes (including sector wide approaches) and leadership of national partners are likely to ensure continuation of gender equality results at country level (Cameroon, Ethiopia, India, Morocco and Nepal);
- Women farmers are using revenues from farm activities to buy seeds and fertilizer for subsequent season planting, paying for additional training and contributing to set-up rural

fund to help establish self-help groups among vulnerable population in their communities (Kyrgyzstan);

- Development of technical capacity at ministries and relevant CSOs and involving CSOs in planning and implementation has led to activities being continued (by some CSOs) after project completion (Sudan); and,
- Exit strategies (when in place) have facilitated continuation of project activities and benefits (Kenya).

3.4.3 Contributing Factors

Several factors reported in the evaluations were seen to facilitate the sustainability of UN Women supported projects and programs. Positive factors noted by some evaluations include:

- Participatory methodologies used in the strengthening the capacity of targeted individuals and organizations at various levels, including at the community level (20 evaluations). This included training, meetings with experts and organizational development;
- Engaging existing structures for the integration of program components and processes and thereby enhancing institutionalization of the program delivery. This included working with government institutions, universities, civil societies, women's groups and existing coalitions and networks and also providing interventions that are complementary to existing programs (13 evaluations);
- National ownership and commitment (12 evaluations) – examples of which include allocation of budgets, implementation of policies, integration of program framework/components; and,
- Support and ownership of the program at the community and decentralized levels – this was seen through active uptake of program results, the creation of local coalitions, financial independence of some partners and/or their ability to mobilize resources (10 evaluations) – see also Box 6.

Other positive factors noted by one or two evaluations include, women's leadership and increased participation in decision making due to UN Women programming and the practice of local NGOs and CSOs seeking advice from religious leaders, leading to buy-in and enhanced the probability to sustain benefits after UN Women program was completed. Involvement of religious leaders, especially in communities with socio-cultural mores and religious norms that may hinder gender equality and empowerment of women, ensured continuity of benefits.

Box 6: Community Capacity Sustainability in Kenya

Sustainability due to community capacity built was evident in Kenya.

- The young women and youth who were trained on participating in peace and conflict management processes in their communities, became advocates in their communities and were very involved in advocating for peace during the 2013 elections. Most are also now involved in the 350 sub-county peace committees established by the National Steering Committee on peace building and conflict management

Final evaluation of Gender & Governance Programme (GGP III Kenya), Gender and Governance Component, 2013, p.22

The main factors cited as detracting from sustainability for the most part reflected the program's failure to optimize those factors cited for enhancing sustainability. It should also be noted that these factors tend to be inter-related so that, for example, lack of engagement with appropriate partners could lead to lack of ownership/commitment on the part of government and hence limited or non-existent budget allocations.

Factors detracting from sustainability and noted by some evaluations include:

- Excessive reliance on donor and/or external support for funding due to limited financial and operational capacity and the government's inability or lack of interest to continue the program. This included instances of "parties resist the promotion of women in politics" (22 evaluations);
- Lack of ownership has reflected in low involvement/engagement by national and provincial actors, lack of support from the ministries for gender focal persons, non-allocation of budgets (11 evaluations). Two evaluations also cited language as an issue for lack of ownership – "dominance of English" (many key documents were reported to be in English);
- The use of a project-to-project approach due to insufficient core and/or other resources which implies less continuity and less ability to develop a long-term approach for dealing with deep-rooted, structural and highly challenging issues. It was also noted that at times programs were designed to lay the foundation for more rapid progress in the future and hence institutionalization and/or sustainability were not key priorities in the design. The approach was also seen to affect sustained engagement of the government (10 evaluations); and,
- Lack of an exit strategy - this could mean that no appropriate phase-out plan was designed, local capacities were not developed, institutionalization was inadequate or, even if capacities were built they were at individual and not organizational level (9 evaluations).

Negative factors pointed out by a few evaluations include:

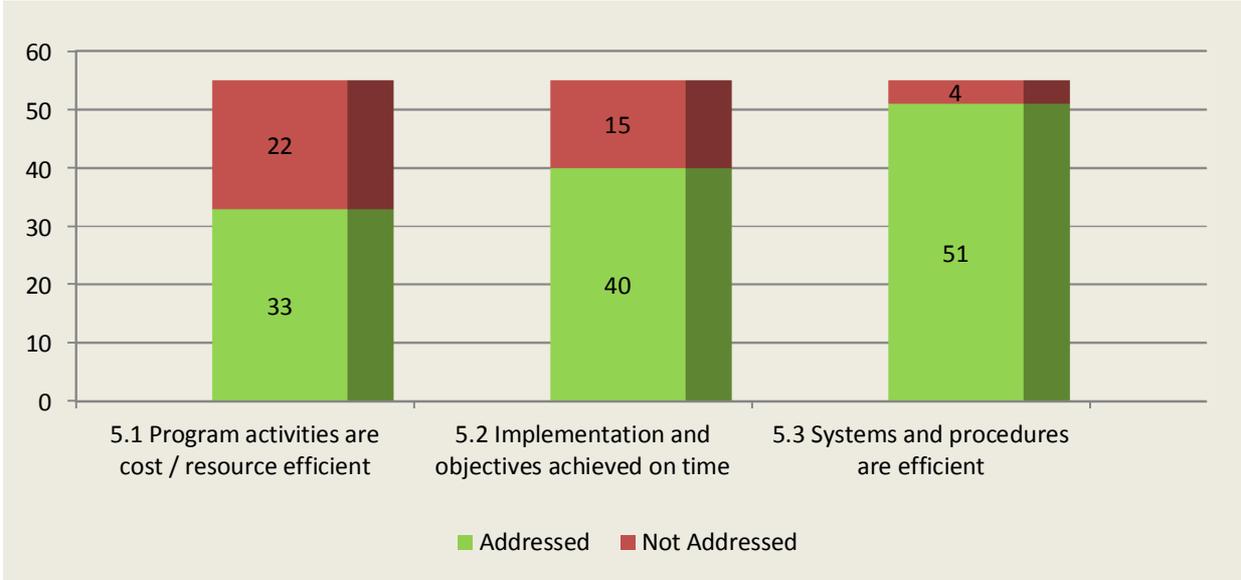
- Ill-designed capacity development approaches not facilitating institutionalization of capacities. For example, the focus on individuals rather than the organization and the design of training without needs assessment or with short-term focus (8 evaluations);
- High staff turnover not only leading to recurring costs of recruitment and training but also loss of institutional memory (6 evaluations);
- Resistance to gender equality due to deep-seated cultural values and wide-spread stigma threaten continuity benefits beyond the program/project life, especially when programs are short-term in nature (5 evaluations); and,
- A weak strategic engagement in terms of partnerships with CSOs, supporting government’s coordination roles, and inconsistent collaboration with development partners (5 evaluations).

3.5 Efficiency

3.5.1 Coverage

Coverage levels for the three sub-criteria of efficiency range from moderate for cost and resource efficiency of programming (sub-criterion 5.1) and implementation and objectives achieved on time (sub-criterion 5.2) to strong for systems and procedures are efficient (sub-criterion 5.3) - Figure 11. It must be noted that the sub-criterion 5.2 is at the highest level of the moderate coverage range (40 evaluations).

Figure 11: Number of Evaluations Addressing Sub-criteria for Efficiency



The moderate level of coverage of sub-criteria on cost efficiency (sub-criterion 5.1) and timelines (sub-criterion 5.2) is an interesting finding. The review team finds that sub-criteria 5.1 and 5.2 tend not to be addressed systematically in many evaluations. Findings on these sub-criteria may be more likely to be included in evaluation reports when problems in cost efficiency or timely delivery of program inputs are encountered during field level evaluations. This may then be presented to explain limitations in the achievement of program objectives.

It is also important to note that moderate coverage of cost/resource efficiency (sub-criterion 5.1) does not mean that UN Women evaluations have directly assessed cost-efficiency. The fact that this criterion has been addressed only indicates that the evaluation has made some judgement (quantitative or qualitative) on cost/resource efficiency and/or reference to unit cost of program inputs or outputs (for example, “*more was achieved with limited resources*”). While information of overall program cost was reported in 49 evaluations, only 33 made some efforts to make a judgement on cost efficiency of the program that was being evaluated. The team has noted that a lack of availability of data, in addition to insufficient time and budget, may have been factors for not having a detailed assessment on cost/resource efficiency in the evaluations.

The strong level of coverage for efficiency of systems and procedures (sub-criterion 5.3) indicates that these are more likely to have been systematically included in the terms of reference of the evaluations.

3.5.2 Key Findings

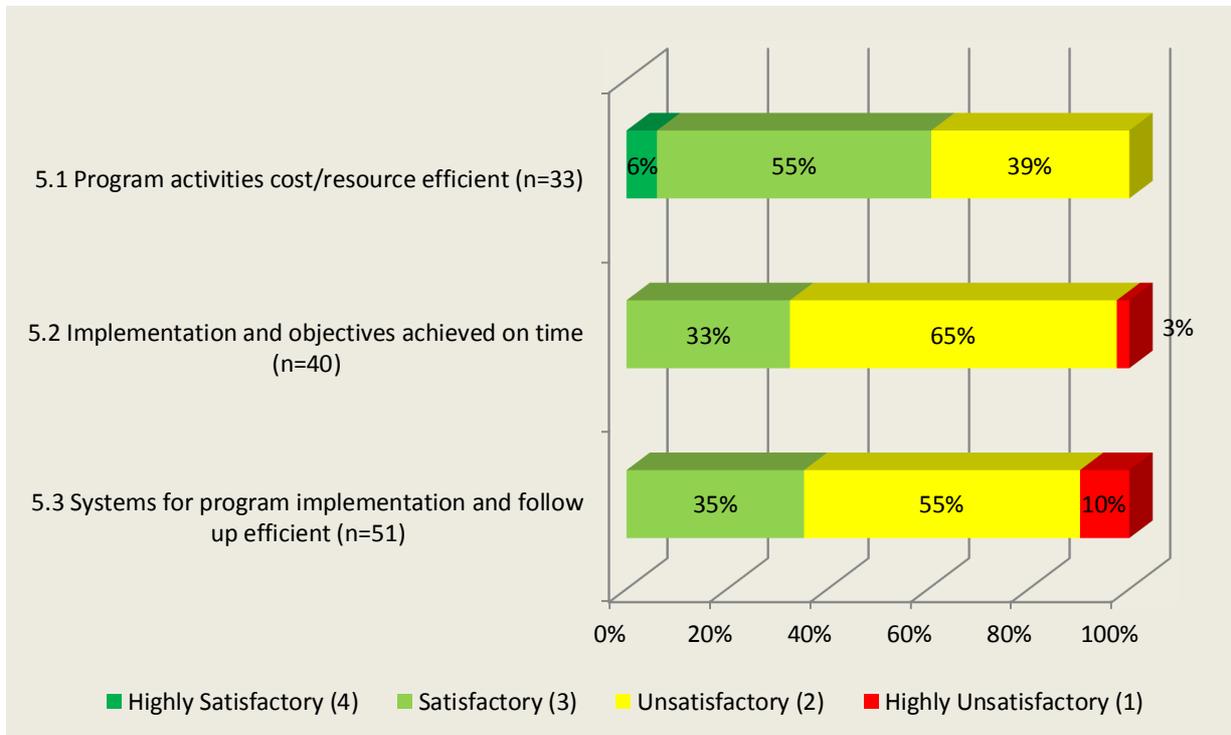
Despite a moderate coverage, a significant 61% (20/33) of evaluations reported satisfactory or better findings on cost/resource efficiency of UN Women programs (sub-criterion 5.1). Also notable is the fact that none of the evaluations reported highly unsatisfactory results for cost/resource efficiency (Figure 12).

Only about one-third of the evaluations reported satisfactory or better findings for program implementation and objectives achieved on time (sub-criterion 5.2) and for systems for program implementation and follow-up are efficient (sub-criterion 5.3) – 33% (13/40) and 35% (18/51) respectively. The review team notes that there may be a tendency to highlight issues and delays in implementation, fund disbursement and administration procedures and under report when there are no problems.

The review team reviewed a sample of 17 randomly selected ToRs (for the evaluations reviewed) to assess if cost-efficiency was a specific ToR requirement. The results were mixed. While in seven ToRs it was explicitly stated (e.g. was it cost-efficient, can it be done at a lower cost, or was the conversion from input to output economical) in seven others, it was blurred, as the requirement was “if the resources were used optimally or efficiently”. In three other ToRs,

neither cost nor resource efficiency assessment was a requirement.

Figure 12: Findings on Efficiency



While the difference between the recent (2014, 2013) and earlier (2012, 2011) evaluations was not significant in cost/resource efficiency and systems and procedures efficiency, the earlier evaluations had marginally more satisfactory or better results than the recent ones. Interestingly, earlier evaluations were more positive in assessing timely implementation than recent reports. The introduction of new Evaluation policy in 2013, new systems and process in the last two years, are seen as reasons for this shift, as it has also increased better coverage and reporting of efficiency (Criterion: Five) in recent evaluations (Annex 6), a key factor for marginally lower results than earlier evaluations. At the same time, efficiency is an area to improve for UN Women in its programming.⁴⁸

Selected examples from evaluation reports on how UN Women programs have tried to improve

⁴⁸ This has been also noted in the recently conducted meta-analysis on evaluation managed in by UN Women in 2013 – “Individual good examples of program efficiency are set against a back drop of fund disbursement delays, delays in appointing people in project-critical positions at field level, and gaps in knowledge management systems”. – UN Women (2014). What can we learn from UN Women Evaluations? p.21

efficiency include:

- Use of semi-volunteer groups for program activities, reduced costs and enabled to achieve more within the limited budget (Georgia);
- Tight cost controls and good financial management (Gaza Strip and West Bank);
- Establishing synergies with other GRB related work by collaborating with other UN agencies and bilateral donor agencies (Rwanda and Nepal); and,
- Working with implementing partners, who have presence at the grassroots level has increased efficiency (Kyrgyzstan).

3.5.3 Contributing Factors

As very different factors contribute to results regarding cost efficiency, when compared to timeliness and the efficiency of administrative systems, they are discussed separately in this section.

Sub-Criterion 5.1: Cost/Resource Efficiency

Positive factors that promote cost / resource efficiency of programs as pointed out by some evaluations include:

- Good use of limited resources to achieve more or complete all activities within the given resource envelope (11 evaluations). This was often reflected in the evaluator judgement statements such as - projects represented good “*value for money*”; and “*UN Women had miniscule funding and managed to do miracles with that.*”

A few evaluations noted:

- The leveraging of existing/limited resources - this included – use of volunteer/semi-volunteer groups, capitalizing on capacities of partner organizations, existing staff taking on program management and creating linkages/synergies with existing programs (5 evaluations); and,
- Savings by internal cost control, thereby keeping administrative overheads low in the total program cost (5 evaluations). Examples include – selective use of international consultants, cutting costs on business lunches, having organizations bearing the expenses of fuel when vehicles are provided and utilizing 100% of the project funding for investment with partners absorbing the overhead costs.

Other positive factors noted by one or two evaluations include – good/transparent financial management and the ability to make well-judged decisions to make implementation cost-effective.

Negative factors that contribute to reduced cost/resource efficiency, as noted by a few

evaluations include:

- Inability manage/work within limited resources and short timeframes contributing to high unit costs (6 evaluations);
- Poor financial monitoring/review and management (4 evaluations). This included “costs not broken down by ATLAS⁴⁹ analysis,” no evidence of progress on the biggest component financially; and lack of transparency in allocation of funds within a project; and,
- Human resource challenges and issues of implementation (3 evaluations). Human resource issues are also discussed in relation to the other efficiency sub-criteria.

Sub-criterion 5.2 – Programs are Implemented on Time and Sub-criterion 5.3 – Systems and Procedures are Effective for Program Implementation

There was a very close relationship between the reported timeliness of implementation and findings regarding the efficiency of UN Women’s systems and procedures for program implementation. Many of the factors supporting or hindering efficiency for these two sub-criteria tended to be identical.

Positive factors include:

- Dedicated staff and focal persons of UN Women who were flexible and willing to adjust the programs to evolving needs or to do more than required by their stated responsibilities and/or were familiar with government procedures (7 evaluations)and,
- The capacity of implementing partners to catch up after initial delays and complete projects on time. This included a local presence and being adequately staffed (5 evaluations);

Factors that detract from timeliness and efficiency often also relate to UN Women’s capacity and processes. Some evaluations noted:

- Weak administrative and financial processes including bureaucracy and internal management challenges (21 evaluations). This included delayed start-up of projects due to delays in planning which contributed to under-spending in initial years and a rush toward the end. This also led to excessive use of no-cost extensions and in one instance led to donor funds returned unspent;
- Absence of or limited staff capacity which included staffing delays, limited access to technical expertise, inadequate administrative staff and, high staff turnover (20 evaluations);

⁴⁹ Automatically Tuned Linear Algebra Software (ATLAS) – is an enterprise resource planning system used UN Women (and other UN agencies – UNFPA, UNOPS) for budget management, accounting and auditing

-
- Issues with slow disbursement and funding delays (17 evaluations). This means that allocation funds for partners to use and implement was not timely; and,
 - Lack of defined roles and responsibility at Country Offices (COs) and more specifically in inter-agency/donor projects (10 evaluations). Examples include :
 - The complexity of communication loops among UN agencies or in UN Women-donor partnerships;
 - Lack of division of labour / structure – as noted in one evaluation “*joint program not more efficient than a bilateral single agency intervention*”;
 - Limited financial and administrative autonomy and/or decision making authority in a few COs; and,
 - Less consistent and not well-defined roles, in partnerships for UN Women COs.

Other negative factors that were pointed out by a few evaluations:

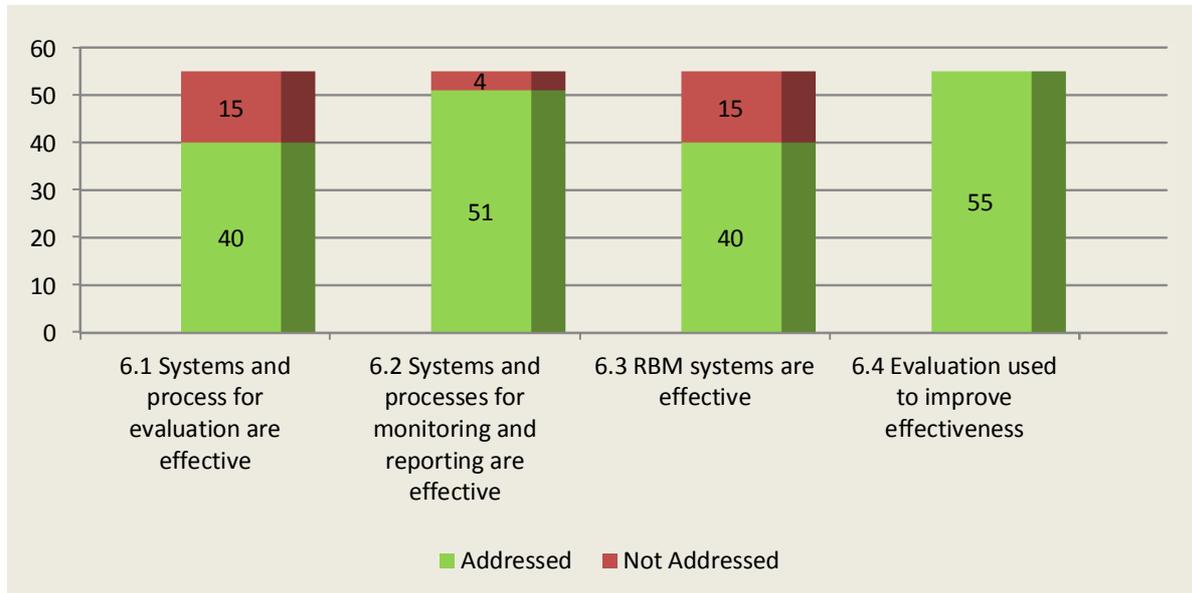
- Flaws in the design of the program (8 evaluations) with regard to short and unrealistic timeframes, weak CSO engagement, and lack of coherence to avoid duplication and/or overlap in implementation;
- Issues in UN Women’s partner mechanisms such as an inability to leverage partner resources, working with too many CSO partners, delays in getting reports from partners, and a mismatch between disbursed and approved budget allocations (5 evaluations); and,
- Problems during the transition from UNIFEM to UN Women (2 evaluations).

3.6 Use of Monitoring and Evaluation

3.6.1 Coverage

Overall, coverage for use of monitoring and evaluation is quite strong. Coverage for two of the four sub-criteria is rated strong – systems and processes for results monitoring and reporting (sub-criterion 6.2) and evaluation used to improve effectiveness (sub-criterion 6.4), with 51 and 55 evaluations respectively addressing them. The coverage was moderate for the other two sub-criteria – systems and process for evaluation are effective (sub-criterion 6.1) and results-based management (RBM) systems are effective (sub-criterion 6.3), with 40 evaluation addressing each (Figure 13). It must be noted that 40 is the maximum threshold for moderate coverage.

Figure 13: Number of Evaluations Addressing Sub-criteria for Using Evaluation and Monitoring



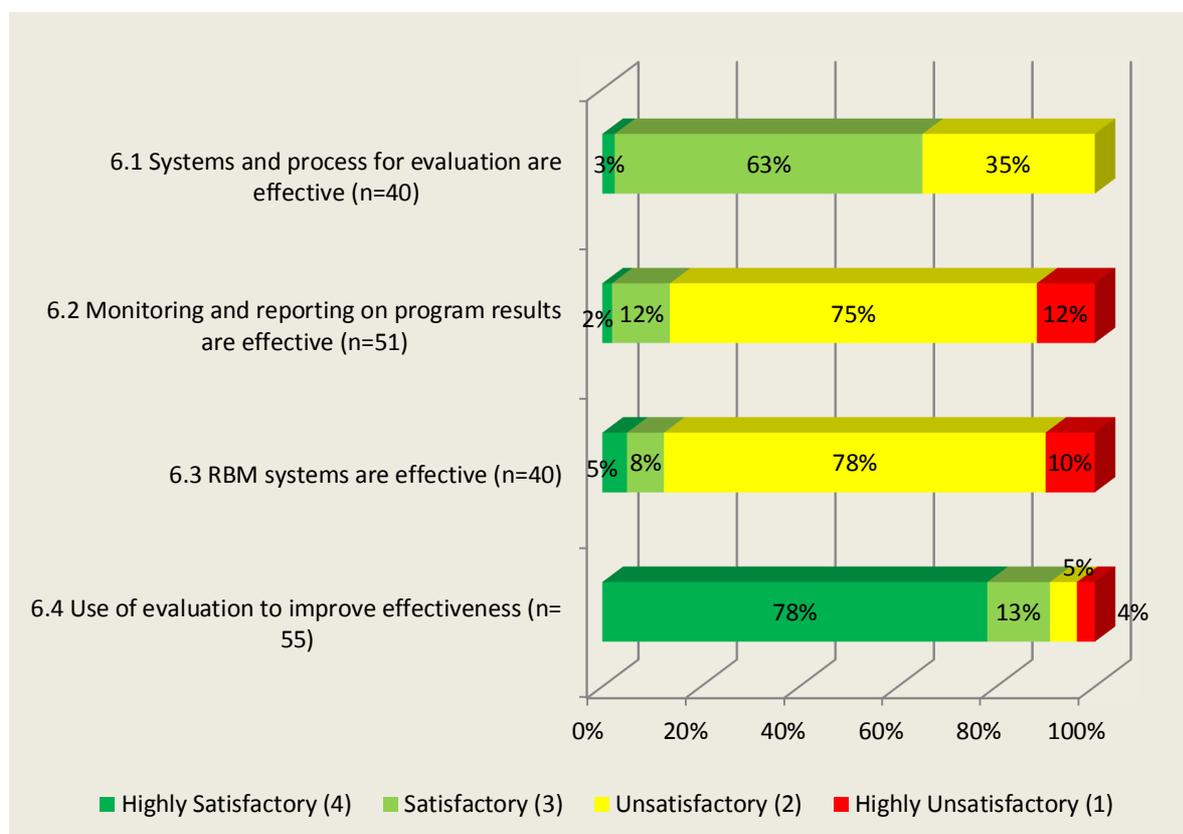
3.6.2 Key Findings

The results for these criteria are mixed. They present a positive view on effective systems and process for evaluation practice (sub-criterion 6.1) and the use of evaluation to improve effectiveness (sub-criterion 6.4). On the other hand, monitoring and reporting on program results and the use of RBM systems have been assessed quite negatively in the evaluations reviewed (Figure 14).

Many evaluations (66% - 26/40) reported findings of satisfactory or better for systems and process for evaluation are effective (sub-criterion 6.1). Notably, results are very positive for use of evaluation to improve development effectiveness (sub-criterion 6.4) - with 91% (50/55) of evaluation reporting satisfactory or better.

It is interesting that, despite findings of weak systems and processes for monitoring and reporting, and ineffective results-based management systems, findings are quite positive for effective evaluation systems and process, and the use of evaluation to improve development effectiveness. This reflects the fact that evaluation as a process seems well integrated into the program planning cycle in UN Women (with mid-term evaluations and/or end of program/project evaluations, which can be mandatory). It also indicates that evaluation teams have been able to access other secondary or primary data on program results where regular monitoring information is not available or weak.

Figure 14: Findings for Using Evaluation and Monitoring



In general, recent (2014, 2013) evaluations showed more satisfactory or better results in all four sub-criteria (6.1, 6.2, 6.3 and 6.4) compared to earlier (2012, 2011) evaluations (Annex 6). More specifically, recent evaluations report significantly more positive results than earlier evaluations for sub-criterion 6.1 - systems and processes for evaluation are effective (with 78% and 55% respectively).

This reflects the recent efforts of UN Women to strengthen evaluation and results reporting systems and process (see Section 1.4).

Sub-criterion 6.2: Effective Systems for Monitoring and Reporting on Program Results and 6.3: RBM Systems are Effective

It must be noted that these two sub-criteria have the fewest number of evaluations reporting

satisfactory or better results among the 19 that measure development effectiveness.⁵⁰ However, the findings reported for sub-criteria 6.2 and 6.3 indicate that UN Women has faced challenges in establishing and strengthening systems for country and regional level in monitoring and RBM during the review period. The preliminary review carried out for this study identified a considerable ongoing effort to strengthen program results definition and monitoring system at country and regional level in UN Women.⁵¹

Strengthening of the monitoring and results reporting function at UN Women is a work-in-progress. In 2014, UN Women developed a strategy to strengthen RBM in the organization and began the process of building a RBM system.⁵² The appointment of regional Evaluation Specialists reporting directly to the Director of IEO, the implementation of GERAAS, and the commission of meta-evaluation and meta-analysis reports are all examples of recent development and efforts to improve monitoring, reporting, and result-based management at UN Women⁵³. Furthermore, in 2014 UN Women has developed a strategy for strengthening, RBM in UN Women, created corporate RBM training modules and is in the process of building a RBM system for UN Women to enable management of country programs and linking country and regional results to Strategic Plan. However, it is still too early to assess whether these measures will indeed lead to more effective RBM systems and more effective systems for Monitoring and Reporting.

Specific positive examples of monitoring, reporting and RBM systems from the evaluations reviewed include:

- Processes and tools for program planning, monitoring and financial documentation and reporting were well designed and used on school canteen project (Palestine);
- At the project level there were relevant indicators that were measurable (Ethiopia);
- M&E system put in place including quarterly self-reporting by partners that was verified by M&E/project officers during partner visits (Zimbabwe); and,
- Performance measurement data used to fine-tune program/projects (Ecuador).

⁵⁰ They are ranked one and two respectively from the bottom in terms of the proportion of evaluations reporting result of for satisfactory or better.

⁵¹ These efforts are very recent and any resulting improvements in results definition and monitoring systems might not be reflected in the findings of the evaluations reviewed (since many of them deal with programming designed and implemented earlier)

⁵² Ten Steps to Strengthened RBM in UN Women (2014)

⁵³ The position of Director of IEO was created in 2013 and most changes have happened since the appointment of the Director.

Sub-criterion 6.4: Use of Evaluations to Improve Effectiveness

The results for sub-criterion 6.4 reflect the extent to which UN Women is preparing management responses for its evaluations and whether these include adequate responses to the recommendations, including an action plan, clearly designated responsibility for its implementation and a system for reporting on the status of progress on actions taken.

The review team was able to find management responses for 53 of the 55 evaluations reviewed and 43 of those were assessed to be of high quality. Ten other evaluations with a management response either rejected many of the recommendations or did not address all recommendations with their action plan.

This sub-criterion has the most highly satisfactory results (78%) among the 19 used to assess development effectiveness. The large number (43/55) of evaluations reporting highly satisfactory results for this sub-criterion reflects the fact that the review team was able to identify and review the formal management response to an evaluation that was complete, responded positively to recommendations or identified reasons for rejecting some recommendations. These management responses also included a plan with assigned responsibilities for implementing the response along with a status report on progress made.

Highly unsatisfactory results reflect the review team's decision to an evaluation report whenever a management response could not be located in UN Women's GATE system. For some of these evaluations a management response may have been prepared and approved but not up-loaded to the database at the time of the review.

3.6.3 Contributing Factors

Many evaluation reports reviewed tended to treat monitoring and evaluation as a single system and as a result cited similar contributing factors for these two areas. As discussed above, very few factors contributed positively to the monitoring and evaluation systems.

- Some evaluations noted that mid-term evaluations were conducted in a way that reflected an effective evaluation system. This included first-phase evaluations of 2-year programs (17 evaluations) with at least 10 of them indicating that some or all of the lessons/ recommendations from the mid-term evaluations were used;
- A few evaluations also pointed to presence of earlier and / or periodic evaluations of projects implemented by implementing partners (6 evaluations) – example of using the evaluation recommendation of project includes using participatory approach to design project.
- The presence of a program baseline, effective situation analysis, a realistic logic model and an operational M&E systems are some of the other positive factors cited in one or two evaluations. For the most part, however, the absence of these elements has been noted as

detracting factors, thereby stressing the need for improvement in these areas.

Evaluations identified several detracting factors for the effective use of evaluation and monitoring. Negative factors pointed out by some evaluations (mostly pertaining to sub-criteria 6.2 and 6.3) include:

- Poor quality or incomplete results / performance measurement frameworks (PMF) which can be attributed to baselines not linked to output indicators in the PMF, weak indicators that are poorly defined or overly ambitious and, indicators that are activity-related and not result-oriented (25 evaluations). This also included instances of not using the PMF even when present and/or weak;
- Weak or absent systematic monitoring systems leading to gaps in or no monitoring data (18 evaluations). This included absent tools for monitoring, no link between the PMF and monitoring, and failure to implement M&E systems;
- Absence of or weak/unreliable baseline information hindered monitoring or evaluating progress made (13 evaluations); and,
- An absent or weak results-based management culture which was linked to a tendency to focus on activity and / or financial and administrative reporting rather than on results, which highlight development effectiveness (11 evaluations).

Other detracting factors noted by a few evaluations include:

- The absence of a logical framework or logic model which also implies that there was no explicit theory of changes (4 evaluations);
- A lack of resources for M&E and reporting - this included absence of M&E staff, lack of RBM training and overall limited staff capacity in the project (4 evaluation); and,
- Weak knowledge of M&E and reporting among partners - government and NGOs (2 evaluations).

4.0 Conclusions

1. UN Women supported programs are **highly relevant** to the needs of target group members and are supportive of the national development plans and priorities of program countries (satisfactory or better in almost all evaluations). UN Women has also had success in developing effective partnerships with governments, national NGOs, civil societies and media in addressing gender equality and women's empowerment - contributing to program relevance. The relevance of UN Women programming has been enhanced by careful use of needs assessments, special studies and research into country conditions and the specific needs of target group members. The overall relevance of programming is supported by the

fact that most of the evaluations reported satisfactory or better results regarding the suitability of programs to the needs of the target group members, the effectiveness of partnerships and alignment to national development goals.

2. UN Women has largely been **effective in achieving the objectives** of its development programs, with 75% of the evaluations reporting findings of satisfactory or better. UN Women programs have also been effective in securing positive benefits for target group members and in supporting positive changes in national policies and programs, (more than 90% evaluations reported findings of satisfactory or better). Underlying factors for the achievement of objectives relate to UN Women’s convening ability to recognize and bring together key actors to work with them to address various gender equality and women’s empowerment issues and develop appropriate legislation, policies and frameworks. UN Women’s advocacy and its active lobbying and its role in building coalitions and networks have also facilitated achievement of objectives. Reasons mentioned for not achieving objectives included weak program design, often linked to over ambitious objectives, unclear causal relationship between various levels of results, lack of a results orientation, excessively short time frames and limited resources.
3. Linked to the previous criterion is UN Women’s performance with respect to **gender equality, which is highly effective**. More than 85% evaluations addressed gender equality and the results were very positive (about 80% evaluations that addressed gender equality reported findings of satisfactory or better). Some of the remaining 15% of the evaluations did report on gender equality results; however, they failed to report if those results met the stated objectives and/or to what extent. Consequently, this led the evaluation team to conclude that for these reports the criterion of gender equality was “not addressed”.⁵⁴ A key factor in these reported results is the mandate of UN Women to address gender equality and women’s empowerment, which means the programs align with gender equality priorities and apply human rights-based approaches. Many evaluations addressed the issue of ending VAW and gender-responsive budgeting and linked their positive findings to the technical expertise of UN Women in the field of gender equality. UN Women’s participatory approach and inclusive processes have also facilitated success in this area.
4. There was almost **no coverage of environmental sustainability** in the evaluations reviewed and hence this evaluation criterion did not merit a presentation of the findings reported. It appears that environmental sustainability has not been a primary focus of UN Women programming.

⁵⁴ See DER 3.1 sub-criteria scale – Annex 7. Furthermore, if an evaluation simply reported the number of women or girl receiving benefits, the evaluation was rated as “not addressed” for this sub-criterion.

-
5. Evaluation findings on **the sustainability of UN Women supported project and program results are mixed**. UN Women achieved positive ratings (more than two-thirds satisfactory or better) for its contributions to strengthening the enabling environment for development and strengthening institutional/community capacity, due to its participatory programming methodologies and inclusive approach. Nevertheless, the results for the likely continuation of benefits after project/program completion were negative and this poses a challenge for UN Women supported programming. Only 33% of the evaluations reported satisfactory or better results regarding the likelihood that program benefits would continue after funding had stopped. A major factor explaining these findings is the continued reliance by government and civil society partners on external program funding. This reflects both the inability of governments and other partners to financially support activities or their lack of interest and commitment to UN Women programs. Evaluations also reported deep-rooted negative cultural values and widespread stigma, high staff turnover at UN Women and partner organizations, and the absence of a clear and well-planned exit strategy as factors that contribute to the relatively poor sustainability of program benefits.
 6. The results for **efficiency of UN Women programming are mixed at best**, but not very positive regarding timeliness and the efficiency of systems and process for program implementation. Although the results for the cost-efficiency of programs were positive with about 60% of the evaluations reporting satisfactory or better, the coverage for this particular sub-criterion was only moderate. This makes it an area for continuing improvement by UN Women. A key factor that contributed to satisfactory findings of cost-efficiency was the ability to work with limited resources - through leveraging of resources and internal cost controls. However, the absence of appropriate and timely cost data collection by the program did not allow reasonable efficiency calculations and/or monitoring. Primarily administrative processes and/or delays in the release of funds negatively affected the timeliness of implementation and this in turn was linked to weak systems and procedures for program implementation and follow-up with about 33% of the evaluations reporting results of satisfactory or better. Key factors further limiting efficiency include weaknesses in staff capacity, staff turnover, excessive bureaucratic process requirements and internal management challenges, in addition to issues in UN Women's partner mechanisms and the transition from UNIFEM to UN Women.
 7. UN Women programs have been evaluated as **positive for evaluation systems and processes but negative for monitoring and results-based management**. The reported evaluation findings on UN Women evaluation systems and processes (67% of evaluations) and the use of evaluations to improve effectiveness (90% evaluations) have most often been found to be satisfactory or better. The use of evaluations to improve program design (and by extension development effectiveness) is supported by an increasing tendency to prepare detailed management responses to evaluation reports that include action plans, assign

management responsibility and include measures for reporting on the status of progress in responding to evaluation recommendations. In contrast, the findings for performance monitoring and reporting, and results-based management systems are predominantly negative with more than 80% of the evaluations reporting unsatisfactory or worse in either case. These are the two sub-criteria with the smallest number of evaluations reporting positive results. Despite UN Women's recent efforts, this has remained an area of concern. Weaknesses in the design of results frameworks, especially in the development of appropriate indicators and the measurement and reporting of baseline information often undermined UN Women's ability to effectively monitor and report on results according to the reviewed evaluations. Reporting often focused more on the completion of activities rather than on progress in achieving program results.

8. **Recent evaluations have shown more positive results** than earlier evaluations in 14 of the 18 sub-criteria reported. More specifically, positive results are significantly higher for recent evaluation for programs and projects making a difference for substantial number of beneficiaries, effectively addressing gender equality, strengthening enabling environment for sustainability and systems and process for evaluation are effective, than earlier evaluations. However, this is not the case for efficiency and sustainability issues: here the 2012/2011 evaluations have shown marginally more positive results than recent evaluations (2014/2013).

5.0 Lessons

Lessons for program and policy improvement based on evaluation reports reviewed for this DER, include:

1. UN Women is respected as an authority in its strategic focus areas of gender equality and women's empowerment. Sustainable changes in UN Women's strategic areas will take several years with the result that a long term focus is required for both its strategy and programming.
2. Many interventions show favourable signs of change; however, with lack of or inadequate monitoring systems, UN Women is not able to capture those results when they do happen. Results-based management is a systemic weakness in UN Women operations at every level. Many of the programs lack explicit theories of change, measurable results frameworks or adequate monitoring. This is inhibiting both accountability and learning.
3. Effective and innovative approaches to enhance inclusiveness in reaching out to women and marginalized groups have helped increase results in gender equality in UN Women supported programs. The approach to include men and religious leaders as active program participants facilitated programs' influence on changes in social and cultural norms, and increased the

possibility to progress and realize rights and gender equality for women by creating an enabling environment.

4. Strengthening the capacity of organizations at national and community level is one of the key characteristics of UN Women interventions. However, in many cases outcomes of the capacity development aspects of the program are not demonstrable due to weak result-based management and monitoring systems. While programs report on completion of activities or numbers of people trained, evidence is not often available on the impact of capacity development efforts.
5. Lack of financial resources and gaps in the technical capacity of national institutions and the political context in a country are threats to the sustainability of the results of UN Women programming. Nonetheless, the commitment of individuals and participatory processes of UN programs have resulted in some examples of locally owned initiatives. Ownership increases when local institutions/stakeholders are involved as early as during the design stages of the program.
6. The need for an exit strategy from UN Women program support was highlighted in many of the evaluations. This is an essential element to ensure sustainability and continuity of activities and benefits after program completion.
7. In order to improve the reporting of cost-efficiency or value for money of UN Women programs and interventions, it is important to have process and systems in place to ensure tracking of cost data throughout the program/project implementation process.

Annex 1: Development Effectiveness Criteria

1. Relevance
1.1 Programs are suited to the needs of target group members 1.2 Programs are aligned with national humanitarian and development goals 1.3 Effective partnerships with government, civil society and humanitarian and development partners
2. Achievement of Development Objectives and Expected Results
2.1 Programs and projects achieve stated objectives 2.2 Positive benefits for target group members 2.3 Substantial numbers of beneficiaries/contribution to national humanitarian and development goals 2.4 Significant changes in national development policies/programs
3. Cross Cutting Themes (Gender Equality and Environmental Sustainability)
3.1 Programs effectively address gender equality 3.2 Changes are environmentally sustainable
4. Sustainability
4.1 Program benefits are likely to continue 4.2 Programs support institutional and community capacity 4.3 Programs strengthen enabling environment for humanitarian and development
5. Efficiency
5.1 Program activities are cost efficient 5.2 Programs are implemented/objectives achieved on time 5.3 Systems for program implementation are efficient
6. Using Evaluation and Monitoring to Improve Development Effectiveness
6.1 Systems and processes for evaluation are effective 6.2 Systems and processes for monitoring are effective 6.3 Results based management systems are effective 6.4 Evaluation results used to improve development effectiveness

Annex 2: Evaluation Sample

Evaluation Title	Type	Office	Year
1.Evaluation of UN Joint Programme on Ending Violence Against Women in Ethiopia	Programme Evaluation	Ethiopia	2014
2.Final Evaluation of Programme "II National Plan of Policies for Women"	Cluster Evaluation	Brazil	2014
3.End of Programme Evaluation of EC-UN Women programme Supporting Gender Equality in the Context of HIV/AIDS	Programme Evaluation	Policy Division	2014
4.Evaluation of UN Women support to International Conference on the Great Lakes Region (ICGLR)	Thematic Evaluation	Regional Office for East and Southern Africa (Kenya)	2014
5.Action to Promote the Legal Empowerment of Women in the Context of HIV/AIDS	Programme Evaluation	Policy Division	2014
6. Consolidating Efforts that Contribute to Gender Equality and Women Empowerment in Sudan	Programme Evaluation	Sudan	2014
7.Final evaluation of the support to the Government of Afghan through the commission on the elimination of VAW to implement the law on eliminating violence against women and assistance to women victims	Programme Evaluation	Afghanistan	2013
8.Project 76742 "Promoting Gender Responsive Policies in the Western Balkans" Final evaluation	Programme Evaluation	Bosnia Herzegovina	2013
9.Project 76756 "Advancing the implementation of UNSCR 1325 in the Western Balkans" Final evaluation	Programme Evaluation	Bosnia Herzegovina	2013
10.Final evaluation of the Young Women Citizens programmes	Programme Evaluation	Brazil	2013
11.Evaluation finale du programme pays 2010-2013	Cluster Evaluation	Burundi	2013
12.Evaluation of Ethiopia UN Joint Programme on Gender Equality and Women's Empowerment	Programme Evaluation	Ethiopia	2013
13.Joint Evaluation on Joint Gender Programmes (JGPs) in the UN System	Thematic Evaluation	IEO	2013
14.Thematic Evaluation on the contribution of UN Women to increasing women's leadership and participation in peace and security and humanitarian response	Thematic Evaluation	IEO	2013
15.Thematic Evaluation on the Contribution of UN Women to Prevent Violence Against Women and Girls and Expand Access to Services	Thematic Evaluation	IEO	2013

Evaluation Title	Type	Office	Year
16.Final Evaluation of Women for Equality Peace and Development in Georgia project	Programme Evaluation	Georgia	2013
17.Final evaluation of the Safe Cities Programme Central American component (El Salvador and Guatemala) (ZONTA funded component)	Programme Evaluation	Guatemala	2013
18.Final evaluation of Gender & Governance Programme (GGP III Kenya), Gender and Governance Component	Programme Evaluation	Kenya	2013
19.Evaluation of Liberia JP for GE and Women's Economic Empowerment	Programme Evaluation	Liberia	2013
20.Evaluation of Gender and Governance Project in Mali	Programme Evaluation	Mali	2013
21.Evaluation of partnership agreement between the National Commission for the Development of Indigenous People and UN-Women	Programme Evaluation	Mexico	2013
22.Intervenções da ONU Mulheres na área de Prevenção e Eliminação da Violência contra a Mulher, com enfoque na Mobilização social no âmbito da Campanha UNiDOS pelo fim da violência contra a Mulher e Rapariga	Programme Evaluation	Mozambique	2013
23.Final Evaluation of Women connect across conflicts: Building accountability for implementation of the UN SCR 1325, 1820, 1888, 1889 Cross-regional project	Programme Evaluation	Multi-Country Office for Central Asia (Kazakhstan)	2013
24.Longitudinal Evaluation of Grenada Man to Man Batterer Intervention Programme	Programme Evaluation	Multi-Country Office for the Caribbean (Barbados)	2013
25.Final Evaluation of Project on Making Politics Work with Women (MP3W) in Nepal	Programme Evaluation	Nepal	2013
26.Final Evaluation of the programme on Sustaining the Gains of Foreign Labour Migration through the Protection of Migrant Workers Rights	Country-level Evaluation	Nepal	2013
27.Mehwar Center for the Protection and Empowerment of women and their families: 2008-2012	Programme Evaluation	Palestine	2013
28.Project 73055 "Advancing Women's Economic Rights in Serbia and Montenegro" Final Evaluation	Programme Evaluation	Regional Office for Europe and Central Asia (Turkey)	2013
29.Final Evaluation of Isange one-stop centres	Programme Evaluation	Rwanda	2013
30.Final Evaluation of the gender support programme	Programme Evaluation	Zimbabwe	2013
31.True, Justice and Reparation Programme for Women	Programme Evaluation	Colombia	2012

Evaluation Title	Type	Office	Year
32. Indigenous Regional Programme Evaluation	Programme Evaluation	Ecuador	2012
33. Project evaluation of Safe Cities in Quito	Programme Evaluation	Ecuador	2012
34. Final evaluation of Gender, Human Rights & Governance Programme (Kenya), Human Rights Component	Programme Evaluation	Kenya	2012
35. Evaluation of UN Women's Work on Gender Responsive Budgeting in India	Thematic Evaluation	Multi-Country Office for India, Bhutan, Maldives, and Sri Lanka	2012
36. Final evaluation of UN Women support to the Positive Women's Network on HIV/AIDS projects in India	Programme Evaluation	Multi-Country Office for India, Bhutan, Maldives, and Sri Lanka	2012
37. Final Evaluation of the project Cultural Heritage and Creative Industries as Vehicles for Development in Morocco	Regional Evaluation	Multi-Country Office for the Maghreb (Morocco)	2012
38. Multi-Sectoral Programme for the Fight Against GBV Through the Empowerment of Women and Girls in Morocco	Programme Evaluation	Multi-Country Office for the Maghreb (Morocco)	2012
39. GEPG Evaluation	Programme Evaluation	Multi-Country Office for the Pacific (Fiji)	2012
40. Establishment, Rehabilitation and Activation of Eight Women's Centres in the Gaza Strip and West Bank	Programme Evaluation	Palestine	2012
41. Final Evaluation Gender Equality and Women's Empowerment in the occupied Palestinian territory	Programme Evaluation	Palestine	2012
42. Final Evaluation of the MDG-F Joint Programme for Culture and Development	Programme Evaluation	Palestine	2012
43. SOCIALLY RESPONSIBLE WOMEN-RUN SCHOOL CANTEENS PROJECT: REPORT OF THE FINAL EVALUATION	Programme Evaluation	Palestine	2012
44. Final Evaluation: Integrating Gender Responsive Budgeting in Aid Effectiveness Agenda	Programme Evaluation	Policy Division	2012
45. Thematic evaluation of VAW across EHARO sub-region	Thematic Evaluation	Regional Office for East and Southern Africa (Kenya)	2012

Evaluation Title	Type	Office	Year
46. Inter-Agency Programme for the Promotion of Gender and Ethnic-Racial Equality Final Evaluation (Summative Evaluation)	Programme Evaluation	Brazil	2012
47. Partnership for Equality and Capacity Enhancement (PEACE): Towards Implementation of UNSCRs 1325 and 1820 Project	Programme Evaluation	Nepal	2012
48. Evaluation of the GRB Programme's UNV component (final evaluation UNV)	Programme Evaluation	Ecuador	2012
49. Evaluation of the Contribution of Italian Funded Projects to Ending Violence against Women and Promoting Gender Equality in ECOWAS Member States	Programme Evaluation	Regional Office for West and Central Africa (Senegal)	2012
50. Building capacities for Gender Equality in Governance and Protection of Women Rights in Sudan 2008-2011	Programme Evaluation	Sudan	2012
51. Final evaluation of supporting gender equality and women's rights in Timor-Leste	Programme Evaluation	Timor Leste	2012
52. Evaluation of the ACT to End Violence against Women (ACT VAW) Project	Project/ Programme	Arab States Regional Office	2011
53. Iraq National Constitution and Referendum Awareness Campaign (INCRAC) Project	Project/ Programme	Arab States Regional Office	2011
54. Final Project Evaluation: Women's rights to land and sustainable livelihoods in Kyrgyzstan	Project/ Programme	Commonwealth of Independent States Regional Office	2011
55. SHiEld Enhancing Prevention and Response to Domestic Violence in Georgia	Project/ Programme	Commonwealth of Independent States Regional Office	2011
56. Defending and securing human rights of women in Darfur	Project/ Programme	East and Horn of Africa Regional Office	2011
57. Programme Evaluation of SAARC Gender Info Base of SAARC-UN Women, South Asia Regional Office	Outcome	South Asia Regional Office	2011
58. Evaluation of SRS Capacity Development Activities (includes South-South Exchange Visits, Trainings and participation in international fora)	Cluster	Southern Africa Regional Office	2011
59. Evaluation of three Southern African regional women networks	Project/ Programme	Southern Africa Regional Office	2011
60. Evaluation of Country Program Liberia	Project/ Programme	West Africa Regional Office	2011
61. Evaluation of Peace Building Fund project in Sierra Leone	Project/ Programme	West Africa Regional Office	2011

Note: The highlighted rows are the six evaluations rejected based on quality review

Annex 3: Methodology

This annex provides a more detailed explanation of the population identification and sampling methodology used for the review of UN Women’s development effectiveness.

Evaluation Universe

The sampling process began with an extraction of UN Women evaluations published between 2011 and 2014 as reported on UN Women’s on-line evaluation database.⁵⁵ This list was augmented with evaluations identified during discussions with UN Women Independent Evaluation Office (IEO). As a result, the review team identified 79 evaluations published by UN Women for the period 2011 – 2014 as the evaluation universe for this review (Box 1).

Year	No. of Evaluations
2014	11
2013	26
2012	27
2011	15
Total	79

The list of 79 evaluation reports⁵⁶ was further scrutinized to arrive at the final sampling frame. The following were criteria used to exclude reports from the evaluation universe:

- Evaluations that were listed but did not have the final report uploaded in the database (2);
- If the reports were related to formative evaluation and / or strategic review (2); and,
- Reports explicitly referring to the work of UNIFEM (2).

Additionally, the two reports pertaining to the meta-evaluation and meta-analysis of evaluations in 2014 were separated from the sampling frame as these will be reviewed for qualitative information separately and relevant findings will be incorporated into the final report.

These exclusions led to a list of 71 evaluation reports, used by the review team as the sampling frame, for the meta-synthesis of evaluation results for UN Women.

Based on inputs and feedback from IEO and IOB, all corporate and thematic evaluations managed by IEO, HQ divisions and the Regional Offices (RO) were included in the sampling frame. These evaluations by their nature address issues of strategic significance on the triple mandate⁵⁷ of UN Women and their inclusion both on the qualitative and quantitative assessment

⁵⁵ <http://gate.unwomen.org/index.html>

⁵⁶ List of Evaluation Reports as of November 25, 2014 in the Evaluation database

⁵⁷ The unique role of UN Women reflects its strong mandate to work at both normative and operational levels as well as to ensure system-wide coordination. This requires an innovative evaluation function that leverages its partnerships and knowledge (UNW/2014/3).

is expected to enrich and provide a broader perspective of UN Women development effectiveness. It was also noted that UN Women corporate evaluations do not duplicate existing decentralized evaluations, as they collect and assess new evidence / data at country, regional and at global.

In identifying an appropriate sample of evaluations – 61 reports (Annex 2), the review team further assessed the sampling frame and excluded 12 mid-term evaluation reports from the sampling frame. This was done so that the evaluations reviewed would all deal with projects, which could reasonably be expected to have achieved their stated goals and produced meaningful development results.

Review Process and Data Analysis

In the course of completing the review, each evaluation report was reviewed by one member of a small review team that included two reviewers and the team leader. Each team member reviewed a set of evaluations. The first task of the reviewer was to assess the quality of the evaluation report to ensure that it was of sufficiently high quality to provide reliable information on development effectiveness. This was done using a quality-scoring grid (Annex 5). If the evaluation met the minimum score required, the reviewer continued to provide a rating on each sub-criterion, based on information in the evaluations and standard review grid (Annex 7). The reviewers also provided evidence from the evaluations to substantiate the ratings.

Significant quality assurance efforts were made to ensure consistency in the ratings by team members. The team leader and the team advisor trained the reviewers. A one-day workshop was held at which all team members reviewed and compared the ratings for the same evaluation. Two mid-term review meetings were also held to address any issues faced by the reviewers. The review team members also had regular communications to iron out any queries and ensure consistency in rating. Following completion of the reviews and the documentation of the qualitative evidence to support the ratings, the team leader reviewed all ratings to ensure there was sufficient evidence provided and it was consistent with the rating of each evaluation findings. These classifications were then cross-checked again by the team advisor. The team leader and the team advisor then reviewed the qualitative evidence for each sub-criterion to identify factors contributing to, or detracting from, the achievement of the sub-criteria.

The review of evaluation reports was also supplemented by a review of UN Women’s corporate documents. These were done to contextualize the results of the review.

Limitations

a) Sample Bias

The sample selected for this review is not a random sample. As with all DER review samples, it is intended as a purposive sample providing an illustration of how UN Women programming has

been assessed by the agency's own evaluation function. As such, the sample does not allow for or require the identification of confidence intervals around the reported findings. Nonetheless, as further illustrated in Annex 2, the 61 evaluations reviewed can be seen as broadly illustrative of UN Women programming during the period because:

- They cover a significant body of programming during the period measured in terms of budget allocations;
- Their distribution reflects the general distribution of UN Women program expenditures by region (although some regions such as Latin America and the Caribbean and Arab States are somewhat over-represented while Asia Pacific is under-represented). This is a reflection not of the sample but of the relatively higher quality of reports available from that region at the time of this review; and,
- The distribution of evaluation reports by year published reflects the overall evaluation output of UN Women in the years under review.

As well as establishing that the reports reviewed are reasonably illustrative of UN Women programming during the period (but not statistically representative), it is important to assess the extent that they actually report on the criteria used by the review. As illustrated in Figure 2 in the report, the review criteria were well covered in the evaluations reviewed, with one exception. Of the 19 sub-criteria used to assess development effectiveness, 12 were rated as strong in coverage (addressed in 41 to 55 evaluation reports) with six rated as moderate (addressed in 26 to 40 evaluation reports). Only one sub-criterion (environmental sustainability) was rated weak in coverage as it was addressed in only two evaluations. As a result, the findings reported in Chapter 3 do not address environmental sustainability. In comparison to those sampled for other Development Effectiveness Reviews, UN Women evaluation reports show a very strong pattern of coverage of the common criteria and sub-criteria.

b) The Challenge of Evaluations

The main body of UN Women evaluation reports reviewed are Project / Program evaluation reports since they feature so strongly in the population of published reports. As such, they provide several strengths to the DER process since they follow a common methodological approach and are consistently structured (hence they consistently cover the evaluation criteria). On the other hand, while country-level do cover all country level programming, they can present a challenge to analysts attempting to determine the evaluation reports overall (as opposed to sector or thematic) findings for a given criteria, especially where the report does not include a statement of the overall findings or conclusions for the criteria.

The review team dealt with this problem by systematically applying the guidance found in the *Operational Guidelines for Classifying Evaluation Findings* especially for the achievement of overall development objectives.

The review was not able to report systematically on the effectiveness of UN Women's programming focus area or classification of the country as the numbers became so small and/or by country classifications as several evaluations covered multiple countries.

As UN Women is a relatively young organization and as requested by IEO, the review team did classify the 55 evaluation reports reviewed by the recent two years (2014, 2013) and initial years of UN Women (2012, 2011) to compare the results across these two classifications. The grouping was done to ensure that there was sufficient number of reports to compare. (Annex 6).

c) Retrospective

Evaluations are, by definition, retrospective and a meta-synthesis is itself a retrospective approach, as it is based on a body of evaluations that address policies and programming implemented over an earlier period of time. UN Women's evaluations published in 2011 – 2014, covered years programming between 2009 and 2014. UN Women policies, strategies and approaches to programming have changed over these years, but the changes will not be reflected in all the evaluations.

Annex 4: Comparison of Population and the Sample

Table 1: Representation of Final Sample by Type of Evaluation

Type of Evaluation‡	Number of Evaluations	% of Sample (n =55)	% in Evaluation Universe (n=79)
Project / Programme†	45	81.8%	79.7%
Cluster	3	5.5%	6.3%
Thematic	4	7.3%	8.9%
Outcome	1	1.8%	1.3%
Regional	1	1.8%	1.3%
Country-level	1	1.8%	2.5%

† Includes evaluations classified as Programme Evaluations and Project / Programme

‡ As classified in evaluation database

Table 2: Representation of Sample by Year

Year	Number of Evaluations	% of Sample (n =60)	% in Evaluation Universe (n=79)
2014⁵⁸	6	10.9%	13.9%
2013	23	41.8%	32.9%
2012	18	32.7%	34.2%
2011	8	14.6%	19.4%

Table 3: Representation of Final Sample by Office / Unit Conducting the Evaluation

Office / Unit	Number of Evaluations	% of Sample (n =55)	% in Evaluation Universe (n=79)
Independent Evaluation Office	3	5.5%	6.3%
HQ Policy Division	3	5.5%	6.3%
Regional Office	11	20.0%	26.6%
CO / MCO†	38	69.1%	62.0%

† CO

– Country Office; and MCO – Multi-Country Office (since 2012)

⁵⁸ Out of the 15 “evaluation reports” listed in the database as of November 25, 2014 for 2014 – 4 of them do not have final reports uploaded, 1 is an evaluability assessment, 1 is a formative evaluation, 2 were meta evaluation / meta-analysis, 1 is a mid-term evaluation, and 2 of them are included in 2013 (as reports are dated 2013).

Table 4: Distribution of the Final Sample by Region / HQ

Region	Number of Evaluations	% of Sample (n=55)	% in Evaluation Universe (n=79)
Arab States	9	16.4%	12.7%
Asia and Pacific	7	12.7%	19.0%
EECA	7	12.7%	12.7%
ESA	13	23.6%	22.8%
WCA	4	7.3%	8.9%
LAC	9	16.4%	12.7%
HQ	6	10.9%	11.4%

EECA – Eastern Europe and Central Asia; ESA – East and Southern Africa; WCA – West and Central Africa; LCA – Latin American and Caribbean; and HQ – Headquarters

Table 5: Programme Expenses and Final Sample Distribution

Region	Programme Expenses (%)	% of Sample† (n=49)
Arab States	9.8%	18.4% (9)
Asia and Pacific	27.4%	14.3% (7)
EECA	9.0%	14.3% (7)
ESA	38.7%	26.5% (13)
WCA		7. (4)
LAC	15.1%	18.4% (9)

†Six evaluations of HQ IEO and Policy Division has not been included (55-6=49) to make the comparison with regional expenses

Annex 5: Evaluation Quality Scoring Guide

Quality Review

Each reviewer reviewed the quality of the evaluation, using a standard quality review guide (Table 1).⁵⁹ The guide reflects the criterion being rated and how the maximum number of points is allocated for each criterion. This ensured that:

1. Evaluations being used to provide information on UN Women programming were of sufficiently overall high quality to be credible evaluations. This resulted in an overall quality score that had a maximum of 40 points. An evaluation had to score a minimum of 25 points to be included in the review; and,
2. Even if the evaluations were generally of high quality, the ratings for Criteria G (multiple lines of evidence), Criteria H (evaluation design) and Criteria I (evaluation findings and conclusions that are relevant and evidence) were also sufficiently high to ensure the evaluation would provide solid information specifically with respect to measuring effectiveness. A total of 12 points was available for these three criteria and an evaluation had to receive a minimum of nine points in on these three criteria.

Quality Review Grid

	Points for criteria scored	Maximum Points	Score
A	Purpose of the evaluation <ul style="list-style-type: none"> - why the evaluation was done (1) - what triggered the evaluation (including timing in the project/program cycle) (1) - how evaluation is to be used (1) 	3	
B	Evaluation objectives <ul style="list-style-type: none"> - evaluation objectives are clearly stated (1) - objectives logically flow from purpose (1) 	2	

⁵⁹ The quality scoring guide used was the one revised for UNICEF (presented above) as it was more appropriate for the UN Women DER also. In this guide on criteria H (Evaluation design) – “baseline data (quantitative or qualitative) on conditions prior to program implementation” and “comparison of conditions after program delivery to those before” were combined into a single sub-criteria based on past experience. This reduced the maximum points that can be scored for Evaluation Design from 5 to 4. So the maximum total for Criteria G, H, and I is 12 (as against the original total of 13). Due to this change the minimum required for Criteria G, H and I was revised downward from 10 to 9.

At the same time, in Criteria F (Evaluation Criteria), the cross-cutting issues of gender equality and environmental sustainability were separated. Hence the total point for H - evaluation criteria became 6 from 5. In the original guide they were scored together, in which case most UN Women evaluation will not have scored any point on cross-cutting issues. The total remained the same at 40 due to these changes and so the minimum required also remained the same at 25.

	Points for criteria scored	Maximum Points	Score
C	<p>Organization of the evaluation</p> <ul style="list-style-type: none"> - logical structure to the organization of the evaluation (1) - evaluation is well written (1) - clear distinction between evidence, findings, conclusions, and recommendations (1) 	3	
D	<p>Subject evaluated is clearly described</p> <p>Evaluation describes:</p> <ul style="list-style-type: none"> - the activity/program being evaluated (1) - the program's expected achievements (for humanitarian: identification of needs of target population and agency specific role in addressing needs) (1) - how the program addresses the development problem (1) - the implementation modalities used (1) 	4	
E	<p>Scope of the evaluation</p> <p>Evaluation defines the boundaries of the evaluation in terms of:</p> <ul style="list-style-type: none"> - time period covered (1) - implementation phase under review (1) - geographic area (1) - dimensions of stakeholder involvement being examined (1) 	4	
F	<p>Evaluation criteria</p> <p>Evaluation criteria include:</p> <ul style="list-style-type: none"> - the achievement of development objectives and expected results (including impacts) (1) - cross-cutting issues: inclusive development which is gender sensitive (1) - cross-cutting issues: inclusive development which is environmentally sustainable (1) - the sustainability of benefits and positive results achieved (1) - the relevance of MO activities and supported projects and programs (1) - the efficiency of MO operations in support of projects and programs (1) 	6	
G	<p>Multiple lines of evidence</p> <ul style="list-style-type: none"> - one point (1) for each line of evidence used (case studies, surveys, site visits, interviews, etc.), up to a maximum of four points (4) 	4	
H	<p>Evaluation design</p> <p>Elements of a good evaluation design include:</p> <ul style="list-style-type: none"> - an explicit theory of how objectives and results were to be achieved (1) - specification of the level of results achieved (output, outcome, impact) (1) - baseline data (quantitative or qualitative) on conditions prior to program implementation and/or comparison of conditions after program delivery to those before (1) - a qualitative or quantitative comparison of conditions among program participants and a control group (1) 	4	
I	<p>Evaluation findings and conclusions are relevant and evidence based</p> <p>Evaluation report includes:</p> <ul style="list-style-type: none"> - evaluation findings relevant to the assessment criteria (1) - findings that are supported by the chosen methodologies (1) - a clear logical link between the evidence and the finding (1) - conclusions which are linked to the evaluation findings as reported (1) 	4	

	Points for criteria scored	Maximum Points	Score
J	Evaluation limitations - statement of the limitations of the methodology (1) - impact of limitations on evaluation (1) - remedies of limitations (1)	3	
K	Evaluation Recommendations - evaluation contains recommendations that flow from findings and conclusions (1) - recommendations are directed to one or more organization (1) - recommendations are aimed at improving development effectiveness (1)	3	
	Total (required to have a minimum of 25 points)	40	
	Total for Criteria G, H and I (required to have minimum of 9 points)	12	

Results of Quality Review Scoring

Overall evaluation quality scores			Development effectiveness key criteria		
Max points = 40 (Min required = 25)	No. of evaluations	% of evaluations	Max points = 12 (Min required = 9)	No. of evaluations	% of evaluations
36 - 40	19	34.5%	12	3	5.5%
31 - 35	29	52.7%	11	15	27.3%
25 - 30	7 + (4)	20.0%	10	17	30.9%
21 - 24	(2)	3.6%	9	20	36.4%
16 - 20	-	-	8	(3)	5.5%
11 - 15	-	-	7	(3)	5.5%
Total	61 55 + (6)	100.0%	Total	61 55 + (6)	100.0%

Out of the 61 reports reviewed for quality, 6 were rejected based on quality scores. The figures in brackets indicate the number of reports that failed quality screening. From the table above it could be seen that although there were four reports that had an overall score of 25 or more, they failed because they did not meet the minimum score of 9 required for Criteria, G, H and I.

The 55 reports that passed the quality were used to complete the review and analysis.

Annex 6: Comparison Between Recent and Initial Years

- Total number of 2014/2013 (recent) evaluations – 29; and total number of 2012/2011 (earlier) evaluations – 26
- Valid # - Number of evaluations that addressed the sub-criteria
- Legend – HUS – Highly Unsatisfactory; US – Unsatisfactory; S – Satisfactory; HS – Highly Satisfactory

	Years	HUS	US	S	HS	Total	Valid #	S&HS
1.Relevance								
1.1 Programs and projects are suited to the needs and/or priorities of the target group	2014/2013	-	7.4%	63.0%	29.6%	100%	27	92.6%
	2012/2011	-	16.7%	70.8%	12.5%	100%	21	83.3%
1.2 Projects and programs align with national development goals	2014/2013	-	-	48.0%	52.0%	100%	25	100%
	2012/2011	-	5.0%	70.0%	25.0%	100%	15	95.0%
1.3 Effective partnerships with governments, bilateral and multilateral development and humanitarian organizations and Non-governmental organizations	2014/2013	-	21.4%	60.7%	17.9%	100%	28	78.6%
	2012/2011	-	34.8%	65.2%	-	100%	23	65.2%
2.Achievement of Development Objectives								
2.1 Programs and projects achieve their stated humanitarian and development objectives and attain expected results	2014/2013	-	20.0%	68.0%	12.0%	100%	25	80.0%
	2012/2011	-	31.8%	68.2%	-	100%	22	68.2%
2.2 Programs and projects have resulted in positive benefits for target group members.	2014/2013	-	-	92.9%	7.1%	100%	28	100%
	2012/2011	-	8.0%	92.0%	-	100%	25	92.0%
2.3 Programs and projects made differences for a substantial number of beneficiaries and where appropriate contributed to national humanitarian and development goals.	2014/2013	-	26.7%	66.7%	6.7%	100%	15	73.4%
	2012/2011	-	63.6%	36.4%	-	100%	11	36.4%
2.4 Programs contributed to significant changes in national development policies and programs (policy impacts) and/or to needed system reforms	2014/2013	-	3.8%	88.5%	7.7%	100%	26	96.2%
	2012/2011	-	16.7%	83.3%	-	100%	18	83.3%
3.Cross-cutting Theme (Gender Equality)								
3.1 Extent to which multilateral organization supported activities effectively address the cross-	2014/2013	-	11.5%	80.8%	7.7%	100%	26	88.5%
	2012/2011	-	31.8%	68.2%	-	100%	22	68.2%

	Years	HUS	US	S	HS	Total	Valid #	S&HS
cutting issue of gender equality.								
4.Sustainability								
4.1 Benefits continuing or likely to continue after project or program completion	2014/2013	3.8%	73.1%	19.2%	3.8%	100%	26	23.1%
	2012/2011	-	54.5%	45.5%	-	100%	22	45.1%
4.2 Projects and programs are reported as sustainable in terms of institutional and/or community capacity.	2014/2013	-	25.9%	74.1%	-	100%	27	74.1%
	2012/2011	-	31.8%	59.1%	9.1%	100%	22	68.2%
4.3 Programming contributes to strengthening the enabling environment for humanitarian and development	2014/2013	5.0%	5.0%	90.0%	-	100%	20	90.0%
	2012/2011	-	37.5%	62.5%	-	100%	16	62.5%
5.Efficiency								
5.1 Program activities are evaluated as cost/resource efficient	2014/2013	-	40.0%	50.0%	10%	100%	20	60.0%
	2012/2011	-	38.5%	61.5%	-	100%	13	61.5%
5.2 Implementation and objectives achieved on time (given the context, in the case of humanitarian programming).	2014/2013	-	75.0%	25.0%	-	100%	20	25.0%
	2012/2011	5.0%	55.0%	40.0%	-	100%	20	40.0%
5.3 Systems and procedures for project/program implementation and follow up are efficient (including systems for engaging staff, procuring project inputs, disbursing payment, logistical arrangements etc.).	2014/2013	3.7%	63.0%	33.3%	-	100%	27	33.3%
	2012/2011	16.7%	45.8%	37.5%	-	100%	24	37.5%
6.Use of Evaluation and Monitoring								
6.1 Systems and process for evaluation are effective	2014/2013	-	22.2%	72.2%	5.6%	100%	18	77.8%
	2012/2011	-	45.5%	54.5%	-	100%	22	54.5%
6.2 Systems and processes for monitoring and reporting on program results are effective	2014/2013	7.7%	76.9%	11.5%	3.8%	100%	26	15.4%
	2012/2011	16.0%	72.0%	12.0%	-	100%	25	12.0%
6.3 Results based management systems are effective.	2014/2013	4.8%	81.0%	9.5%	4.8%	100%	21	14.1%
	2012/2011	16.7%	73.7%	5.3%	5.3%	100%	18	10.5%
6.4 Evaluation is used to improve humanitarian and development effectiveness.	2014/2013	6.9%	-	10.3%	82.8%	100%	29	93.1%
	2012/2011	-	11.5%	15.4%	73.1%	100%	26	88.5%

Annex 7: Guide for Review Team to Classify Evaluation Findings

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
Common Development Evaluation Assessment Criteria				
1. Relevance				
1.1 MO supported programs and projects are suited to the needs and/or priorities of the target group	Substantial elements of program or project activities and outputs were unsuited to the needs and priorities of the target group.	No systematic analysis of target group needs and priorities took place during the design phase of developmental or relief and rehabilitation programming or there is some evident mismatch between program and project activities and outputs and the needs and priorities of the target group.	MO supported activity, program or project is designed taking into account the needs of the target group as identified through a process of situation or problem analysis (including needs assessment for relief operations) and the resulting activities are designed to meet the needs of the target group.	MO supported programs and projects are suited to the needs and/or priorities of the target group
1.2 MO supported development projects and programs align with national development goals:	Significant elements of MO supported development program and project activity run counter to national development priorities with a resulting loss of effectiveness.	<u>Significant portion</u> (1/4 or more) of the MO supported development programs and projects are not aligned with national plans and priorities, but there is no evidence that they run counter to those priorities.	<u>Most</u> MO supported development programs and projects are fully aligned with national plans and priorities as expressed in national poverty eradication and sector plans and priorities. Wherever MO supported programs and projects are reported in the evaluation as not directly supportive of national plans and priorities they do not run counter to those priorities.	MO supported development projects and programs align with national development goals:
1.3 MO has developed an effective partnership with	MO experiences significant divergence in priorities from	MO has experienced significant difficulties in	MO has improved the effectiveness of its partnership	MO has developed an effective partnership with governments,

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
governments, bilateral and multilateral development and humanitarian organizations and NGOs for planning, coordination and implementation of support to development and/or emergency preparedness, humanitarian relief and rehabilitation efforts.	those of its (government, NGO or donor) partners and lacks a strategy or plan which will credibly address the divergence and which should result in strengthened partnership over time.	developing an effective relationship with partners and that there has been significant divergence in the priorities of the MO and its partners.	relationship with partners over time during the evaluation period and that this partnership was effective at the time of the evaluation or was demonstrably improved.	bilateral and multilateral development and humanitarian organizations and NGOs for planning, coordination and implementation of support to development and/or emergency preparedness, humanitarian relief and rehabilitation efforts.
2. Achievement of Objectives and Expected Results				
2.1 MO supported programs and projects achieve their stated development and/or humanitarian objectives and attain expected results.	Less than half of stated output and outcome objectives have been achieved including one or more very important output and/or outcome level objectives.	Half or less than half of stated output and outcome level objectives are achieved.	MO supported programs and projects either achieve at least a majority of stated output and outcome objectives (more than 50% if stated) or that the most important of stated output and outcome objectives are achieved.	MO supported programs and projects achieve all or almost all significant development and/or humanitarian objectives at the output and outcome level.
2.2 MO supported programs and projects have resulted in positive benefits for target group members.	Problems in the design or delivery of MO supported activities mean that expected positive benefits for target group members have not occurred or are unlikely to occur.	MO supported projects and programs result in no or very few positive changes experienced by target group members. These benefits may include the avoidance or reduction of negative effects of a sudden onset or protracted emergency.	MO supported projects and programs have resulted in positive changes experienced by target group members (at the individual, household or community level). These benefits may include the avoidance or reduction of negative effects of a sudden onset or protracted emergency.	MO supported projects and programs have resulted in widespread and significant positive changes experienced by target group members as measured using either quantitative or qualitative methods (possibly including comparison of impacts with non-program participants). These benefits may include the avoidance or reduction of negative effects of a sudden onset or protracted emergency.
2.3 MO programs and projects made differences for a substantial number of beneficiaries and where appropriate contributed to	MO supported projects and programs have not contributed to positive changes in the lives of beneficiaries as measured quantitatively or qualitatively.	MO supported projects and programs have contributed to positive changes in the lives of only a small number of beneficiaries	MO supported projects and programs have contributed to positive changes in the lives of substantial numbers of beneficiaries as measured	MO supported projects and programs have contributed to positive changes in the lives of substantial numbers of beneficiaries. Further, they have contributed to the achievement of

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
national development goals.		(when compared to project or program targets and local or national goals if established).	quantitatively or qualitatively. These may result from development, relief, or protracted relief and rehabilitation operations and may include the avoidance of negative effects of emergencies.	specific national development goals or have contributed to meeting humanitarian relief objectives agreed to with the national government and/or national and international development and relief organizations.
2.4 MO activities contributed to significant changes in national development policies and programs (including for disaster preparedness, emergency response and rehabilitation) (policy impacts) and/or to needed system reforms.	National policies and programs in a given sector or area of development (including disaster preparedness, emergency response and rehabilitation) were deficient and required strengthening but MO activities have not addressed these deficiencies.	MO activities have not made a significant contribution to the development of national policies and programs in a given sector or area of development, disaster preparedness, emergency response or rehabilitation. (Policy changes in humanitarian situations may include allowing access to the effected populations).	MO activities have made a substantial contribution to either re-orienting or sustaining effective national policies or programs in a given sector or area of development disaster preparedness, emergency response or rehabilitation.	MO activities have made a substantial contribution to either re-orienting or sustaining effective national policies or programs in a given sector or area of development disaster preparedness, emergency response or rehabilitation. Further, the supported policies and program implementation modalities are expected to result in improved positive impacts for target group members.
3. Cross Cutting Themes: Inclusive Development Which can be Sustained				
3.1 Extent MO supported activities effectively address the cross-cutting issue of gender equality.	MO supported activities are unlikely to contribute to gender equality or may in fact lead to increases in gender inequalities.	MO supported activities either lack gender equality objectives or achieve less than half of their stated gender equality objectives. (Note: where a program or activity is clearly gender-focused (maternal health programming for example) achievement of more than half its stated objectives warrants a satisfactory	MO supported programs and projects achieve a majority (more than 50%) of their stated gender equality objectives.	MO supported programs and projects achieve all or nearly all of their stated gender equality objectives.

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
		rating).		
3.2 Extent changes are environmentally sustainable.	MO supported programs and projects do not include planned activities or project design criteria intended to promote environmental sustainability. In addition changes resulting from MO supported programs and projects are not environmentally sustainable.	MO supported programs and projects do not include planned activities or project design criteria intended to promote environmental sustainability. There is, however, no direct indication that project or program results are not environmentally sustainable. OR MO supported programs and projects include planned activities or project design criteria intended to promote sustainability but these have not been successful.	MO supported programs and projects include some planned activities and project design criteria to ensure environmental sustainability. These activities are implemented successfully and the results are environmentally sustainable.	MO supported programs and projects are specifically designed to be environmentally sustainable and include substantial planned activities and project design criteria to ensure environmental sustainability. These plans are implemented successfully and the results are environmentally sustainable.
4. Sustainability				
4.1 Benefits continuing or likely to continue after project or program completion or there are effective measures to link the humanitarian relief operations, to rehabilitation, reconstructions and, eventually, to longer-term developmental results.	There is a very low probability that the program/project will result in continued intended benefits for the target group after project completion. For humanitarian relief operations, the evaluation finds no strategic or operational measures to link relief, to rehabilitation, reconstruction and, eventually, to development.	There is a low probability that the program/project will result in continued benefits for the target group after completion. For humanitarian relief operations, efforts to link the relief phase to rehabilitation, reconstruction and, eventually, to development are inadequate. (Note, in some circumstances such linkage may not be possible due to the context of the	Likely that the program or project will result in continued benefits for the target group after completion. For humanitarian relief operations, the strategic and operational measures to link relief to rehabilitation, reconstruction and, eventually, development are credible.	Highly likely that the program or project will result in continued benefits for the target group after completion. For humanitarian relief operations, the strategic and operational measures to link relief to rehabilitation, reconstruction and, eventually, development are credible. Further, they are likely to succeed in securing continuing benefits for target group members.

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
		emergency. If this is stated in the evaluation, a rating of satisfactory can be given)		
4.2 Extent MO supported projects and programs are reported as sustainable in terms of institutional and/or community capacity.	The design of MO supported programs and projects failed to address the need to strengthen institutional and/or community capacity as required. In the case of humanitarian operations, the design of programs and projects failed to take account of identified needs to strengthen local capacities for delivery of relief operations and/or for managing the transition to rehabilitation and/or development.	MO programs and projects may have failed to contribute to strengthening institutional and/or community capacity or, where appropriate, to strengthen local capacities for delivery of relief operations and/or for managing the transition to rehabilitation and/or development.	MO programs and projects may have contributed to strengthening institutional and/or community capacity but with limited success.	Either MO programs and projects have contributed to significantly strengthen institutional and/or community capacity as required or institutional partners and communities already had the required capacity to sustain program results.
4.3 Extent MO development programming contributes to strengthening the enabling environment for development.	For development programs, there were important weaknesses in the enabling environment for development (the overall framework and process for national development planning; systems and processes for public consultation and for participation by civil society in development planning; governance structures and the rule of law; national and local mechanisms for accountability for public expenditures, service delivery and quality; and necessary improvements to supporting structures such as	MO development activities and/or MO supported projects and programs have not made a notable contribution to changes in the enabling environment for development.	MO development activities and/or MO supported projects and programs have made a notable contribution to changes in the enabling environment for development including one or more of: the overall framework and process for national development planning; systems and processes for public consultation and for participation by civil society in development planning; governance structures and the rule of law; national and local mechanisms for accountability for public expenditures, service delivery and quality; and necessary	MO development activities and/or MO supported projects and programs have made a significant contribution to changes in the enabling environment for development including one or more of: the overall framework and process for national development planning; systems and processes for public consultation and for participation by civil society in development planning; governance structures and the rule of law; national and local mechanisms for accountability for public expenditures, service delivery and quality; and necessary improvements to supporting structures such as capital and labour markets. Further,

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
	capital and labour markets). Further, the MO activities and support provided to programs and projects failed to address the identified weakness successfully, further limiting program results.		improvements to supporting structures such as capital and labour markets.	these improvements in the enabling environment are leading to improved development outcomes.
5. Efficiency				
5.1 Program activities are evaluated as cost/resource efficient:	Credible information indicating that MO supported programs and projects (development, emergency preparedness, relief and rehabilitation) are not cost/resource efficient.	MO supported programs and projects under evaluation (development, emergency preparedness, relief and rehabilitation) do not have credible, reliable information on the costs of activities and inputs and therefore the evaluation is not able to report on cost/resource efficiency. OR MO supported programs and projects under evaluation present mixed findings on the cost/resource efficiency of the inputs.	Level of program outputs achieved (development, emergency preparedness, relief and rehabilitation) when compared to the cost of program activities and inputs are appropriate even when the program design process did not directly consider alternative program delivery methods and their associated costs.	MO supported (development, emergency preparedness, relief and rehabilitation) programs and projects are designed to include activities and inputs that produce outputs in the most cost/resource efficient manner available at the time.

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
5.2 Evaluation indicates implementation and objectives achieved on time (given the context, in the case of humanitarian programming)	Less than half of stated output and outcome level objectives of MO supported programs and projects are achieved on time, there is no credible plan or legitimate explanation found by the evaluation which would suggest significant improvement in on-time objectives achievement in the future.	Less than half of stated output and outcome level objectives of MO supported programs and projects are achieved on time but the program or project design has been adjusted to take account of difficulties encountered and can be expected to improve the pace of objectives achievement in the future. In the case of humanitarian programming, there was a legitimate explanation for the delays.	More than half of stated output and outcome level objectives of MO supported programs and projects are achieved on time and that this level is appropriate to the context faced by the program during implementation, particularly for humanitarian programming.	Nearly all stated output and outcome level objectives of MO supported programs and projects are achieved on time or, in the case of humanitarian programming, a legitimate explanation for delays in the achievement of some outputs/outcomes is provided.
5.3 Evaluation indicates that MO systems and procedures for project/program implementation and follow up are efficient (including systems for engaging staff, procuring project inputs, disbursing payment, logistical arrangements etc.)	Serious deficiencies in agency systems and procedures for project/program implementation that result in significant delays in project start-up, implementation or completion and/or significant cost increases.	Some deficiencies in agency systems and procedures for project/program implementation but does not indicate that these have contributed to delays in achieving project/program objectives.	Agency systems and procedures for project implementation are reasonably efficient and have not resulted in significant delays or increased costs.	Efficiency of agency systems and procedures for project implementation represent an important organizational strength in the implementation of the program under evaluation.

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
6. Using Evaluation and Monitoring to Improve Development Effectiveness				
6.1 Systems and process for evaluation are effective.	Evaluation practices in use for programs and projects of this type (development, emergency preparedness, relief and rehabilitation) are seriously deficient.	No indication that programs and projects of this type (development, emergency preparedness, relief and rehabilitation) are subject to systematic and regular evaluations.	Program being evaluated is subject to systematic and regular evaluations or describes significant elements of such practice. No mention of policy and practice regarding similar programs and projects. This may include specialized evaluation methods and approaches to emergency preparedness, relief and rehabilitation programming.	Program being evaluated (along with similar programs and projects) is subject to systematic regular evaluations or describes significant elements of such practice.
6.2 Systems and processes for monitoring and reporting on program results are effective	Absence of monitoring and reporting systems for the development and humanitarian programming. This would include the absence of adequate monitoring of outputs during the implementation of humanitarian programming.	While monitoring and reporting systems for the development and humanitarian programming exist, either they do not report on a regular basis or they are inadequate in frequency, coverage or reliability.	Monitoring and reporting systems for development and humanitarian programming as appropriate are well established and report regularly.	Monitoring and reporting systems for the program are well established and report regularly. The quality of regular reports is rated highly by the evaluation and results are reportedly used in the management of the program.
6.3 Results Based Management (RBM) systems are effective	No evidence of the existence of an RBM system for the program and no system is being developed.	While an RBM system is in place, or being developed, it is unreliable and does not produce regular reports on program performance.	RBM system is in place and produces regular reports on program performance.	RBM system is in place for the program and there is evidence noted in the evaluation that the system is used to make changes in the program to improve effectiveness.
6.4 MO makes use of evaluation to improve development/humanitarian effectiveness	Report does not include a management response and does not have one appended to it or associated with it. There is no indication of how the	Report includes a management response (or has one attached or associated with it) but it does not indicate which	Report includes a management response (or has one attached or associated with it) that indicates which recommendations have been accepted.	Report includes a management response (or has one attached or associated with it) describes a response to each major recommendation which is

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
	<p>evaluation results will be used. There is no indication that similar evaluations have been used to improve effectiveness in the past.</p>	<p>recommendations have been accepted. OR There is some, non-specific indication that similar evaluations have been used to improve program effectiveness in the past.</p>	<p>OR There is a clear indication that similar evaluations in the past have been used to make clearly identified improvements in program effectiveness.</p>	<p>appropriate and likely to result in the organizational and programmatic changes needed to achieve their intent.</p>

Annex 8: UN Women Corporate Documents

- Annual report 2013-2014
- Annual report 2012-2013
- UN Women Strategic Plan 2011-2013 (May 2011)
- UN Women Strategic Plan 2014-2017
- UN Women 2014 -2017 Evaluation Strategic Plan
- Report on the Evaluation Function of UN Women, 2012 (April 2013);
- UN Women Evaluation Policy (UNW/2012/12)
- Report of the Advisory Committee for period January 1 to December 31, 2013 (April 2014)
- Joint Statement to the UN Women Executive Board by Netherlands on behalf of Canada, Norway and Netherlands (June 2013)
- Report of the Under-Secretary-General/Executive Director on the progress made in implementing the Strategic Plan, 2011–2013 for the year 2013 (UNW/2014/2)
- Data Companion to the Report of the Under-Secretary-General/Executive Director on the progress made in implementing the Strategic Plan, 2011-2013
- Progress made on the UN-Women Strategic Plan 2011–2013 (2012/4)
- Report of the Under-Secretary-General/Executive Director on the regional architecture review (UNW/2012/5)
- Statements by other board members on the evaluation Function/Meta-analysis (Ex-Board June 2014 meeting)
- Statement of Switzerland on meta-analysis
- Meta-Analysis- What can we learn from UN Women evaluations?, July 2014
- GERAAS Meta-evaluation, March 2014
- MOPAN 2014 – Synthesis Report Executive Summary – UN Women
- MOPAN 2015 – Technical Report – UN Women – Volume 1 Results by Micro-indicator and by Country
- Professional Peer Review of the Evaluation Function of UN Women – Peer Review Panel Final Report – September 2014
- Report of the Audit Advisory Committee for the period 1 January to 31 December 2013 (UNW2014/4/Add.1)
- Report of the Global Evaluation Advisory Committee on the external assessments of the evaluation function of the United Nations Entity for Gender Equality and the Empowerment of Women, January 2015 (UNW/2015/CRP.1)
- UN Women (2014) Ten Steps to Strengthened RBM in UN Women